

Genetic screens to identify factors  
pertinent to host defence against  
bacterial infection

**Josephine Duncombe-Moore**

**University of Strathclyde**

**Strathclyde Institute of Pharmacy and Biomedical Sciences**

A thesis presented in fulfilment of the requirements for the degree  
of

**Doctor of Philosophy**

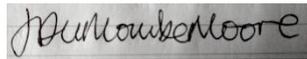
December 2020

## DECLARATION

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

A rectangular box containing a handwritten signature in black ink, which appears to read "J. H. Moore".

Date: 23/02/2021

## ABSTRACT

Antibiotic resistance is a growing concern for healthcare providers across the world. Indeed, resistance is now being found to those ‘last-ditch’ antibiotics reserved for antibiotic-resistant infections. There is great need for research into alternative therapies for globally important pathogens, as well as those that pose a bioterror threat due to their infectivity and high morbidity (e.g. *Francisella tularensis*). Host-directed therapy as a concept for infection treatment seeks either to block bacterial invasion and growth within host cells or enhance host bactericidal activity. This approach has the potential to treat a broad range of bacterial infections, as well as to reduce the likelihood of developing resistance.

As a first step, we explored a host infection response network (including the gene *PCDH7*) defined on the basis of results from a previous HEK-293 host infection screen. Cellular knockout of human *PCDH7* showed reduced intracellular *Salmonella enterica* serotype Typhimurium (STM) and *Shigella sonnei* burden *in vitro*, suggesting resistance to bacterial growth.

To identify further targets, a ‘gene trap’ mutation library was generated in a macrophage-like human cell line, U937, which was then differentiated and infected with STM and *F. tularensis* LVS in independent screens. RNA-Seq was performed on the infected and control populations to identify functionally vital host defence gene mutations. The most statistically significant gene mutations were assessed using pathway analysis tools and literature searches. Multiple pathway analyses converged on electron transport chain subunits *MT-ND5*, *MT-ND6* and *MT-CO1*, the trapped versions of which were identified as protective in the STM screen. Furthermore, another trapped protective hit, *SLC7A11*, shows promise from initial validation using a CRISPR knockout (lower intracellular STM burden) as well as pharmacological modulation with the inhibitor sulfasalazine.

This work has provided a starting point for the investigation of human genes and cellular processes that might be amenable to pharmacological manipulation to provide protection against, or recovery from, bacterial infection. Therefore host-directed therapies merit further exploration as a novel route to counter the potentially devastating impact of bacteria largely resistant to current antimicrobial drugs.

## ACKNOWLEDGMENT

I would like to thank my supervisors Dr Benjamin Pickard, Prof Craig Roberts, Dr Jun Yu and my DSTL contact, Dr Riccardo D’Elia, for their advice and guidance during the course of my PhD. I would also like to thank these people, as well as DSTL, for giving me the opportunity and funding to start this PhD, and for being patient when I struggled with it. I would especially like to thank Dr Benjamin Pickard for his persistence and patience in correcting my writing, as well as giving me laboratory support when I had my hands full.

I would like to thank my parents for being there when I wanted to talk and Lorraine Glass, the student counsellor who showed me how to pick myself up again. I would like to thank Dr Wenqin Li and Intisar Albandar for their continued companionship, shared experiences and technical aid through my time at Strathclyde.

I would like to thank the contributors (including Dr Rasha Mahmoud, Dr Wenqin Li and Dr Stuart Woods) to the Centre for Defence Enterprise project for generating the results that allowed this project to come into being. I would like to thank the Masters and undergraduate students who contributed to the work in this thesis, Jason Dick, Robyn Maitland and Yiwen Ding. I would like to thank Helen Flick-Smith and Dr Ruth Holloway for their teaching and laboratory collaboration while at DSTL and Strathclyde respectively. I would like to thank both for the method development collaboration to improve the bacterial burden assay and testing CRISPR control cell infection phenotypes at DSTL. I would like to thank everyone in the CBR Services department, especially Riccardo’s team, for making me feel welcome and chaperoning me around the laboratories.

I would also like to thank Stuart Woods for showing me how to perform *in vitro Toxoplasma* infections and giving me advice on flow cytometry. I would like to thank Dr Thaer Hasan for showing me the protocol for bacterial infections and giving me advice about working with *Shigella sonnei*.

I would like to thank Juudit Gross for my being my climbing buddy and later my flat mate. Finally, I would like to give special thanks to my boyfriend Tom, for being wonderful, funny, calm, kind and supportive – as a flat mate, a rubber duck, a friend, and perhaps most importantly as a partner.

## ABBREVIATIONS

<b>ATP</b>	Adenosine triphosphate
<b>ANOVA</b>	Analysis of variance
<b>ATG</b>	Autophagy-related
<b>ATRA</b>	All-trans-retinoic acid
<b>BMDM</b>	(Mouse) bone marrow-derived macrophage
<b>bp</b>	Base pairs
<b>BSL</b>	Biosafety level
<b>CC BY</b>	Creative Commons Attribution
<b>CDC</b>	Centers for Disease Control and Prevention
<b>CFU</b>	Colony forming unit
<b>CPDB</b>	Consensus Pathway Database
<b>CRISPR</b>	Clustered Regularly Interspaced Short Palindromic Repeat
<b>DAMP</b>	Danger-associated molecular pattern
<b>DAPI</b>	4',6-diamidino-2-phenylindole
<b>DC</b>	Dendritic cell
<b>DMSO</b>	Dimethyl sulfoxide
<b>DNA</b>	Deoxyribonucleic acid
<b>FACS</b>	Fluorescence-activated cell sorting
<b>FC</b>	Fold change
<b>FCS</b>	Fetal calf serum
<b>FDR</b>	False Discovery Rate
<b>GBP</b>	Guanylate-binding protein
<b>GFP</b>	Green fluorescent protein
<b>GT</b>	Gene trap
<b>GTPase</b>	Guanosine triphosphatase
<b>HEK-293</b>	Human embryonic kidney 293 (cell)
<b>IFN</b>	Interferon
<b>IL (e.g. IL-1<math>\beta</math>)</b>	Interleukin
<b>KO</b>	Knock-Out
<b>LB</b>	Luria Bertani
<b>LC3</b>	Microtubule-associated proteins 1A/1B light chain
<b>lncRNA</b>	Long non-coding RNAs
<b>LPS</b>	Lipopolysaccharide
<b>LVS</b>	Live vaccine strain
<b>MDM</b>	Human monocyte-derived macrophage
<b>MOI</b>	Multiplicity of Infection
<b>mRNA</b>	Messenger RNA
<b>mROS</b>	Mitochondrial ROS
<b>MTT</b>	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
<b>N/A</b>	Not applicable
<b>NADPH</b>	Nicotinamide adenine dinucleotide phosphate
<b>NEB</b>	New England BioLabs
<b>Neo<sup>R</sup></b>	Neomycin resistance
<b>NF-<math>\kappa</math>B</b>	Nuclear factor kappa B
<b>NGS</b>	Next generation sequencing
<b>NHEJ</b>	Non-homologous end joining
<b>NLR</b>	Nucleotide-binding domain-like receptor
<b>OD</b>	Optical density

<b>OD<sub>600</sub></b>	Optical density at 600 nm
<b>PCV</b>	Pathogen-containing vacuole
<b>PAGE</b>	Polyacrylamide gel electrophoresis
<b>PAM</b>	Protospacer adjacent motif
<b>PAMP</b>	Pathogen-associated molecular pattern
<b>Pam<sub>3</sub>CSK<sub>4</sub></b>	Pam3-Cys-Ser-Lys4
<b>PBS</b>	Phosphate-buffered saline
<b>PCD</b>	Programmed cell death
<b>PCR</b>	Polymerase chain reaction
<b>PIPES</b>	1,4-Piperazinediethanesulfonic acid
<b>PIT</b>	Pore-induced intracellular trap
<b>PMA</b>	Phorbol 12-myristate 13-acetate
<b>PFA</b>	Paraformaldehyde
<b>PRR</b>	Pattern recognition receptor
<b>RACE</b>	Rapid amplification of cDNA ends
<b>RIN</b>	RNA integrity number
<b>RNA</b>	Ribonucleic acid
<b>RNAi</b>	RNA interference
<b>RNA-Seq</b>	RNA sequencing
<b>RNS</b>	Reactive nitrogen species
<b>ROS</b>	Reactive oxygen species
<b>SCV</b>	<i>Salmonella</i> containing vacuole
<b>SDS</b>	Sodium dodecyl sulfate
<b>sgRNA</b>	Single guide RNA
<b>siRNA</b>	Short interfering RNAs
<b>SPI</b>	<i>Salmonella</i> pathogenicity island
<b>spp.</b>	Referring to multiple species within a single genus
<b>SSC</b>	Side scatter
<b>STM</b>	<i>Salmonella enterica</i> serovar Typhimurium
<b>STRING</b>	Search Tool for Recurring Instances of Neighbouring Genes
<b>Subsp.</b>	Subspecies
<b>T3SS</b>	Type III Secretion System
<b>T6SS</b>	Type VI Secretion System
<b>TCA</b>	Tricarboxylic acid
<b>V-ATPase</b>	Vacuolar type adenosine triphosphatase
<b>VD3</b>	1,25-dihydroxyvitamin D3
<b>WHO</b>	World Health Organisation
<b>WT</b>	Wild type

# TABLE OF CONTENTS

Declaration.....	ii
Abstract.....	iii
Acknowledgment.....	iv
Abbreviations.....	v
Table of contents.....	vii
Table of Figures.....	xiv
List of Tables.....	xviii
1. Introduction.....	2
1.1. The necessity for research into alternative treatment strategies for infection.....	2
1.1.1. Antimicrobial resistance and the ‘discovery void’.....	2
1.1.2. Threat of bioterrorism.....	3
1.1.3. Alternative approaches to treating infection.....	3
1.2. Pathogen-host evolution.....	4
1.2.1. Innate vs adaptive immune responses.....	4
1.2.2. Classification of infectious bacteria by pathogenic niche.....	4
1.2.3. Survival of intracellular bacteria and host responses.....	5
1.2.4. Innate responses.....	5
1.2.4.1. Recognition of bacteria by plasma membrane and intracellular receptors.....	5
1.2.4.2. Phagocytosis, lysosome-mediated pathogen degradation and reactive oxygen species.....	8
1.2.4.3. Xenophagy.....	8
1.2.4.4. Programmed cell death.....	11
1.2.4.5. Interferon-induced GTPases.....	12
1.3. Rationale for selection of pathogenic bacteria for infection screens.....	12
1.4. Infection lifecycle of bacteria of interest.....	16
1.4.1. <i>Salmonella enterica</i> serotype Typhimurium infection.....	16
1.4.2. <i>Francisella tularensis</i> infection.....	17
1.4.3. <i>Shigella sonnei</i> .....	18
1.5. Phagocytes as important innate host defence cells.....	19
1.5.1. Classical and alternative activation of macrophages.....	20
1.6. Macrophages as a pathogen niche.....	21
1.6.1. Macrophage responses to <i>S. enterica</i> ser. Typhimurium.....	23
1.6.2. Macrophage responses to <i>F. tularensis</i> Schu S4.....	23

1.6.3.	Macrophage responses to <i>S. sonnei</i> .....	24
1.7.	The molecular tools to assess host cell gene function in the context of pathogen infection 25	
1.8.	Creation of mutant cell line libraries for screening purposes – potential tools .....	25
1.9.	Other infection-based genetic screens .....	28
1.10.	Previous work carried out as part of a Centre for Defence Enterprise (CDE) funded project identified a number of candidate host resistance genes .....	29
1.10.1.	Gene trap identified genes .....	32
1.10.1.1.	<i>CFTR</i> .....	33
1.10.1.2.	<i>ARID4B</i> .....	34
1.10.1.3.	<i>PCDH7</i> .....	34
1.10.1.4.	<i>PRKG1</i> .....	35
1.10.1.5.	<i>PSMA1</i> .....	35
1.10.1.6.	<i>SENP5</i> .....	35
1.11.	Aims and hypotheses .....	37
2.	Materials and Methods.....	40
2.1.	Conditions for cell culture maintenance and archiving .....	40
2.2.	Generation of CRISPR cell line and Gene trap library mutants.....	40
2.2.1.	Design or ordering of CRISPR constructs in order to knock out candidate genes .....	40
2.2.2.	Transfection .....	42
2.2.3.	Gene trap library construction.....	45
2.3.	Genomic DNA extraction from mutated cells.....	45
2.4.	Characterisation of CRISPR gene mutations using PCR, cloning and sequence analysis .....	45
2.4.1.	Western blot .....	48
2.5.	Final established differentiation protocol for U937 cells .....	48
2.6.	Flow cytometry to establish composition of differentiated cells .....	49
2.7.	Bacterial culture conditions .....	50
2.8.	Bacterial infection assay .....	50
2.9.	Lactate dehydrogenase (LDH) activity assay to determine the extent of cell death after infection .....	52
2.10.	Immunofluorescence .....	53
2.11.	Screening protocol for differentiated U937 GT library.....	53
2.11.1.	STM U937 screen .....	53
2.11.2.	<i>F. tularensis</i> U937 screen .....	54
2.12.	Molecular processing of RNA from gene trap library samples .....	54
2.13.	Bioinformatics analysis of ‘RNA-seq’ data from gene trap screens.....	56

2.14.	Independent validation of a gene hit (SLC7A11) from the STM U937 infection screen, through use of a CRISPR KO pool and pharmacological modulation of the protein .....	59
2.14.1.	Counting intracellular bacteria .....	59
2.15.	Statistical analysis .....	62
3.	Creating and characterising cell line mutants to confirm host candidate gene roles in the infection process.....	64
3.1.	Genetic validation of isolated CRISPR KO cell lines.....	64
3.1.1.	Confirming CRISPR mutations in <i>SENP5</i> KO cell lines .....	65
3.1.2.	Confirming CRISPR mutations in <i>PRKG1</i> KO cell lines.....	71
3.1.3.	Confirming CRISPR mutations in <i>PCDH7</i> KO cell lines.....	75
3.2.	Failure to observe significant effects on protein expression in CRISPR <i>ARID4B</i> HEK-293 cell lines	80
3.3.	Lack of altered <i>SENP5</i> protein expression in CRISPR Cas9 <i>SENP5</i> targeted cell line .....	85
3.4.	Phenotypic characterisation of <i>PCDH7</i> and <i>PRKG1</i> CRISPR KO cell line infection by <i>Salmonella</i> and <i>Shigella</i> .....	90
3.4.1.	Determining accurate multiplicity of infection for <i>S. enterica</i> ser. Typhimurium: correlation of optical density measurements with colony forming units .....	92
3.4.2.	Survival of CRISPR <i>PCDH7</i> and <i>PRKG1</i> KO cell lines during <i>S. enterica</i> ser. Typhimurium infection as a means to validate the survival phenotype originally identified in the HEK-293 gene trap survival screen.....	94
3.4.3.	Assessment of the intracellular bacterial burden of CRISPR KO cell lines during infection with intracellular pathogens.....	97
3.4.3.1.	<i>S. enterica</i> ser. Typhimurium infections .....	97
3.4.3.2.	<i>Shigella sonnei</i> infections.....	101
3.5.	Discussion.....	105
3.5.1.	Overview .....	105
3.5.2.	Summary of the genetic characterisation of CRISPR KO cell lines.....	105
3.5.3.	Are CRISPR KO HEK-293 cell lines heterogeneous? .....	105
3.5.4.	Were the ‘mutants’ functionally mutated? .....	106
3.5.5.	Summary of the phenotypic characterisation of CRISPR KO cell lines.....	107
3.5.6.	Validated gene <i>PCDH7</i> .....	109
3.5.7.	Initial invasion protection may affect burden later .....	109
3.5.8.	Cross-species similarities and differences .....	109
3.5.9.	Critique of CFU to OD.....	110
3.5.10.	Critique of the LDH assay .....	111
3.5.11.	Critique of the bacterial burden assay.....	111
3.5.12.	Problems with the CRISPR sham ( <i>SEMA3A</i> KO) control.....	112
3.5.13.	Future work.....	113

3.5.13.1.	Creation of a better CRISPR control .....	113
3.5.13.2.	Clarification and further empirical analysis of infection phenotypes.....	113
3.5.13.3.	Generation of new CRISPR KO <i>SENP5</i> and <i>ARID4B</i> cell lines for validation of infection phenotypes .....	114
3.5.14.	Conclusion and rationale for work described in the next chapter .....	114
4.	Construction, differentiation, and screening of a gene trap library in U937 macrophage-like cells	118
4.1.	Creation of a differentiation protocol for the U937 cell line to generate macrophage-like adherent cells.....	119
4.1.1.	Differentiation Strategy #3 .....	119
4.1.2.	Can macrophage-like cells be “retrodifferentiated”?.....	123
4.2.	Characterisation of the differentiated U937 macrophage-like cell phenotype.....	126
4.2.1.	Cell surface markers.....	126
4.2.2.	Bacterial phagocytosis .....	128
4.2.2.1.	Live imaging of differentiated U937 cells phagocytosing <i>Escherichia coli</i> DH5α	132
4.3.	Generation of a U937 GT library .....	133
4.4.	Determining suitable infection conditions for the differentiated U937 gene trap library as preparation for an infection-host cell survival screen .....	134
4.5.	Use of the U937 gene trap library in a host cell survival screen during <i>S. enterica</i> ser. Typhimurium infection .....	137
4.6.	Use of the same U937 gene trap library in a host cell survival screen using <i>F. tularensis</i> subsp. <i>holarctica</i> Live Vaccine Strain infection.....	140
4.7.	Discussion.....	143
4.7.1.	An optimal infection MOI and time frame for stringent selection .....	143
4.7.1.1.	The selection procedure was not too stringent - the total RNA was still of good quality	144
4.7.2.	Limitations of U937 differentiation characterisation .....	145
4.7.2.1.	Exposure of cocktail differentiated U937 cells to bacteria .....	146
4.7.3.	Retrodifferentiation is not possible in drug cocktail differentiated U937 cells.....	146
4.7.4.	Future work.....	146
4.7.4.1.	<i>S. sonnei</i> strain 86 as candidate pathogen for a future infection screen .....	147
4.7.4.2.	Iterative rounds of selection as an additional approach in a monocyte-like model	147
5.	Bioinformatics analysis and phenotypic validation of U937 GT library infection screens.....	149
5.1.	Summary of pre-sequencing and sequencing processes.....	149
5.2.	Bioinformatics analysis of next generation sequencing (NGS) results from the <i>S. enterica</i> ser. Typhimurium infection-host cell survival screen with the U937 gene trap library .....	152
5.2.1.	Raw DNA sequence quality assessment .....	153

5.2.1.1.	Base call quality and read length .....	153
5.2.1.2.	Quality check for species and vector contamination.....	155
5.2.1.3.	GC content quality check .....	156
5.2.1.4.	Per sample sequence overrepresentation .....	157
5.2.2.	Empirical comparison of quality-based processing techniques for use prior to analysis	158
5.2.2.1.	Trimming to an average Phred of 20 is appropriate .....	158
5.2.2.2.	GT vector removal from sequence reads leads to loss of mappable sequences	161
5.2.3.	Statistics of differentially represented human genes during STM infection and identification of the most significantly differentially represented genes .....	161
5.2.3.1.	Selection stringency #1: FDR threshold of 0.1 .....	161
5.2.3.2.	Selection stringency #2: A reduced FDR threshold of 0.3.....	165
5.2.4.	Gene ontology analysis of STM screen-derived genes .....	170
5.2.4.1.	GO Process - Electron transport chain .....	172
5.2.4.2.	GO Process - Response to oxidative stress .....	173
5.2.5.	Intergenic trapped sequences with altered representation after infection were identified using <i>de novo</i> transcript assembly .....	174
5.3.	Bioinformatics analysis of the <i>F. tularensis</i> subsp. <i>holarctica</i> live vaccine strain (LVS) infection-host cell survival screen using the U937 gene trap library .....	178
5.3.1.	Raw DNA sequence quality assessment .....	178
5.3.2.	Quality processing.....	181
5.3.3.	Comparison of GT library representation between different passages of the GT library – greater passage number expected to affect representation.....	181
5.3.4.	Identified Genes of Interest .....	182
5.3.4.1.	Differentially represented genes with gene products involved in protein trafficking	187
5.3.4.2.	Differentially represented genes with gene products involved in immune response regulation .....	187
5.3.5.	Gene ontology analysis of the <i>F. tularensis</i> U937 screen .....	188
5.3.6.	Protein interaction networks from combined query gene list .....	188
5.3.6.1.	Consensus Pathway Database .....	188
5.3.6.2.	Search Tool for Recurring Instances of Neighbouring Genes (STRING) database	191
5.4.	Assessment of phenotype in an independent mutant model .....	195
5.4.1.	Creation of CRISPR constructs and transfection to create U937 CRISPR mutant pools	195
5.4.2.	Effects of genetic ablation or pharmacological inhibition of SLC7A11 on STM burden	198

5.4.2.1.	Microtubule changes .....	201
5.5.	Discussion.....	203
5.5.1.	Summary/Overview .....	203
5.5.2.	Involvement of the Electron Transport Chain.....	204
5.5.3.	Novel gene associations with STM and or <i>F. tularensis</i> infection.....	205
5.5.4.	Lack of overlap with published <i>Salmonella</i> infection screens? .....	206
5.5.5.	Comparison to other infection or immune response-based genetic screens: .....	206
5.5.6.	Why does the screen have a comparatively low gene coverage? .....	208
5.5.7.	Technical critique of U937 screen.....	208
5.5.7.1.	Trimming .....	208
5.5.7.2.	Genome Alignment .....	209
5.5.7.3.	<i>De novo</i> vs reference-based assembly vs an alternative .....	209
5.5.7.4.	<i>F. tularensis</i> U937 screen: low statistical power.....	209
5.5.7.5.	Repeated or independent measures designs for statistical analysis? .....	210
5.5.8.	Future work.....	210
5.5.8.1.	Experimental .....	210
5.5.8.2.	Computational .....	211
5.5.9.	Conclusion.....	211
6.	Discussion.....	213
6.1.	Summary .....	213
6.2.	Assessing the success of the functional screening approach through the biological actions of the identified host processes.....	214
6.3.	A universal infection drug target – does it exist? .....	217
6.3.1.	A gene-centric approach.....	217
6.3.2.	A pathway-centric approach.....	217
6.3.3.	Exploration of one such pathway; different survival strategies to intraphagosomal ROS 218	
6.3.4.	Differences in pathogen responses to differing intracellular environments within professional and non-professional phagocytic cells .....	219
6.3.5.	Stochastic variation in the screening process – detecting the signal through the noise 219	
6.3.6.	Lack of shared host infection factors between HEK-293 and U937 cells – what technical limitations were at play? .....	219
6.3.7.	What technical limitations prevented identification of host infection factors common to STM and <i>F. tularensis</i> ?.....	220
6.4.	How could discovery screens be optimised? .....	221

6.4.1.	Identification of infection resistance or susceptibility genes through host cell survival selection requires context – host cell survival is not conclusive of simple infection resistance	221
6.4.2.	Host cell survival or Bacterial burden – methods of selection of infection resistance	222
6.4.3.	Mutant library generation – GT or CRISPR?.....	223
6.4.4.	What is the optimal route to identify GTs of interest in a screen – clonal isolation of cells, or NGS of pooled populations?.....	224
6.4.5.	A solution for iterative rounds of infection selection .....	225
6.5.	Future work.....	225
6.5.1.	Further development of a pathway-centric approach through screening with a broader range of pathogens .....	225
6.5.2.	Taking the most promising host infection factors forward .....	225
6.5.3.	Potential druggability of identified proteins and pathways .....	226
6.6.	Concluding remarks .....	227
	References .....	228
Appendix A	Investigating the suitability of differentiation strategies of U937 cells for screening	290
A.1.	Differentiation Strategy #1 .....	290
A.2.	Differentiation Strategy #2 .....	295
Appendix B	Quality assessment of total RNA samples derived from STM and <i>F. tularensis</i> U937 screens	304
B.1.	RNA extracted from STM U937 screen samples: Quality assessment and processing for next generation sequencing (NGS) .....	304
B.2.	RNA extracted from <i>F. tularensis</i> LVS U937 screen samples: Quality assessment and processing for next generation sequencing (NGS) .....	309
Appendix C	Investigation of suitable GT NGS pre-alignment procedures .....	311
C.1.	Cutting.....	311
C.2.	Simply trimming retains more mappable reads than (Cut x3-Trim) x2 .....	316
C.3.	StringTie successfully incorporates single-mapped reads and paired end reads during transcript assembly and enumeration.....	319

## TABLE OF FIGURES

Figure 1.1 Summary of canonical autophagy, xenophagy and non-canonical (LC3 associated phagocytosis) xenophagy.....	11
Figure 1.2 Three diverse intracellular bacterial pathogens ( <i>S. enterica</i> , <i>F. tularensis</i> , <i>S. sonnei</i> ) share the macrophage as a replicative niche – thus the macrophage is used as an <i>in vitro</i> model to screen for shared host infection factors. ....	22
Figure 1.3 CRISPR- Cas9 causes double-stranded breaks in target DNA using a complementary guide RNA, which can be repaired by host non-homologous end joining (NHEJ), thus generating mutations. ....	27
Figure 1.4 Polyadenylation-enhancer trap mutagenesis. ....	28
Figure 1.5 Postulated host response network based on positive hits from screens #1 and #2 and relationships identified in the literature. ....	32
Figure 2.1 pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid with sgRNA insert. ....	42
Figure 2.2 Generation and genetic validation of CRISPR cell line knock outs.....	44
Figure 2.3 Schematic of the PCR programme used to amplify CRISPR target loci. ....	46
Figure 2.4 Process to identify the genetic mutation and predicted consequences to the amino acid sequence of a CRISPR KO cell line.....	47
Figure 2.5 Final differentiation protocol for U937 cells. ....	49
Figure 2.6 Gating strategy for cytometric analysis of differentiated and unstimulated U937 cells. ....	49
Figure 2.7 Equation to calculate MOI of <i>Shigella</i> sp. 2-hour subcultures. ....	51
Figure 2.8 Equation to calculate MOI of <i>Salmonella</i> sp. 2-hour subcultures.....	51
Figure 2.9 Bacterial infection assay. ....	52
Figure 2.10 PCR programme used for RACE1 and RACE2. ....	55
Figure 2.11 Overview of bioinformatics pipeline in Galaxy. Galaxy workflow (automated pipeline of tools where the output of one is fed into another) for LVS U937 GT screen used as an example. ....	58
Figure 3.1 PCR products of some mutated <i>SENP5</i> lines are visibly different in size compared to wild type PCR product.....	66
Figure 3.2 HEK-293 CRISPR <i>SENP5</i> Line 3 contains multiple deletions and insertions at the CRISPR target site. ....	68
Figure 3.3 HEK-293 <i>SENP5</i> Line 3 contains multiple copies of the <i>SENP5</i> gene with mutations predicted to create non-functional protein. ....	70
Figure 3.4 HEK-293 CRISPR <i>PRKG1</i> Line 12 contains multiple deletions at the CRISPR target site. ...	72
Figure 3.5 HEK-293 <i>PRKG1</i> Line 12 contains multiple non-functional copies of the <i>PRKG1</i> protein. ....	74
Figure 3.6 HEK-293 CRISPR <i>PCDH7</i> Line 2 contains multiple deletions at the CRISPR target site. ....	77
Figure 3.7 HEK-293 <i>PCDH7</i> Line 2 contains multiple non-functional copies of protocadherin 7. ....	80
Figure 3.8 HEK-293 CRISPR KO <i>ARID4B</i> Lines K and 6 have comparable <i>ARID4B</i> protein expression to CRISPR control (cont) <i>SEMA3A</i> and HEK-293 wild type (WT).....	82
Figure 3.9 <i>ARID4B</i> Lines 6 and K show different patterns of protein expression compared to wild type cells, which may indicate the presence of mutations. ....	85
Figure 3.10 CRISPR KO <i>SENP5</i> Line 3 has similar <i>SENP5</i> protein expression to CRISPR control <i>SEMA3A</i> regardless of STM-infected status. ....	86
Figure 3.11 Comparison of the protocols that were used to isolate, and subsequently validate, infection-resistant genes in HEK-293 cells.....	92

Figure 3.12 STM growth in subculture, as measured by OD <sub>600</sub> over time: an OD <sub>600</sub> of 0.9 is still in exponential phase. ....	93
Figure 3.13 OD: CFU correlations for STM and <i>S. sonnei</i> . ....	94
Figure 3.14 Equation to determine percent cytotoxicity provided by the Promega Cytotox96 Non-radioactive cytotoxicity assay manual. ....	95
Figure 3.15 Infected HEK-293 CRISPR KO cell lines show increased cytotoxicity at 48 hours post gentamicin maintenance. ....	96
Figure 3.16 Infection experiment 1 - HEK-293 CRISPR KO <i>PRKG1</i> Line 12 has significantly reduced intracellular STM burden at 24 hours; a difference that disappears when normalised to 0-hour burdens. ....	99
Figure 3.17 Infection experiment 2 - HEK-293 CRISPR KO <i>PCDH7</i> Line 2 has significantly reduced intracellular STM burden at 24 hours; a difference which is retained when normalised to 0-hour burdens. ....	100
Figure 3.18 Infection experiment 1 - HEK-293 CRISPR KO <i>PRKG1</i> Line 12 and <i>PCDH7</i> Line 2 have significantly lower intracellular <i>S. sonnei</i> burdens than wild type (WT) at 24 hours; a difference that is inverted for HEK-293 CRISPR KO <i>PRKG1</i> Line 12 when normalised to 0-hour burdens. ....	102
Figure 3.19 Infection experiment 2 - HEK-293 CRISPR KO <i>PCDH7</i> Line 2 has significantly reduced intracellular <i>S. sonnei</i> burden at 24 hours compared to wild type (WT); a difference which is retained when normalised to 0-hour burdens. ....	104
Figure 4.1 Differentiation strategies including ATRA in combination with PMA with or without VD3 result in what appears to be apoptotic cell death. ....	122
Figure 4.2 Combined PMA and VD3 treatment enhances differentiation as assessed by substantial morphological changes. ....	123
Figure 4.3 FCS concentration increases U937 cell detachment after PMA treatment, though this was limited at desired PMA concentrations. ....	125
Figure 4.4 The differentiation drug-cocktail treatment causes increased CD11b surface expression of U937 cells. ....	127
Figure 4.5 <i>Escherichia coli</i> DH5α bacteria colocalise with heterogeneous U937 differentiation drug-cocktail treated cells. ....	130
Figure 4.6 Z-Stack: Non-invasive <i>E. coli</i> are actively internalised by drug-cocktail differentiated U937 cells. ....	132
Figure 4.7 Differentiated U937 cells move, engulf bacteria, and eventually die in response to bacterial inoculation. ....	133
Figure 4.8 STM infection cytotoxic effects on differentiated U937 cells assessed by LDH release: Cytotoxic effect increases with increasing MOI and time. ....	137
Figure 4.9 Cytotox96 reagent breaks down over multiple freeze-thaws, reducing the OD activity range. ....	137
Figure 4.10 Final GT screen RACE PCR material prior to DNA library preparation and sequencing. ....	139
Figure 4.11 Final GT screen RACE PCR material prior to DNA library preparation and sequencing. ....	142
Figure 5.1 Summary of RACE PCR amplicon composition and technical processes required for Next Generation Sequencing. ....	152
Figure 5.2 STM U937 screen raw NGS files have an average of 30 million paired reads, as requested during procurement. ....	153
Figure 5.3 The read length is as requested and the majority of reads in all samples have good quality (Phred) scores averaged across the read. ....	154

Figure 5.4 At least 40% of sequence reads across all infected (experimental), uninfected (control), and original U937 GT library samples match the human genome. ....	156
Figure 5.5 The average per sequence GC content for the raw STM screen sequence samples is around 40% as expected, but the bell curve is interrupted by numerous sharp peaks. ....	157
Figure 5.6 At least 12.5% of reads per raw sample contain overrepresented sequences. ....	158
Figure 5.7 Q20 trimmed and size-filtered sequences from STM U937 screen samples retain a greater majority (30 million surviving read pairs) of the input sequences than Q30 for genome alignment. ....	160
Figure 5.8 Successful identification of trapped genes that show a strong effect on U937 cell survival or death (FDR threshold of 0.1) as measured by read depth differences between Infected and Control samples across the six paired replicates. ....	162
Figure 5.9 Sequence analysis resolved 84 mutant genes differentially represented in the STM infected U937 GT population compared to the control at a more liberal FDR threshold (0.3). ....	166
Figure 5.10 Flow chart of processes used to identify over- or underrepresented trapped intergenic sequences and additional overrepresented genes in control vs infected U937 GT cells through <i>de novo</i> transcript assembly. ....	175
Figure 5.11 Raw NGS read quality for U937 GT library <i>F. tularensis</i> LVS screen is comparable to NGS read quality from U937 GT library STM screen. ....	180
Figure 5.12 Q20 trimmed and filtered LVS U937 screen sequencing samples retain more than 20 million surviving read pairs for genome alignment. ....	181
Figure 5.13 <i>F. tularensis</i> differentiated U937 GT library screen resolves 1,867 mutant gene sequences differentially present in infected U937 population. ....	183
Figure 5.14 Differentially represented gene sequences in <i>F. tularensis</i> LVS infected U937 samples reduced to 39 when further constrained. ....	184
Figure 5.15 Protein-protein interactions between both screen gene products converge on Amyloid-beta precursor protein (APP). ....	190
Figure 5.16 Products of genes identified in STM and LVS differentiated U937 GT screens may have functional associations. ....	194
Figure 5.17 U937 CRISPR KO <i>SLC7A11</i> pool #1 has much less SLC7A11 protein expression than wild type or Miscullin control. ....	196
Figure 5.18 U937 CRISPR KO <i>SLC7A11</i> #1 pool has reduced SLC7A11 staining. ....	198
Figure 5.19 Flattened cocktail differentiated CRISPR KO <i>SLC7A11</i> pool #1 U937 cells and wild type cells treated with 10 ng/mL sulfasalazine tend to have fewer intracellular STM than wild type cells. ....	201
Figure 5.20 STM infection causes microtubule remodelling of differentiated U937 cells at 2-hours post inoculation. ....	202
Figure 5.21 Example microscopy images to indicate what was seen during data collection at 24 hours post-inoculation for Figure 5.19. ....	203
Figure 6.1 Most promising potential host infection factors identified through genetic macrophage screens utilising two evolutionarily distant bacterial pathogens. ....	216
Figure 6.2 Increased host cell survival may be the result of a resistance to cell death or a resistance to bacterial entry/growth. ....	222
Figure A.1 Differentiation strategy #1. ....	291
Figure A.2 PMA treatment of U937 cells promotes a variety of morphological changes, creating a heterogeneous population of attached cells. ....	294
Figure A.3 Timeline for differentiation strategy #2- comparing effects of VD3, PMA and ATRA; secondly for retrodifferentiation of PMA treated cells. ....	298
Figure A.4 U937 cells continuously treated with PMA cluster together but do not change shape. ....	301

<b>Figure A.5 PMA-treated U937 cells undergo diverse morphological changes after a subsequent 'rest' period of culturing without PMA. ....</b>	<b>303</b>
<b>Figure B.1 Selection of infection-Control pair replicates that produced the highest quality RNA. ....</b>	<b>305</b>
<b>Figure B.2 The RNA quality of OR1 and OR2 is good. ....</b>	<b>307</b>
<b>Figure B.3 U937- F. tularensis LVS GT library screen RNA samples are good quality. ....</b>	<b>310</b>
<b>Figure C.1 Cases in which Cutadapt will remove or not remove a matching region of a read.....</b>	<b>312</b>
<b>Figure C.2 Both per base sequence content and GC count distribution approach expected values after iterative cutting and trimming.....</b>	<b>316</b>
<b>Figure C.3 The percentage of mapped reads was higher in the (Cut x3-Trim) x2, filtered sample, but the total number of reads was lower, hence trimming and filtering seems to be the best choice.</b>	<b>318</b>
<b>Figure C.4 StringTie makes use of single-mapped reads to produce gene counts.....</b>	<b>320</b>

## LIST OF TABLES

Table 1.1 List of pertinent PRRs that recognise bacterial PAMPs. ....	6
Table 1.2: Characteristics of pathogens used in gene trap screens and phenotyping of CRISPR mutant cells described in Table 1.4. ....	14
Table 1.3 Symptoms of, and human populations susceptible to, diseases caused by pathogens of interest. ....	15
Table 1.4 Summary of infection screens and infections in CRISPR mutants done as part of the CDE project. ....	30
Table 1.5 Genes of Interest identified from HEK-293 gene trap library intracellular infection screen and now the focus of this PhD. ....	33
Table 2.1. Bacterial strains and plasmids used. ....	50
Table 2.2 Antibody combinations used for immunofluorescence. <i>All secondary antibodies were the Invitrogen brand from Thermo Fisher.</i> ....	53
Table 2.3 Size distribution of DNA fragment libraries created from each experimental sample. ....	56
Table 2.4 Primers and CRISPR oligo sequences. ....	59
Table 3.1 List of CRISPR cell lines created, isolated and genetically characterised for the genes of interest from the HEK-293 screen. ....	87
Table 3.2 Summary of infection results using isolated and genetically characterised CRISPR KO cell lines. ....	108
Table 3.3 Side by side comparison of predicted STM numbers from an initial OD reading is largely the same for my equation and that produced by Gonzalez et al. ....	110
Table 4.1 Attempts to retrodifferentiate macrophage-like cells after PMA treatment: differentiation drug and FCS concentrations influence the number of cells that return to suspension. ....	124
Table 4.2 Description of RACE PCR amplicon samples from the STM U937 screen presented in Figure 4.10. ....	140
Table 4.3 Summary details of <i>F. tularensis</i> LVS U937 GT infection. ....	141
Table 4.4 Description of RACE PCR amplicon samples from the <i>F. tularensis</i> U937 screen presented in Figure 4.11. ....	143
Table 4.5 Summary of STM infection of cocktail differentiated wild type U937 cells to find optimal infection conditions. ....	145
Table 5.1 Molecular function of differentially represented genes at a FDR threshold of 0.1. ....	163
Table 5.2 Molecular function of additional differentially represented genes at a FDR threshold of 0.3 in STM infected U937 GT cells. ....	166
Table 5.3 Gene Ontology enrichment analysis of a lower stringency selection of 35 overrepresented mutant genes identified enriched GO terms including the GO process ‘response to oxidative stress’. ....	172
Table 5.4 Intergenic mutant transcripts with altered representation in STM-infected samples. ....	176
Table 5.5 Molecular function of differentially represented gene sequences in <i>F. tularensis</i> LVS infected U937 samples constrained by post-limma filters. ....	185
Table 5.6 Comparison of STM and <i>F. tularensis</i> U937 GT screens with other infection or immune response based genetic screens. ....	207
Table 5.7 Alternative or additional programmes suggested as options to improve upon the current analyses. ....	210
Table A.1 PMA treatment promotes adherence and changes in morphology in U937 cells. ....	292

<b>Table B.1 Quality of total RNA extracted from chosen samples is sufficient for RNA-Seq. ....</b>	<b>306</b>
<b>Table B.2 DNA concentration and quality of purified GT screen RACE PCR material prior to DNA library preparation and sequencing. ....</b>	<b>308</b>

# Chapter 1

## Introduction

# 1. INTRODUCTION

## 1.1. The necessity for research into alternative treatment strategies for infection

### 1.1.1. Antimicrobial resistance and the ‘discovery void’

The World Health Organization produced a 2014 report into the global burden of (acquired) antimicrobial resistance, focusing on the antimicrobial resistance of eight bacterial genera (WHO, 2014). Although the report was necessarily incomplete (due to lack of available data from some developing countries), antibiotic resistance was found to be widespread in all bacterial species surveyed. Acquired antimicrobial resistance is the development of resistance in a pathogen that was once sensitive to the antimicrobial compound, whether by mutation or from an external source (e.g. from a plasmid with an antimicrobial resistance gene). More recently, plasmid-mediated antibiotic resistance to polymyxins, a class containing colistin, has been found in bacterial isolates from humans and animals (Liu et al., 2016). This form of antibiotic resistance can spread quickly due to the episomal location of the resistance gene. The spread of colistin resistance increases the potential inability to treat infections caused by multi-drug resistant bacteria as the antibiotic is normally reserved for these (WHO, 2019). The increased level of acquired antimicrobial resistance is the result of anthropogenic antibiotic activity – heavy duty antibiotic agricultural use (Patel et al., 2020), antibiotic pollution of the environment (Kumar et al., 2020; Nnadozie and Odume, 2019; Rodriguez-Mozaz et al., 2015; Sun et al., 2020) and high rates of global medical prescribing (Klein et al., 2018). These activities increase the selection pressure on bacteria develop antibiotic resistance, thus increasing the rate at which resistance is acquired and spread.

An estimated 700,000 people die from drug-resistant infections annually (although this includes HIV and malarial as well as bacterial infections) (O’Neil, 2016). Indeed, the Chief Medical Officer for England, Professor Dame Sally Davies has described routine medical operations as “risky” without effective antibiotics against antimicrobial-resistant pathogens (The BBC, 2017). To encourage research into, and development of, novel antibacterial treatments, the WHO published a list of antibiotic resistant priority bacteria; seen as the biggest threats to public health (Tacconelli and Magrini, 2017; Willyard, 2017). The most critical of these are *Acinetobacter baumannii* and *Pseudomonas aeruginosa* (both of which are carbapenem resistant). Lower priority pathogens include fluoroquinolone resistant *Salmonellae* and *Shigella*.

No new classes of antibiotics targeting gram-negative bacteria have been licenced in more than forty years (Brown, 2015) and the WHO has named the period following the discovery of Daptomycin (gram positive-targeting antibiotic) as a “discovery void” (WHO, 2014). During the “golden age of antibiotic discovery” pharmaceutical companies identified soil bacteria as sources of plentiful efficacious antimicrobial compounds; with most existing antibiotics based on these. At the turn of the century, research efforts turned to targeted approaches, using high-throughput assays to identify compounds effective against purportedly essential targets within bacteria. This strategy produced few successful results; with difficulties finding compounds with suitable characteristics for the treatment of infections, and the ever-present risk of further microbial resistance evolving due to the mutation of the targeted protein (Payne et al., 2007; Tommasi et al., 2015). Antibiotic discovery research is now mainly focused on identifying natural products produced by micro-organisms from

less studied parts of the environment – this has been reviewed elsewhere (Hutchings et al., 2019; Lewis, 2020). A big bottleneck for this type of discovery is the unknown quantity of micro-organisms with novel potential antibacterial agents in the sampling space. Additionally, most large pharmaceutical companies have left the antibiotic discovery space (Hutchings et al., 2019), making it more difficult to scale up investigations and throughput.

### 1.1.2. Threat of bioterrorism

Another cause for concern is the potential for the intentional use of biological agents as weapons. In recent years, there have been numerous terrorist attacks across Europe and the USA, including the 2017 vehicular attacks in Spain and England, the 2017 Manchester suicide bombing as well as the 2016 suicide bombings in Paris. Biological attacks have the potential to be as destructive/devastating as conventional weaponry (Riedel, 2004) and recent biotechnological advances (such as Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9) has reduced the technological barrier to purposefully weaponising non-genetically modified pathogens (MacIntyre, 2020). The Centers for Disease Control and Prevention (CDC) published a report on strategic planning for biological and chemical terrorism (Centers for Disease Control and Prevention, 2000) after the USA became aware of bioweapons programmes under development in other countries. There are few undisputed incidents of bioterrorism (Riedel, 2004), but those that occur, such as the anthrax attack in 2001 (WHO, 2004) are reminders of the ease of implementation, the potential for great devastation, and the ability for covertness. Well-planned biological attacks could prove a heavy burden for countries where the infrastructure is already under strain, such as those affected by natural disasters. Therefore, new treatment approaches are needed to combat both potential bioweapons and the existing natural epidemic infection risks to the general population. Examples of bacteria of particular concern include; *Bacillus anthracis*, *Yersinia pestis* and *Francisella tularensis*; due to their infectivity, potential to generate lethal infection, potential for social disruption and necessity for specific preparation (Centers for Disease Control and Prevention, 2018; MacIntyre et al., 2006; Pappas et al., 2009).

### 1.1.3. Alternative approaches to treating infection

Due to the 'discovery void' and increasing (and wide-ranging) prevalence of antibiotic resistance, alternative methods of treating bacterial infection are being sought, including use of bacteriophages, predatory bacteria, antimicrobial peptides (Allen et al., 2014), metals (H.-Z. Li et al., 2017) and host response-targeting (Kaufmann et al., 2017). Previously, little attention had been paid to host targets as anti-infectives; undoubtedly because the complexity of the human genome and the human protein interactome made this an unrealistic strategy. Now in the post-genomic era, host-directed therapies are a realistic prospect. Enhancement of host immune responses and obstruction of host proteins required for intracellular bacterial survival are just some of the therapeutic options explored to reduce infection generated morbidity (Kaufmann et al., 2017). There are advantages to targeting host proteins, such as the potential to find broad-specificity treatments and to reduce the likelihood of therapy resistance (Kaufmann et al., 2017). These are issues that modern single-target high-throughput assays fail to address. Naturally, there are also drawbacks to this approach, including the potential to interfere with vital human cellular processes.

Many bacterial infections do not produce unique constellations of symptoms (e.g. non-typhoidal-salmonellosis and shigellosis present similar symptoms to other gastric infections). Therefore, many of these are difficult to treat with targeted antibiotics in a timely fashion because the species and in some cases, the strain, must be identified first. Hence, alternative broad-spectrum treatments would

be extremely useful; particularly in instances of epidemics, for the treatment of critically ill individuals, or treatment of people deliberately exposed to an unknown pathogen.

## 1.2. Pathogen-host evolution

### 1.2.1. Innate vs adaptive immune responses

To identify potential host infection factors that may be pharmacologically modulated, one must determine which features of the host infection response are appropriate to investigate. The human immune system employs different mechanisms to destroy and clear intracellular and extracellular pathogens depending on their physical location. Both **adaptive** (antigen-specific, not genetically encoded) and **innate** (non-specific, genetically encoded) responses are employed in responding to pathogens in the extracellular and intracellular space. The adaptive immune system has very powerful mechanisms of identifying and targeting pathogens but requires time to develop. The adaptive immune response is also highly specific – distinguishing between species and often strain of bacteria. Therefore, it is not an appropriate target for the generation of a broad spectrum (or universal) host-directed therapy. In contrast, the innate immune system is able to broadly and rapidly identify pathogens, before setting in motion a range of effector mechanisms that can kill or restrict pathogen growth but are self-deactivating. Therefore, the innate response is an appropriate aspect of the focus for this project.

Human innate immune responses encompass extracellular components (such as complement) as well as a cellular aspect, which can be subdivided into cell-autonomous and non-autonomous functions. The non-autonomous characteristics refer to any intercellular signals and functions resulting from these, such as signals required for cellular activation and recruitment. The autonomous characteristics refer to the intracellular capabilities of a host cell, as well as autocrine signalling, but without the requirement for other host cells. Investigating these aspects of the innate immune response requires different experimental contexts (e.g. the presence of serum, tissue explants, multiple host cell co-culture, or single cell type culture). To determine the appropriate host context required to identify potential host infection factors of interest, it is helpful to define the natural accessibility of the bacteria to the immune system as well as the predominant aspects of its infection life cycle.

### 1.2.2. Classification of infectious bacteria by pathogenic niche

Pathogenic bacteria are classically classified by their ability to multiply in a cell-free environment or intracellularly inside a susceptible host: extracellular (e.g. *Staphylococcus aureus*, *Streptococcus pyogenes*, *Pseudomonas aeruginosa*), facultative intracellular (e.g. *Salmonella enterica*, *Francisella tularensis*, *Shigella sonnei*) and obligate intracellular bacteria (e.g. *Coxiella burnetii*, *Anaplasma phagocytophilum*, *Rickettsia prowazekii*) (Fields et al., 2011). This classification, in light of new experimental data has become controversial, as bacteria generally thought of as extracellular appear to have some intracellular phases *in vivo* see Silva (2012) for review. Currently there are no other standardised criteria available to classify pathogens, but the current criteria serve our purposes sufficiently as the majority of the productive infection by putatively extracellular bacteria are in the extracellular space. In contrast, obligate intracellular bacteria can only survive within host cells, thus escaping from much of the host immune response. Facultative intracellular bacteria are those that have evolved to generate productive infections inside host cells but are also able to multiply extracellularly. In addition to evolving mechanisms to survive in both environments, these bacteria must also gain access to the host cell, either through passive or active mechanisms.

Intracellular bacteria are some of the most virulent, indeed several of the bacteria identified as potential bioterror threats on the CDC website are intracellular (Balali-Mood et al., 2013; Centers for Disease Control and Prevention, 2018). Furthermore, many virulent intracellular bacteria (e.g. *Yersinia pestis*, *Bacillus anthracis*, *Brucella* species, *Burkholderia pseudomallei*, *Mycobacterium tuberculosis* and *F. tularensis*) cause diseases that lack either vaccines, effective treatments, or both (Doganay and Demiraslan, 2015; Figueiredo et al., 2015; Jia and Horwitz, 2018; Morici et al., 2019; Pai et al., 2016; Pechous et al., 2016).

### 1.2.3. Survival of intracellular bacteria and host responses

When inside host cells, intracellular pathogens need to create an environment that will facilitate survival, multiplication and ultimately transmission. Three key ways in which they survive in this hostile niche are by evading, or subverting, the host immune response, as well as by obtaining nutrients directly from the host cells. Consequently, intracellular pathogens have evolved mechanisms to import important nutrients and release molecules that alter host cell signalling, immune mechanisms, and metabolism. Hosts cells, in turn, have evolved mechanisms to detect pathogens and alter their own fatty acid and glucose metabolism to increase their inflammatory and antibacterial capacity (Jha et al., 2015; Naujoks et al., 2016; Russell et al., 2019; Tannahill et al., 2013). Host cells have also evolved mechanisms to produce toxic molecules, including reactive oxygen species (ROS) and reactive nitrogen species (RNS) that can kill intracellular pathogens. Host factors within any of these processes or within the canonical innate immune response could be potential pharmacological targets; provided temporary pharmacological inhibition of the target factor is not a detriment to the host. Potentially targetable host factors may also be involved in as yet undiscovered host-pathogen interactions.

### 1.2.4. Innate responses

There are several types of innate response effector mechanisms that target intracellular bacterial and parasitic infections, including phagocytosis, production of ROS or RNS, xenophagy, programmed host cell death, Interferon-induced guanosine triphosphatases (GTPases), and the restriction of host metabolic processes (such as metal withholding) (Wessling-Resnick, 2015). Additionally, recognition of bacterial infection, or even damage generated by infection, by the innate immune system is an integral feature of innate response initiation. Some of these innate recognition and host effector mechanisms will be described in further detail below.

#### 1.2.4.1. Recognition of bacteria by plasma membrane and intracellular receptors

Bacteria have extracellular pathogen-associated molecular patterns (PAMPs), such as gram-negative Lipopolysaccharide (LPS), or gram-positive peptidoglycan, as well as intracellular PAMPs (e.g. nucleic acids) - which may be released during lysis. Host cells use innate immune receptors known as pattern recognition receptors (PRRs) to recognise PAMPs. A list of examples of some plasma membrane-associated PRRs, intracellular PRRs, examples of their ligands, signalling locations and downstream functions are presented in **Table 1.1**.

**Table 1.1 List of pertinent PRRs that recognise bacterial PAMPs.** *This list is not exhaustive but provides examples of pertinent PRRs. Nuclear factor kappa B (NF-κB) is a nuclear transcription factor that stimulates a range of inflammatory cytokines and antimicrobial peptides. Pam<sub>3</sub>CSK<sub>4</sub> (Pam3-Cys-Ser-Lys4) is a synthetic bacterial lipopeptide.*

Receptor family	Receptor examples	Examples of Ligands	Signalling cellular locations	Downstream function
Toll-like receptors (TLRs)	TLR2	Diacetylated and triacetylated lipopeptides (components of bacterial membranes), ligand specificity depends on co-receptor (Buwitt-Beckmann et al., 2006; Jin et al., 2007; Kang et al., 2009)	Plasma membrane	Stimulation with Pam <sub>3</sub> CSK <sub>4</sub> increased mitochondrial reactive oxygen species (mROS) and production of proinflammatory cytokines (Papadopoulos et al., 2013; Quero et al., 2017; Sjöstrand et al., 2020; West et al., 2011)
	TLR4	Lipopolysaccharide (LPS)	Plasma membrane, endosomes (Kagan et al., 2008)	NF-κB activation, downstream inflammatory cytokine and antimicrobial peptide production and contributes to macrophage activation
Nucleotide-binding domain-like receptors (NLRs)	Nucleotide-binding oligomerization domain-containing protein 1 (NOD1)	γ-D-glutamyl-meso-diaminopimelic acid, specific to Gram-negative peptidoglycan (Chamaillard et al., 2003; Girardin et al., 2003a)	Cytoplasmic. localises to endosomal vesicles during signalling (Irving et al., 2014)	Autophagy (Irving et al., 2014; Travassos et al., 2010), NF-κB signalling (Girardin et al., 2003a)
	Nucleotide-binding oligomerization domain-containing protein 2 (NOD2)	Muramyl dipeptide, a component of Gram-negative and Gram-positive bacterial peptidoglycan (Girardin et al., 2003b; Inohara et al., 2003)	Cytoplasmic, localises to endosomal membranes during signalling (Barnich et al., 2005; Nakamura et al., 2014)	Autophagy (Travassos et al., 2010) and NF-κB signalling (Barnich et al., 2005)
	NLR family apoptosis inhibitory protein (NAIP)	Flagellin (Kortmann et al., 2015) and Type III Secretion System (T3SS) needle protein (Grandjean et al., 2017; Yang et al., 2013)	Predicted cytoplasmic (Kortmann et al., 2015; Miao et al., 2006; Poyet et al., 2001; Yang et al., 2013)	NAIP- NLR family CARD domain containing 4 (NLRC4) inflammasome activation (Kortmann et al., 2015; Yang et al., 2013) and pyroptosis (Grandjean et al., 2017; Kortmann et al., 2015)

Receptor family	Receptor examples	Examples of Ligands	Signalling cellular locations	Downstream function
	NLR family pyrin domain containing 3 (NLRP3)	Specific ligand undetermined, activated in response to cellular stress, see Swanson et al., (2019) for review.	Within endoplasmic reticulum prior to activation (Misawa et al., 2013; Zhou et al., 2011)	NLRP3 inflammasome activation (Misawa et al., 2013; Zhou et al., 2011) and pyroptosis (Beckwith et al., 2020; Kayagaki et al., 2015; Platnich et al., 2018)
	Cyclic GMP–AMP synthase (CGAS)	B-form double stranded Deoxyribonucleic acid (DNA) (Sun et al., 2013)	Cytoplasmic (Sun et al., 2013)	Activation of stimulator of interferon response cGAMP interactor 1 (STING1) which leads to the production of interferon (IFN) (Ablasser et al., 2013; Diner et al., 2013; Sun et al., 2013; Wu et al., 2013; Zhang et al., 2013)
Pyrin and HIN domain family (PYHIN)	Absent in melanoma 2 (AIM2)	Double stranded DNA (Morrone et al., 2015)	Cytoplasmic (Choubey et al., 2000)	AIM2 inflammasome activation (Dick et al., 2016; Morrone et al., 2015), pyroptosis
	Caspase 4 and 5	LPS (Shi et al., 2014)	Caspase 4 has been found in the cytoplasm and associated with the plasma membrane (The Human Protein Atlas, 2017) and Caspase 5 is predicted to be present in the cytoplasm based on its activity (Shi et al., 2014).	Caspase 4 activation and subsequent pyroptosis (Shi et al., 2014)

Together, these are some of the cell-intrinsic mechanisms of host cell recognition of bacteria and they in part determine the initiation of cell intrinsic host responses. Besides PAMP recognition, mammalian cells have an additional mechanism of recognising infection – through the exposure of Damage Associated Molecular Patterns (DAMPs), released by host cells or organelles in response to cellular or tissue damage. DAMPs (such as high mobility group box 1 - HMGB1 (Huebener et al., 2015; Wen et al., 2019; Wu et al., 2012), heat shock protein family A (Hsp70) member - HSP70 (Asea et al., 2000), or mitochondrial Deoxyribonucleic acid (DNA)) can be recognised through classical PRRs (Aarreberg et al., 2019) or through non-PRR DAMP sensors (El Mezayen et al., 2007).

#### 1.2.4.2. Phagocytosis, lysosome-mediated pathogen degradation and reactive oxygen species

Professional phagocytes (neutrophils, monocytes, macrophages, and dendritic cells — DCs) efficiently phagocytose bacteria (Gordon, 2016). Phagocytosis is activated by multiple mechanisms including pathogen recognition by plasma membrane PRRs (Peiser et al., 2000). Though TLRs do not directly mediate phagocytosis of bacteria (Gordon, 2016), TLR4 coordinates with phagocytic receptors to greatly enhance uptake (Blander and Medzhitov, 2004; Seixas et al., 2018). Plasma membrane remodelling and changes in the actin cytoskeleton generate the formation of pseudopods (finger-like projections formed from the plasma membrane) (Masters et al., 2013; Michl et al., 1979) that encircle the bacterium. Fusion of these finger-like projections creates a complete single membrane vacuole called an early phagosome. The early phagosome fuses with endocytic vesicles to recruit proteins necessary for maturation and allows sorting vesicles to fission from it for protein recycling (Lee et al., 2020). As part of this recycling process, a protein complex called retromer ferries proteins to the *trans*-Golgi network (Rojas et al., 2008). The early phagosome thus matures into a late phagosome. Throughout the maturation of the phagosome into the phagolysosome (the final antimicrobial form), vacuolar type adenosine triphosphatases (V-ATPases) are recruited and the vacuole becomes steadily more acidified (Tabata et al., 2020). To become a phagolysosome, lysosomes (vesicles containing degradative enzymes such as cathepsins or lysozymes) fuse with the phagosome, which then has an optimal acidic pH of 5 - 5.5. Depending on the host cell type and activation status, the intraphagosomal nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex (produces reactive oxygen species, ROS) may be recruited to generate an oxidising intravacuolar environment (Nauseef, 2019). The combination of these factors creates a highly antimicrobial vacuole, destroying the majority of the bacteria captured within it; however, some bacteria have evolved to efficiently evade phagolysosomes-mediated destruction.

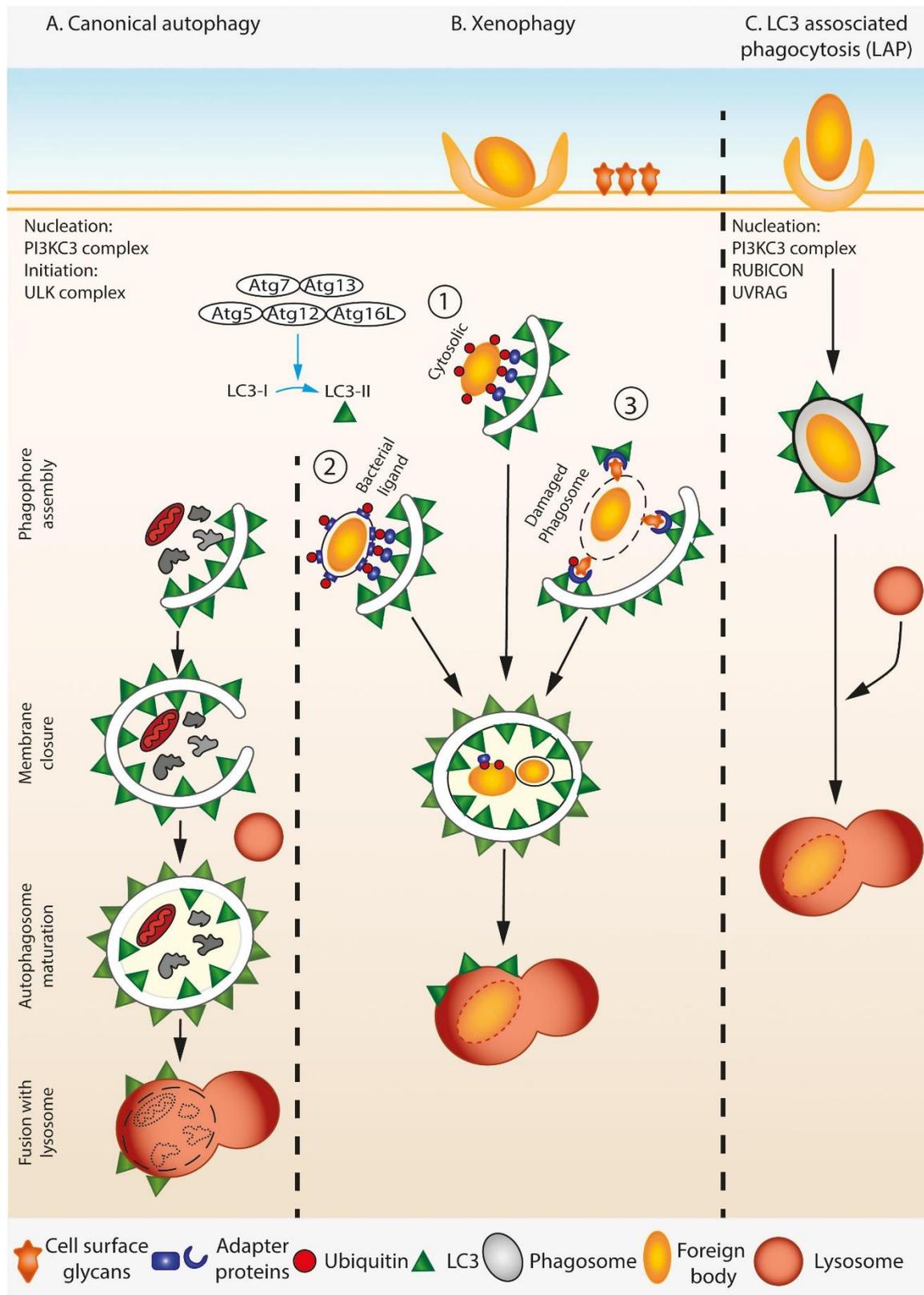
#### 1.2.4.3. Xenophagy

Xenophagy is a category of autophagy, which is the process of catabolic degradation of cytoplasmic and membrane components of the cell (Mitchell and Isberg, 2017). Canonical xenophagy plays a crucial role in pathogen elimination in many cell types, professional and non-professional immune cells alike (Fujita et al., 2009; Lapaquette et al., 2009; Nozawa et al., 2017). For example, canonical xenophagy is initiated in response to *S. enterica* ser. Typhimurium (henceforth, referred to as STM for brevity) infection in mouse embryonic fibroblasts and 3T3 (mouse cell line, fibroblast-derived) cells (Kageyama et al., 2011). A schematic presenting some of the major differences between classical autophagy and different types of xenophagy can be found in **Figure 1.1**.

Canonical xenophagy begins when the cell recognises the pathogen after it has invaded, either through intracellular recognition of PAMPs (Cooney et al., 2010; Homer et al., 2010; Irving et al., 2014; Travassos et al., 2010), or damaged pathogen vacuole membranes (Mansilla Pareja et al., 2017; Meunier et al., 2014; Mitchell et al., 2018). The pathogen, or pathogen vacuole membrane,

becomes labelled with chains of multiple ubiquitin molecules, targeting it for autophagy (Noad et al., 2017). Initiation complexes composed of autophagy-related (ATG) proteins accumulate at the fledgling autophagophore (a double membrane isolation vacuole) membrane, which is localised next to the ubiquitinated pathogen. Additional ATG protein containing complexes form in association with the autophagophore including the Autophagy related 16 like 1 (ATG16L) complex which is composed of Autophagy related 5 (ATG5), Autophagy related 12 (ATG12) and ATG16L1 (Fujita et al., 2013; Kageyama et al., 2011; Mizushima et al., 2001). The ATG16L complex is the enzyme required for the final reaction to ligate a protein from the Microtubule-associated proteins 1A/1B light chain 3 (LC3) family to the autophagophore (Fujita et al., 2013). LC3 proteins are required for fusion of double layer autophagosome membranes (Nakatogawa et al., 2007) to complete the autophagosome vacuole. The autophagosome matures through fusion with lysosomes to degrade the pathogen.

Sometimes the xenophagic pathway can be activated in response to plasma membrane PRR activation by bacterial PAMPs – indicating the presence of bacteria – on the external surface of professional phagocytes (Gluschko et al., 2018; Huang et al., 2009; Hubber et al., 2017). This type of xenophagy is known as non-canonical xenophagy (or microtubule-associated proteins 1A/1B light chain (LC3)-associated phagocytosis). Instead of a double-membraned autophagosome, the bacteria become enclosed in a single membrane phagosome which is then decorated directly with LC3 without ubiquitin (Lam et al., 2013). LC3-associated phagocytosis is morphologically and mechanically distinct from canonical xenophagy and requires only a subset of the proteins (including the ATG16L complex) required in the canonical process (Fletcher et al., 2018). Indeed, the LC3 attached to these single membrane autophagosomes may be required for fusion with lysosomes for bacterial degradation (Martinez et al., 2015), a different function to that performed in canonical xenophagy.



**Figure 1.1 Summary of canonical autophagy, xenophagy and non-canonical (LC3 associated phagocytosis) xenophagy.** **A**, Nucleation and initiation complexes composed of autophagy-related (ATG) proteins coordinate the autophagy machinery during canonical autophagy. LC3 is added to the double membrane as it grows and the membrane engulfs the cellular cargo. Lysosome-autophagosome fusion occurs once the autophagosome is mature, causing the degradation of the contents by the acidic environment. **B**, during (canonical) xenophagy, many of the same autophagy-related (ATG) proteins bring LC3 proteins to the double-layered isolation membrane. The host cell specifically recognises the pathogen (foreign body), both **1**. Cytosolic pathogens and **2**. pathogen-containing vacuoles (PCV) are identified through PAMPs either directly on the pathogen or on the PCV surface. Once recognised, the PAMPs are ubiquitinated, allowing the binding of adaptor proteins which recruit LC3. Intracellular pathogens can also be identified through host glycans (normally on the outer leaflet of the plasma membrane) found on the inside of a perforated PCV. This allows the perforated membrane to be adorned with LC3. In all three cases LC3 targets the pathogen to the autophagosome for degradation. **C**, non-canonical xenophagy requires the PI3KC3 complex involved in nucleation as required in canonical autophagy/xenophagy as well as RUBICON and UVRAG (proteins missing in the other forms). LC3 is added to the outside of the single layer isolation membrane, allowing direct fusion with a lysosome. This image is reprinted from Evans et al. (2018) under a Creative Commons Attribution (CC BY) licence.

#### 1.2.4.4. Programmed cell death

There are multiple mechanisms of programmed cell death (PCD) in which the cell actively destroys itself, and some of these (pyroptosis and necroptosis) are directly caused by recognition of pathogen invasion. Intracellular recognition of PAMPs or DAMPs can cause inflammasome activation, leading to pyroptosis and the concomitant release of inflammatory cytokines interleukin (IL)-1 beta (IL-1 $\beta$ ) (abbreviated to IL-1 $\beta$ ) and IL-18. Examples of inflammasome sensors and their activating ligands include Absent in melanoma 2 (AIM2) which is activated by cytosolic DNA (Bürckstümmer et al., 2009; Fernandes-Alnemri et al., 2009; Hornung et al., 2009; Roberts et al., 2009) and NLR family pyrin domain containing 3 (NLRP3), which is activated in response to signs of cellular stress, such as mitochondrial ROS (mROS), potassium ion efflux, lysosomal damage as well as multiple other signals (Compan et al., 2012; Franchi et al., 2007; Groß et al., 2016; Hornung et al., 2008; Muñoz-Planillo et al., 2013; Perregaux and Gabel, 1994; Pétrilli et al., 2007; Shimada et al., 2012; Zhou et al., 2011). The NLRP3 sensor is also activated through cleavage by caspases 4 and 5 (humans), or 11 (mouse) in response to cytosolic LPS (Baker et al., 2015; Casson et al., 2015; Schmid-Burgk et al., 2015; Shi et al., 2014).

Pyroptosis and necroptosis also cause inflammation in the surrounding tissue, through the release of danger-associated molecular patterns (DAMP)s and, in the case of pyroptosis, inflammatory cytokines IL-1 $\beta$  and IL-18. The DAMPs and cytokines produced result in the recruitment of other immune cells, such as neutrophils. During pyroptotic (and potentially necroptotic) membrane rupture, the soluble cytoplasmic contents are released but bacteria and organelles are trapped within the membranous debris known as a pore-induced intracellular trap (PIT) (Jorgensen et al., 2016). This allows the concomitantly recruited neutrophils (and potentially macrophages) to phagocytose the pathogens (Jorgensen et al., 2017, 2016).

Apoptosis (a type of non-inflammatory PCD) can also occur in response to some pathogen infections. Apoptotic bodies form (with intracellular contents trapped inside) and recruit professional phagocytes (such as macrophages), or neighbouring tissue layer cells, to phagocytose and remove the bodies (known as efferocytosis) – review by Poon et al., (2014). Apoptotic cells actively inhibit efferocytosis by neutrophils (Bournazou et al., 2009) to reduce the potential for inflammation. Necroptotic, apoptotic and pyroptotic pathways have multiple signalling connections ensuring PCD

can occur even if one form is inhibited by the pathogen (Christgen et al., 2020; Doerflinger et al., 2020; Jorgensen et al., 2017). For a review of PCD in infection, see (Jorgensen et al., 2017). The three mechanisms of PCD described are important features of professional and non-professional immune cell responses (Hefele et al., 2018; Miao et al., 2010; Shi et al., 2014). Indeed, besides promoting the recruitment of further immune cells, PCD removes the availability of the intracellular niche. For example, infected epithelial cells undergo expulsion from the tissue layer prior to PCD generated cell lysis (Knodler et al., 2014; Sellin et al., 2014).

#### 1.2.4.5. Interferon-induced GTPases

Interferon (IFN)-induced GTPases are a superfamily of guanosine triphosphatases which include guanylate-binding proteins (GBPs) (Ngo and Man, 2017). Their expression can be induced in response to IFN- $\alpha$ , - $\beta$ , and - $\gamma$  in most, but not all, cell types (Cheng et al., 1983; The Human Protein Atlas, 2017; Tretina et al., 2019; Uhlén et al., 2015). Infected cells produce IFNs in response to stimulation of various PRRs including Toll-like receptor 4 (TLR4), Nucleotide-binding oligomerization domain-containing protein 1 (NOD1) (Watanabe et al., 2010), Nucleotide-binding oligomerization domain-containing protein 2 (NOD2) (Leber et al., 2008; Pandey et al., 2009) and many DNA sensors (such as cyclic GMP- AMP synthase (CGAS)) (Man et al., 2015; Meunier et al., 2015; Storek et al., 2015). GBPs are mainly localised to the cytosol or intracellular vesicles (Britzen-Laurent et al., 2010; Kravets et al., 2016; Tripal et al., 2007; Vestal et al., 2000) and participate in intracellular bacterial clearance in three ways. Firstly, they can disrupt bacteria-containing vacuoles to expose the bacteria (and PAMPs, such as LPS) to the cytosol (Feeley et al., 2017; Fisch et al., 2019; Meunier et al., 2014; Pilla et al., 2014). Secondly, they can recruit NADPH oxidase and autophagic agents to the bacteria-containing vacuole for bacterial degradation (B.-H. Kim et al., 2011). Last but not least, they can lyse cytosolic bacteria to expose bacterial DNA to promote inflammasome activation and subsequent cell death (P. Li et al., 2017; Man et al., 2015; Meunier et al., 2015; Wandel et al., 2017).

### 1.3. Rationale for selection of pathogenic bacteria for infection screens

Three important intracellular pathogens were chosen to assess the potential of host response gene products as therapeutic targets at various stages of target discovery and validation; STM, *Shigella sonnei* and *F. tularensis*. Species of *Salmonella* and *Shigella* are listed as high priority and priority level pathogens respectively, requiring further research into antibacterial treatments due to concerning levels of antibiotic resistance (Tacconelli and Magrini, 2017). *S. sonnei* has widespread fluoroquinolone resistance (Chung The et al., 2019, 2016; De Lappe et al., 2015). All of the selected species are Gram negative bacteria, thus they have an intrinsic mechanism of resistance to some antibiotics due to a double cell membrane (hydrophilic and hydrophobic) (Acosta-Gutiérrez et al., 2018). Some species of bacteria can become temporarily resistant through the modulation of gene expression – for example, STM forms persistent colonies within macrophages (Stapels et al., 2018). Additionally, STM persisters can function as a reservoir of antibiotic resistance plasmids (Bakkeren et al., 2019).

STM is a globally important pathogen and while it causes relatively low morbidity for the majority of the population, the infection can increase in severity in immune-compromised individuals. *Shigella* spp. are also globally prevalent, causing shigellosis (a form of dysentery), which can be life threatening in the elderly. *F. tularensis* (the agent of tularaemia) is mainly found in the Northern hemisphere in prairie- or farmlands (Rodríguez-Pastor et al., 2017). The infective dose of *F. tularensis* is extremely low and can have a high morbidity depending on the type of exposure as well as the

infecting species. A summary of the prevalence and characteristics of these pathogens can be found in Table 1.2, symptoms can be found in Table 1.3. Based on the low infective dose and high morbidity, the CDC listed *F. tularensis* as category A (the highest threat category) on its potential bioterrorism agents list (Centers for Disease Control and Prevention, 2018). STM and *S. sonnei* were also classed as potential bioterrorism agents (Category B) requiring research into treatment for better preparedness. As the most virulent strain of *F. tularensis*, Schu S4 (subspecies (subsp.) *tularensis*) is a Biosafety level (BSL) 3 pathogen a BSL 2 model of *F. tularensis*, the subsp. *holarctica* live vaccine strain (LVS) was chosen to identify host infection factors that might translate to the virulent Schu S4 strain.

Another pathogen, the parasitic apicomplexan *Toxoplasma gondii*, is also listed in **Table 1.2** and **Table 1.3** as it was used in some preliminary validation prior to this work (described in Section 1.10). *T. gondii* causes toxoplasmosis, a widespread zoonotic disease (Robert-Gangneux and Dardé, 2012) that can be particularly harmful in utero. *T. gondii* is listed as a Category B potential bioterrorism agent (NIH: National Institute of Allergy and Infectious Diseases, 2016) and has a slow growing form (bradyzoite) during chronic infection that lacks effective antimicrobials (Dunay et al., 2018).

**Table 1.2: Characteristics of pathogens used in gene trap screens and phenotyping of CRISPR mutant cells described in Table 1.4. N/A means not applicable**

Pathogen	Gram	Motility	Invasion strategy	niche	Host cell types	geography
<i>Salmonella enterica</i> serotype Typhimurium	negative	motile	Type III secretion system	<i>Salmonella</i> containing vacuole, extracellularly	Epithelial cells, macrophages, monocytes	Likely global
<i>Shigella sonnei</i>	Negative	Non-motile (has a flagellum)	Type III secretion system	cytosol	Epithelial cells, macrophages	Global, most infections in low- and middle-income countries
<i>Francisella tularensis</i>	negative	Non-motile	Clathrin and actin are required for invasion of non-phagocytic cells (Law et al., 2011). Phagocytic engulfment through pseudopod loops (Clemens et al., 2005).	cytosol	wide variety of host cell types including monocytes, macrophages	USA and northern hemisphere (subspecies dependent)
<i>Toxoplasma gondii</i>	N/A	motile	Dependent on parasite secreted proteins	parasitophorous vacuole	all nucleated cells	Likely global, zoonotic, reservoirs in many mammalian species, difficult to detect

**Table 1.3 Symptoms of, and human populations susceptible to, diseases caused by pathogens of interest.** *Gastroenteritis refers to the following constellation of symptoms; nausea, vomiting, diarrhoea, and abdominal pain. Anorexia in this context refers to the loss of appetite as a medical condition. Acute respiratory distress syndrome is a form of severe inflammatory lung trauma frequently precipitated by infection (Laffey et al., 2017). Infection of monocytes by S. enterica ser. Typhimurium (STM) is described in Serbina et al., (2008) and Vazquez-Torres et al. (1999). Information on the epidemiology of shigellosis (the disease caused by Shigella species) is from Kotloff et al., (2018). Information on symptoms of pneumonic tularemia (caused by species of Francisella) is from Adalja et al., (2015) and information on other types of tularemia is from Kingry and Petersen (2014). Information on infection characteristics of tularemia at the cellular level is from Rowe and Huntley (2015) and Meyer et al., (2015). Information on symptoms and susceptibility to melioidosis (caused by Burkholderia pseudomallei) is from Currie et al., (2010). Information regarding molecular infection characteristics of melioidosis is from Vander Broek and Stevens (2017) and Willcocks et al. (2016). Information on epidemiology and molecular characteristics of infection of toxoplasmosis (caused by T. gondii) are from Robert-Gangneux and Dardé (2012).*

Pathogen	Symptoms	Susceptible population
STM	Self-limiting acute gastroenteritis, with watery diarrhoea, bacteraemia in immunocompromised individuals	Detected in young children, the elderly, and immunocompromised individuals, but milder unreported infections occur in healthy individuals
<i>Shigella sonnei</i>	Most cases: fever, headache, malaise, anorexia, and vomiting followed by watery diarrhoea. Some asymptomatic. In others symptoms progress to small frequent bloody stools, abdominal cramps and tenesmus	All ages, most infections in children 1-4 yrs age
<i>Francisella tularensis</i>	Dependent on route of transmission. Pneumonic tularemia (caused by inhalation of aerosols); starts with symptoms similar to pneumonia, leads to cytokine storm, 'acute respiratory distress syndrome', and respiratory failure. Other modes of transmission result in ulcers at the site of infection, swollen glands and fever.	All
<i>Toxoplasma gondii</i>	Otherwise healthy individuals (including most pregnant women) present as asymptomatic, otherwise fever or enlarged cervical lymph nodes, some experience muscle pain, physical weakness, ocular inflammation	People at risk are pregnant women, foetuses, immunocompromised individuals

## 1.4. Infection lifecycle of bacteria of interest

The interactions of various bacteria of interest with host cells during the course of infection is described below to give some context to the cell autonomous host responses to be investigated.

### 1.4.1. *Salmonella enterica* serotype Typhimurium infection

Infection with non-typhoidal *Salmonella* — those causing gastroenteritis — (e.g. STM) begins with ingestion of contaminated substances (normally food). Once swallowed, STM in food is carried through the gastrointestinal system to the small intestine through peristalsis. *Salmonella* spp. modify Lipid A (part of LPS) in response to host-invasion cues so these bacteria can avoid detection by Toll-like receptor (TLR) 4 (Needham and Trent, 2013). *Salmonella* can invade the intestinal epithelia via non-phagocytic epithelial cells (Tahoun et al., 2012) by inserting a Type III Secretion System (T3SS) — encoded by *Salmonella* pathogenicity island (SPI)-I — into the plasma membrane to transfer effectors into the host cell. These effectors (virulence proteins adapted to control host processes) promote actin polymerisation and membrane ruffling, facilitating bacterial uptake (Raffatellu et al., 2005; Zhang et al., 2018). One of these effectors, Guanine nucleotide exchange factor SopE (SopE), activates host Ras-related C3 botulinum toxin substrate 1 (RAC1) the activity of which greatly enhances membrane ruffling and invasion (Criss et al., 2001; Friebel et al., 2001; Patel and Galán, 2006; Rudolph et al., 1999). *Salmonella* is also frequently taken up by phagocytic microfold cells — an epithelial cell adapted to sample the intestinal interior (Jones et al., 1994; Radtke et al., 2010) — which are localised in Peyer's patches (specialised regions of the intestinal wall). In either case, a vacuole forms around the STM. STM promotes maturation of this vacuole (to make it suitable for bacterial replication) by secreting further effectors through a different T3SS encoded by SPI-II. The mature vacuole is known as the *Salmonella* containing vacuole (SCV). Mouse NLR family apoptosis inhibitory proteins (NAIPs) recognise *Salmonella* T3SS components and flagella (Kofoed and Vance, 2011; Rayamajhi et al., 2013; Yang et al., 2013; Zhao et al., 2011). In some *Salmonella*-infected epithelial cells within the intestinal epithelium, one or more of these components are recognised, causing the activation of the NAIP-NAIP-NLR family CARD domain containing 4 (NLRC4) inflammasome, extrusion and pyroptosis of *Salmonella*-infected epithelial cells (Sellin et al., 2014). In mice, this results in reduced *Salmonella* replication within the gut at early time points. The human NAIP also recognises T3SS needle and flagella proteins (Grandjean et al., 2017; Kortmann et al., 2015; Yang et al., 2013), though the potential effects of this have not been investigated.

A small proportion of SCVs become perforated due to intra-vacuolar SP-I-T3SS mediated damage (Birmingham et al., 2006; Knodler et al., 2010), releasing the bacterial contents into the cytoplasm (Laughlin et al., 2014; Liss and Hensel, 2015; Thurston et al., 2016). Caspase 4 activates the inflammasome in response to liberated, cytosolic LPS in *Salmonella*-infected human epithelial cells (Knodler et al., 2014; Shi et al., 2014). The inflammasome activation causes epithelial cell extrusion from the epithelium into the intestinal lumen and undergo pyroptosis (Knodler et al., 2014, 2010; Shi et al., 2014). Intracellular *Salmonella* within pyroptotic host cells remain secluded within the PIT (as described above) and may be

phagocytosed by recruited neutrophils or macrophages (consequences described below) (Jorgensen et al., 2017, 2016).

Phagocytosis by macrophages results in further bacterial multiplication through the establishment of a sustained population of STM inside SCVs (Miao et al., 2010), though a small population of bacteria can also be found in the cytosol as a result of a host cell response (Perrin et al., 2004). This occurs in large part through the process of GBP-mediated SCV lysis (Meunier et al., 2014). Chromosome 3 GBPs also mediate autophagy of STM in perforated SCVs (**Figure 1.1**

). Galectin-8 (LGALS8) (a cytosolic beta-galactoside-binding lectin) recognises and binds to host glycans found on the inside of perforated SCVs. Galectin-8 recruits microtubule-associated proteins 1A/1B light chain (LC3), a key autophagy protein, to newly cytosolic STM in a GBP dependent manner (Meunier et al., 2014); thus, promoting xenophagic digestion of STM. Exposure of STM to the cytosol can also activate the inflammasome and stimulate pyroptosis (Meunier et al., 2014). Macrophages can also respond to the recognition of intracellular *Salmonella* infection by initiating cell death. Macrophages are able to recognise *Salmonella* and activate the inflammasome to initiate pyroptosis through many mechanisms (as introduced in Sections 1.2.4.1 and 1.2.4.4) including the NAIP-NLRC4 inflammasome (Bierschenk et al., 2019; Miao et al., 2010) and caspase 4 activation (Bierschenk et al., 2019). Further detail describing the macrophage as a replicative niche for STM is in Section 1.6.1 and a summary schematic of STM macrophage infection is presented later in Figure 1.2.

STM can outcompete colonic microbiota in the lumen under inflammatory conditions (Stecher et al., 2007) by using the sulphur-containing electron acceptor tetrathionate (Winter et al., 2010). Winter et al., (2010) discovered that tetrathionate is formed in the inflamed colonic lumen by oxidation of thiosulphate, mainly due to NADPH oxidase (Cybb)-dependent release of oxygen radicals. Thiosulphate is generated by colonic mucosal tissue from toxic hydrogen sulphide (Furne et al., 2001; Levitt et al., 1999) which is itself produced by microbiota within the colon. If invading neutrophils phagocytose extracellular STM, they efficiently kill the bacteria with ROS (Miao et al., 2010) making neutrophils a key effector cell type in *Salmonella* clearance from the gut. Human STM infection normally produces self-limiting gastroenteritis but can become systemic in immunocompromised individuals (Gordon, 2008).

#### 1.4.2. *Francisella tularensis* infection

Human *F. tularensis* infection occurs via infected animals or vectors through multiple routes. The pathogen localises to a variety of organs, including the lungs, skin, small intestine and eye; though the focus depends on the route of transmission (Cross et al., 2019; Maurin and Gyuranecz, 2016). It has an extracellular phase (Forestal et al., 2007) but the intracellular phase appears to be the most important for productive infection. *F. tularensis* enters, survives and proliferates inside many types of host cells, including but not limited to, neutrophils (McCaffrey and Allen, 2006), lung epithelial cells (Hall et al., 2007) and macrophages (Hall et al., 2008). Information on the mechanisms of infection come from a combination of pathogen subspecies (*F. tularensis* subsp. *novicida* U112 strain, subsp. *holarctica* LVS, subsp. *tularensis* Schu S4 strain), host species (mouse, human) and cell types; which can change the host-pathogen interactions substantially (Celli and Zahrt, 2013). Thus, where possible information garnered from experiments using the subsp. *holarctica* strain LVS and the subsp. *tularensis* strain Schu S4 in human cells will be used.

In mouse bone marrow-derived macrophage (BMDM) infection, *F. tularensis* subsp. *tularensis* Schu S4 transiently appears in late endosomes or phagolysosomes (as measured by lysosomal-associated membrane protein 1 (LAMP-1) and LysoTracker Red DND-99 bacterial co-localisation) before escaping (Chong et al., 2008; Geier and Celli, 2011). The *novicida* subsp. also transiently co-localises with the host proton v-ATPase complex

(required for acidification and phagosome maturation) and LysoTracker Red DND-99 in human monocyte derived macrophages (MDMs) before also entering the cytosol (Santic et al., 2008). The Type VI Secretion System (T6SS)-like proteins (such as VrgG) in *Francisella* are predicted to puncture host plasma- and vacuole- membranes (Pukatzki et al., 2007) to introduce bacterial effectors into the cytoplasm. The cytosol is permissive for replication of *Francisella*. Indeed the secretion system is required for intracellular proliferation and phagosome escape in the *novicida* subsp. (Brodmann et al., 2017) and the *holarctica* subsp. LVS (Lindgren et al., 2004). However, cytosolic bacteria are also exposed to the cell-autonomous innate immune response. For example, infection of mouse BMDMs with the *novicida* or *holarctica* (LVS) subsp. results in AIM2 inflammasome activation with concomitant IL-1 $\beta$  production (Fernandes-Alnemri et al., 2010; Rathinam et al., 2010). Infection of phorbol 12-myristate 13-acetate (PMA)-treated human THP-1 (macrophage-like) cells with the *novicida* or *holarctica* subsp. LVS strain produces NLRP3 inflammasome activation in addition to the AIM2 inflammasome (Atianand et al., 2011). *F. tularensis* has evolved to avoid or resist other cell-autonomous innate immune responses, as would be expected of a cytosol specialised pathogen. For example, the *novicida* subsp. U112 and the *holarctica* subsp. LVS are somewhat resistant to BMDM produced ROS (Llewellyn et al., 2011). Additionally, BMDMs fail to produce ROS upon *tularensis* subsp. Schu S4 infection, despite a normal capacity to produce ROS under other conditions (Geier and Celli, 2011), suggesting some form of pathogen-mediated suppression. *Novicida* subsp. strain U112 and *tularensis* subsp. strain Schu S4 avoid the xenophagic pathway in human embryonic kidney (HEK)-293T and human monocyte-derived macrophage (MDM) cells despite lengthy periods of replication in the cytosol (Akimana et al., 2010; Edwards et al., 2010). Respiratory tularaemia (caused by aerosolised *F. tularensis*) has a high mortality rate (~40%) in humans (Foley and Nieto, 2010). The macrophage as a replicative niche for *F. tularensis* is described further in Section 1.6.2, along with a summary schematic in Figure 1.2.

#### 1.4.3. *Shigella sonnei*

The vast majority of research literature on *Shigella* infection is about *Shigella flexneri*, so most of this information is inferred from *S. flexneri* research. Human *Shigella* infection begins with ingestion. The bacteria move through the intestinal epithelium by transcytosis through microfold cells (Ranganathan et al., 2019), allowing access to the basolateral side. Macrophages and DCs found in the microfold cell pocket phagocytose the bacteria, which quickly escape the phagosome into the cytosol through a T3SS dependent mechanism (Paetzold et al., 2007). *Shigella* replicates in the cytosol, before instigating the pyroptotic cell death of its host cell (Arizmendi et al., 2016; Kayagaki et al., 2015; Suzuki et al., 2014) allowing the bacteria access to the basolateral face of the intestinal epithelial wall. This brief intracellular interaction with macrophages is summarised later in Figure 1.2. *Shigella* invade the basolateral side of the epithelial cells (Ranganathan et al., 2019), using the T3SS to secrete effectors (including IpgB1) directly into the host cell cytoplasm, inducing bacterial uptake. *Shigella* effector IpgB1 localises Rac1 activation – the same host GTPase *Salmonella* uses for invasion – to enable host cell invasion (Bulgin et al., 2010; Criss et al., 2001; Ohya et al., 2005; Patel and Galán, 2006; Weigele et al., 2017). Once in the epithelial endocytic vacuole *Shigellae* escape into the cytosol in a T3SS dependent manner (Du et al.,

2016). The intestinal epithelial cell cytosol is the optimal host cell niche and the bacteria replicate intensely. To maintain this cellular niche, *Shigella* regulates many immune responses in host epithelial cells (Schnupf and Sansonetti, 2019) including the downregulation of inflammatory chemokines and inhibition of programmed cell death (Bergounioux et al., 2012; Kobayashi et al., 2013; Sperandio et al., 2008). Another method of host response escape is the use of actin tails to physically push through into neighbouring host cells; generating a double membrane from which it escapes (Michard et al., 2019). Uninfected bystander epithelial cells instigate an inflammatory response, secreting IL-8 (Kasper et al., 2010) which recruits neutrophils. Infected epithelial cells eventually die by necrosis (Carneiro et al., 2009) and pyroptosis (Dupont et al., 2009), upon which Shigellae are likely retained in PITs (Jorgensen et al., 2016). Neutrophils or macrophages may phagocytose the Shigellae containing-PITs (Jorgensen et al., 2016). Neutrophils, unlike macrophages, are able to efficiently kill phagocytosed *Shigella* (Li et al., 2010) and are well-established as the key effector cells in *Shigella* clearance (Hermansson et al., 2016) and disease resolution.

## 1.5. Phagocytes as important innate host defence cells

Professional phagocytes include macrophages, dendritic cells, monocytes and neutrophils. Macrophages are found resident in individual tissues or are recruited from the bone marrow as monocytes, which then differentiate into macrophages. Tissue macrophages and DCs act as resident sentinels of the host tissue; the former scanning for deviations in homeostasis, the latter sampling the environment and educating the adaptive immune system. Tissue macrophages have expression profiles that are unique to their tissue microenvironment (Gautier et al., 2012); but broadly maintain homeostasis as well as mediating tissue repair responses in post-infection or chronic inflammatory settings. As these phagocytes are already present in the tissue, they are uniquely situated to respond to invading pathogens. Macrophages, unlike DCs are well equipped to kill or coordinate the destruction of most microorganisms (though this is dependent to some degree on their activation state) (Gordon, 2016; McCurley and Mellman, 2010).

Human blood circulating monocytes have the capacity, through further differentiation, to contribute to tissue macrophage populations (which endocytose debris) (Bajpai et al., 2018; Bujko et al., 2017), resident DC populations (which sample the local environment) (Richter et al., 2018) and inflammatory DCs (Goudot et al., 2017; Segura et al., 2013; Wilsmann-Theis et al., 2013) - reviewed in Tang-Huau and Segura (2019) in response to differential signals. There are numerous circumstances under which monocytes extravasate (move across a vessel wall) into tissues including, most relevantly, bacterial infection. In mouse blood, monocytes respond to signals characteristic of bacterial infection by differentiating into 'effector monocytes', a term coined by Mildner et al., (2013). These effector monocytes express tumour necrosis factor (Tnf) and inducible nitric oxide synthase (Nos2) (produces reactive nitrogen species - RNS), are inflammatory and thus generate tissue damage to both host and pathogen (Serbina et al., 2003). Human macrophages also express NOS2 (Bogdan, 2015; Hedl et al., 2019).

### 1.5.1. Classical and alternative activation of macrophages

Macrophages have a continuous spectrum of phenotypes *in vivo* broadly categorised as classically- (inflammatory) or alternatively- (anti-inflammatory) activation states. *In vitro* macrophages can be modelled by culturing and differentiating human blood monocytes into macrophages – known as MDMs – or by culturing mouse BMDMs. For practical purposes the macrophage phenotype spectrum is modelled by a distinct number of activation states *in vitro* (Murray, 2017; Murray et al., 2014), of which IFN- $\gamma$  + LPS stimulated macrophages – M1(IFN- $\gamma$  + LPS) – and IL-4 stimulated macrophages – M2(IL-4) – will be described for contrast. The combination of IFN- $\gamma$  and LPS induces a strong inflammatory phenotype, replete with secreted inflammatory cytokines, such as TNF- $\alpha$ , IL-12p40, IL-6, IL-8 and IFN- $\gamma$  (Tarique et al., 2015; Vogel et al., 2014). M1(IFN- $\gamma$  + LPS) MDMs also secrete ROS (Sanjurjo et al., 2018), an important component of the microbicidal arsenal of inflammatory macrophages (Lv et al., 2017; Yang et al., 2009). Human blood monocytes and M1(IFN- $\gamma$  + LPS) MDMs also generate intraphagosomal ROS through the action of NADPH oxidase (Canton et al., 2014; Foote et al., 2019). Additionally, IFN- $\gamma$  stimulated mouse BMDMs are more effective than unstimulated BMDMs at *Escherichia coli* and *Burkholderia pseudomallei* (a facultative intracellular bacterium) clearance *in vitro* (Eske et al., 2009). In contrast, M2(IL-4) MDMs have reduced proinflammatory cytokine (*IL8*, *IL6*, *IL1A*, *IL18*, *TNF*) transcription (Chaitidis et al., 2005; Gharib et al., 2019) and proinflammatory cytokine reduced secretion (of IL-12, TNF- $\alpha$  and IL-6) (Vogel et al., 2014). M2(IL-4) MDMs do not secrete ROS either (Sanjurjo et al., 2018). Instead M2(IL-4) MDMs and mouse BMDMs serve as *in vitro* models to study polarised functions such as the maintenance of homeostasis – such as heat regulation (Nguyen et al., 2011), controlling fat storage (Moratal et al., 2018) – as well as mediating tissue healing (Bosurgi et al., 2017). M2(IL-4) polarised MDMs also efferocytose more apoptotic cells than genotype media matched controls (Gharib et al., 2019).

Macrophage polarity is tightly linked to metabolism. Indeed, the Tricarboxylic acid (TCA) cycle becomes disrupted in M1(IFN- $\gamma$  + LPS) macrophages, generating an increased level of succinate (Jha et al., 2015). The succinate is required for the indirect induction of IL-1 $\beta$  (Tannahill et al., 2013). Furthermore, changes in the activity of the electron transport chain in response to bacterial stimulation or exogenous IFN- $\gamma$  and LPS cause excess mROS generation (Chouchani et al., 2014; Garaude et al., 2016; Guarás et al., 2016; Jha et al., 2015; Jin et al., 2014; Lapuente-Brun et al., 2013; Lopez-Fabuel et al., 2016). The excess mROS is linked to reductions in intracellular bacterial burden (Roca and Ramakrishnan, 2013; West et al., 2011).

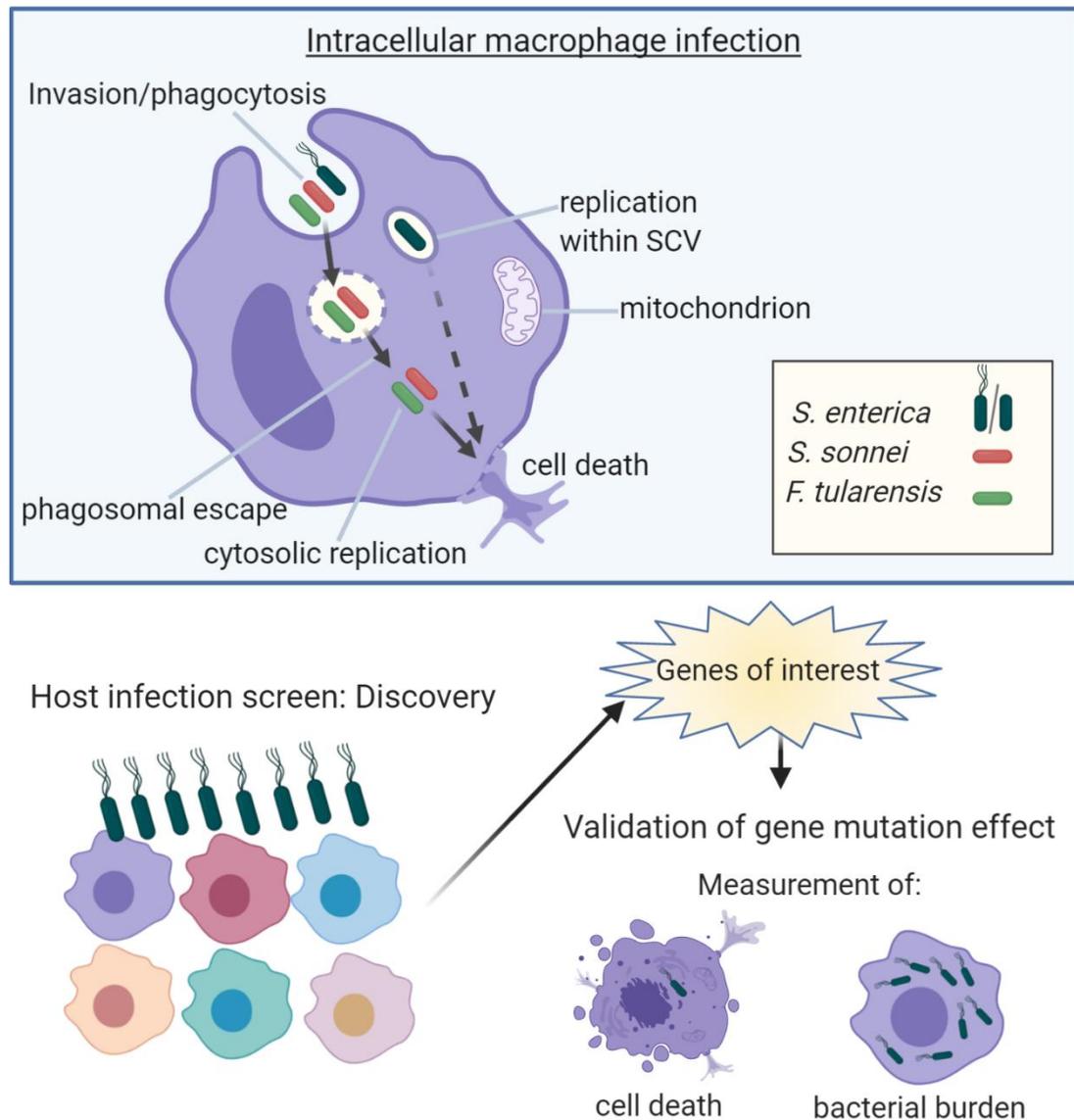
Excess mROS reduces intracellular bacterial burden (Roca and Ramakrishnan, 2013; West et al., 2011) through three possible mechanisms. Firstly, mROS production is an important mechanism of antibacterial activity within macrophages (Roca and Ramakrishnan, 2013; West et al., 2011), but the mechanism for this is still unclear. Furthermore, STM lacking the ability to stabilise the SCV become exposed to the cytoplasm, resulting in exposure to oxidative stress and consequent reduced bacterial replication (Heijden et al., 2015). Indeed, it has been reported that mitochondria are brought in proximity to bacterial PAMP containing phagosomes (West et al., 2011); thus, mROS may be directly bactericidal. A small proportion of wild type STM are found in the cytoplasm (Knodler et al., 2010;

Meunier et al., 2014; Perrin et al., 2004), so this oxidative activity would also affect wild type STM. Secondly, mROS (in response to LPS stimulation) has been shown to stimulate proinflammatory cytokine and *Nos2* gene expression in BMDMs and fibroblasts (Bulua et al., 2011; Jin et al., 2014). The proinflammatory cytokines indirectly aid the infection response and *Nos2* (also expressed by human macrophages (Bogdan, 2015)) produces bactericidal reactive nitrogen species (Serbina et al., 2003). NLRP3 appears to have multiple activating signals, though the mechanism for integrating these is unclear (Groß et al., 2016; Shimada et al., 2012; Zhou et al., 2011).

In contrast, M2 macrophages have a complete TCA cycle and enhanced mitochondrial oxidative phosphorylation which are required for M2 polarised functions.

## 1.6. Macrophages as a pathogen niche

Intracellular invasion of macrophages and epithelial cells by *Salmonella* is required for a productive infection, as this allows avoidance of neutrophil-mediated killing (LaRock et al., 2015). *F. tularensis* and *Shigella* also replicate and produce productive infections inside macrophages (Ashida et al., 2011; Case et al., 2014). The macrophage as a common pathogen niche for these bacteria is summarised in Figure 1.2 – thus, host-pathogen interactions within this cell type are of particular interest.



**Figure 1.2** Three diverse intracellular bacterial pathogens (*S. enterica*, *F. tularensis*, *S. sonnei*) share the macrophage as a replicative niche – thus the macrophage is used as an *in vitro* model to screen for shared host infection factors. *Top* - *S. enterica* and *S. sonnei* utilise similar mechanisms of host cellular entry and the latter shares a cytoplasmic intracellular niche with *F. tularensis*; yet each differs in aspects of the mechanisms of achieving entry, survival, and replication within macrophages. The vast majority of *S. enterica* bacteria replicates inside a Salmonella containing vacuole (SCV), where the environment can be controlled. Meanwhile, the macrophage responds to the bacterial invasion by modifying its metabolic activity to aid intracellular antibacterial activity. Eventually the infection sets off a cascade of events resulting in (often lytic) host cell death. *Bottom* - two macrophage host infection screens are planned to identify shared, perhaps even universal, host infection factors that may be pharmacologically modulated to reduce bacterial burden, thus, reducing morbidity. These genes must be validated in independent knockouts (KOs) through measurement of host cell death in response to infection, as well as intracellular bacterial burden.

### 1.6.1. Macrophage responses to *S. enterica* ser. Typhimurium

Many intracellular pathogens are preferentially found in macrophages with specific phenotypes and have effectors that modulate macrophage responses to promote the preferred phenotype. In *in vitro* infection experiments using STM, mouse BMDMs could be categorised into 4 groups; uninfected bystander macrophages, macrophages containing host-killed STM, macrophages with active non-growing STM and macrophages with growing STM (Stapels et al., 2018). Mouse BMDMs containing host-killed STM expressed M1-like transcriptional markers (nuclear factor kappa B subunit 2 (*Nfkb2*), CD40 molecule (*Cd40*), *Il1b*, *Nlrp3*, and *Tnf*) (Saliba et al., 2016; Stapels et al., 2018). In contrast, mouse BMDMs containing growing STM expressed M2 transcriptional markers (interleukin-4 receptor alpha (*Il4r*), arginase 1 (*Arg1*), ornithine decarboxylase 1 (*Odc1*), peroxisome proliferator-activated receptor delta (*Ppard*), and TIMP metalloproteinase inhibitor 1 (*Timp1*)) (Saliba et al., 2016; Stapels et al., 2018) and BMDMs with intracellular active non-growing STM expressed a combination of M1-like and M2-like transcriptional markers (Stapels et al., 2018). The observation that M2 macrophages (as measured by IL-4 receptor alpha chain – IL4RA – surface expression) with growing STM was additionally sustained in an acute STM model *in vivo* (Stapels et al., 2018). These suggest that a less inflammatory cellular environment aids STM survival, indeed, STM are found to actively drive M2-like activation of BMDMs *in vitro* (Stapels et al., 2018). A recently described mechanism for this is the interaction of a STM effector protein with Glycogen Synthase Kinase 3 Beta (GSK3B), causing GSK3B to phosphorylate and activate signal transducer and activator of transcription 3 (STAT3), thus, triggering polarisation towards an M2-like phenotype (Gibbs et al., 2020; Panagi et al., 2020). Furthermore, the beneficial environment inside M2 macrophages is also seen in human biology, as STM replicates faster in M2(IL-4) MDMs than in MDMs stimulated with IFN- $\gamma$  + LPS (Lathrop et al., 2015). During persistent *in vivo* infection STM are more frequently found within IL4RA (alternatively activated) macrophages than in inflammatory nos2 expressing effector monocytes (Pham et al., 2020); suggesting that the former are an important bacterial reservoir.

### 1.6.2. Macrophage responses to *F. tularensis* Schu S4

*F. tularensis* LVS infection of BMDMs stimulates IL-6, IL-12 (p40 subunit) and IL-10 secretion (Griffin et al., 2013; Navratil et al., 2014). In contrast, the virulent *F. tularensis* Schu S4 strain reduces the transcription of the PRRs and associated adaptors Toll-like receptor 2 (*TLR2*), *TLR4*, CD14 molecule (*CD14*) and MYD88 innate immune signal transduction adaptor (*MYD88*) in infected MDMs; furthermore, infected MDMs are unresponsive to TLR2 or TLR4 stimuli (Butchar et al., 2008). Additionally, Nuclear factor kappa B (NF- $\kappa$ B) activation is prevented during Schu S4 infection of MDMs; thus preventing TNF- $\alpha$  and IL-6 secretion in response to exogenous IFN- $\gamma$  (Melillo et al., 2010). Thus, *F. tularensis* Schu S4 appears to create hyporesponsive host macrophages *in vitro*. Nevertheless, exogenous IFN- $\gamma$  limits *F. tularensis* Schu S4 growth inside murine BMDMs and human MDMs (Edwards et al., 2010), suggesting that IFN- $\gamma$  still has some intracellular effect. *F. tularensis* LVS infection of the mouse peritoneal cavity stimulated the alternative activation (Resistin-like alpha (Retnla) aka FIZZ-1<sup>+</sup>) of peritoneal macrophages (Shirey et al., 2008). However, the importance of

alternatively activated macrophages in *F. tularensis* LVS or Schu S4 infection could not be corroborated when mice were inoculated intranasally to infect the lung (to emulate pulmonary tularemia) (D'Elia et al., 2015; Griffin et al., 2013). Indeed, *F. tularensis* strain Schu S4 and LVS mouse infections lacked a significant infiltration of anti-inflammatory macrophages (characterised in D'Elia et al., (2015) as FIZZ-1<sup>+</sup>, Arg1<sup>+</sup> and FIZZ-1<sup>+</sup>, Arg1<sup>+</sup>, chitinase-like 3 (chl3) aka (Ym-1)<sup>+</sup> in Griffin et al., (2013)) into the site of pathology (D'Elia et al., 2015; Griffin et al., 2013). Therefore, these alternatively activated macrophages were not available as an intracellular niche. Instead the *F. tularensis* Schu S4 infected mouse lung had high numbers of classically activated macrophages (NOS2 staining) (D'Elia et al., 2015). The animals developed high levels of morbidity, despite the high numbers of classically activated macrophages, perhaps due to the resistance of Schu S4 to oxidative stress (Melillo et al., 2010).

### 1.6.3. Macrophage responses to *S. sonnei*

*Shigella sonnei* and *flexneri* evolved at different times from *Escherichia coli* harbouring the pINV virulence plasmid (which contains the type III secretion system and effectors required for intracellular infection) (Hawkey et al., 2020). No literature is available on the effects of *S. sonnei* effectors on macrophage polarisation or activity, so literature on *S. flexneri* will be discussed where relevant. Additionally, most of the literature describing host cell interactions with *Shigella* effectors have been performed in non-macrophage cells (e.g. HeLa cells) so the majority of this section suggests what might happen in macrophages on this basis. A major difference between the response of macrophages and colonic epithelial cells (enterocytes) to *Shigella* is that the former undergoes inflammatory cell death a short time after infection (Arizmendi et al., 2016; Kayagaki et al., 2015; Suzuki et al., 2014; Watson et al., 2019). *S. flexneri* induced macrophage cell death occurs through the engagement of the NLRP3, NLRC4 and NLRP1B inflammasomes through the action of the IpaH7.8 effector (homologous protein expressed by *S. sonnei*, Uniprot accession ID Q3YTT4) (Sandstrom et al., 2019; Suzuki et al., 2014; The UniProt Consortium, 2021). In contrast, enterocytes do not express *NLRP3* or *NLRC4* mRNA, so cannot activate the respective inflammasomes (found at The Human Protein Atlas, <http://www.proteinatlas.org>) (Uhlén et al., 2015).

*S. flexneri* interrupts NF- $\kappa$ B signalling to desensitise epithelial cells to PAMP- and inflammatory cytokine- stimulation and it is expected that these interactions will also occur in macrophages due to shared host cell signalling mechanisms. One *S. flexneri* effector, IpaH0722, prevents NF-kappa-B inhibitor alpha (NFKBIA) degradation in HeLa cells (Ashida et al., 2013). In homeostatic conditions, NFKBIA binds to the nuclear localisation signal of NF- $\kappa$ B, preventing it from translocating to the nucleus and stimulating the transcription of inflammatory cytokines. NFKBIA is degraded in response to *S. flexneri*  $\Delta$ IpaH0722 (IpaH0722 KO) infection within 1 hr 20 minutes (Ashida et al., 2013), indicating that *S. flexneri* (and perhaps *S. sonnei*) have sufficient time modulate the macrophage environment prior to stimulating host cell death (Kayagaki et al., 2015) (and data not shown). Another *S. flexneri* effector, OspZ, blocks TNF- $\alpha$  directed localisation of the NF- $\kappa$ B p65 subunit to the nucleus (Newton et al., 2010); thus somewhat inhibiting the responses of infected cells (e.g. synergistic IL-8 secretion) to other immune cells (Zhang et al., 2016). *S. sonnei* is also predicted to express OspZ protein (Uniprot accession ID A0A6N3N0L6) (The UniProt

Consortium, 2021). If these interactions occur during *S. sonnei* infection of macrophages, then the infected macrophages would be less responsive to some aspects of the infection.

## 1.7. The molecular tools to assess host cell gene function in the context of pathogen infection

To identify potential host factors relevant to the infection process, gene (or protein) phenotype relationships need to be interrogated. So how is a gene-function relationship identified? The three main methods are rooted in classical genetics, ‘-omics’ methodologies, and functional genetics. The first looks at the diversity of human gene variation to discover how phenotypic variation in human traits (such as disease resistance) arises, the second uses transcriptomic datasets to identify genes that respond to a perturbation (such as infection) in a biological system; and the third involves mutating genes and comparing the resulting mutants to the wild type state under conditions of interest (Marchetti et al., 2012). Functional genetics can be further partitioned into *forward* and *reverse* approaches; the former beginning with a phenotype and later identifying the underlying genetic differences, the latter perturbing genes and identifying those that cause the required phenotype. A good example of reverse genetics on a large scale is the mouse phenotyping consortium (<https://www.mousephenotype.org/>) (Dickinson et al., 2016).

Forward genetic, functional screening is the method employed throughout this project. In essence, we are creating a spectrum of gene mutations and using their effect on cell resistance or susceptibility to infection as the phenotype under examination. Gene mutations isolated in this way provide new information on the processes and pathways that are employed by host cells as defences or, conversely, hijacked by pathogens as survival and reproductive strategies. Note that a key advantage of these ‘functional’ screens is that there is a clear cause-effect link: transcriptomic approaches suffer from a lack of clarity in this regard because observed gene expression changes could be directly related to pathogen infection, indirectly related to infection, or even long-term homeostatic response mechanisms.

## 1.8. Creation of mutant cell line libraries for screening purposes – potential tools

There are several (physical, chemical, and molecular) methods for creating mutant or knockdown cell line libraries. Recent popular methods of choice are gene trap mutagenesis, ribonucleic acid (RNA) guided endonucleases (a.k.a. CRISPR-Cas9), and RNA interference (RNAi).

In the CRISPR-Cas9 system (schematic in Figure 1.3), the Cas9 enzyme creates a double-stranded break in the DNA at a designated site defined by a neighbouring protospacer adjacent motif (PAM) as well as by a guide RNA sequence complementary to the target DNA. The guide RNA binds to the target site through homologous DNA base pairing. This double-stranded break will either be repaired by the cell by homologous recombination if a

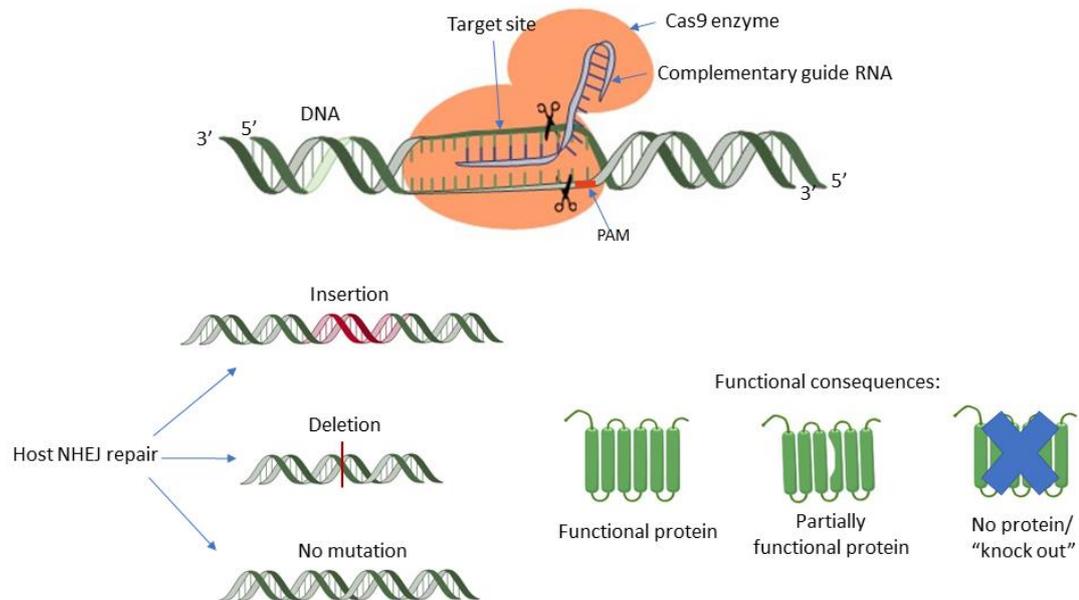
'repair' sequence matching both ends is available (e.g. provided by the researcher) or by non-homologous end joining (NHEJ). NHEJ typically creates an imperfect repair, thus generating the required gene-damaging mutations. By creating a 'library' of guide RNAs, it becomes possible to create a library of functionally mutated cells – each containing a different CRISPR-damaged gene.

Gene trap (GT) mutagenesis can be designed to occur in multiple different ways, but the focus here is on polyadenylation-enhancer trap mutagenesis (schematic in Figure 1.4), as described by Tsakiridis et al., (2009). In this case, the transfected cassette randomly inserts itself into the genome. If it falls within an intron of a gene, it disrupts that gene (heterozygously) while also splicing a neomycin resistance (Neo<sup>R</sup>) gene to the subsequent exons of the 'trapped' target gene. That splicing event not only allows productively mutated cells to be selected by antibiotic treatment (creating a highly efficient library to work with), but also generates fusion messenger RNA (mRNA) transcripts, which can be used to identify the trapped gene as required. This method obviously restricts potential gene targets to those which contain introns, though only about 15% of human genes are without (Sakharkar et al., 2004; Willyard, 2018).

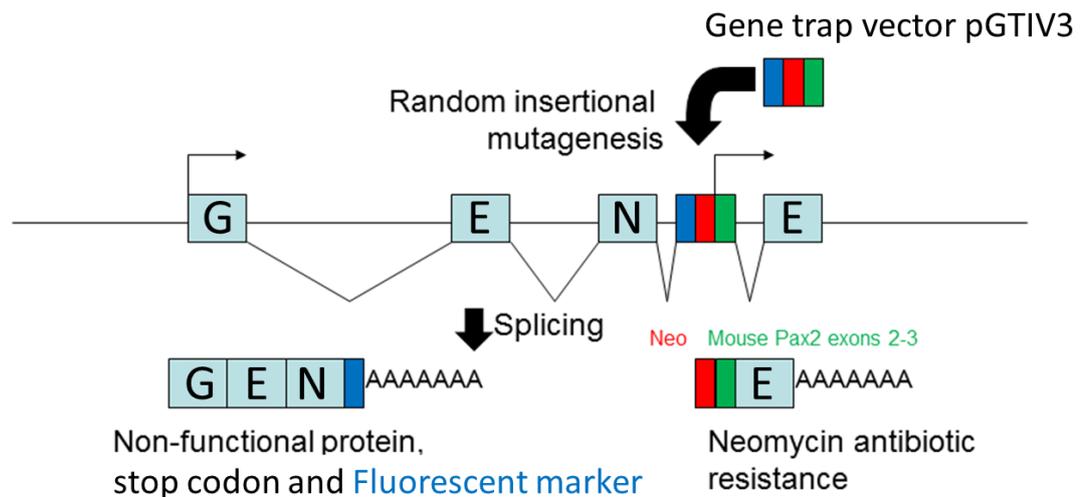
RNAi is the post-transcriptional inhibition of gene expression through sequence-specific recognition of target mRNAs which are then degraded. RNAi is an endogenous cellular mechanism to modulate gene expression – see Saliminejad et al., (2019) for review, as well as an experimental tool to generate transient (Elbashir et al., 2001) or long-term knockdown of genes (McManus et al., 2002; Paddison et al., 2002). Double-stranded short interfering RNAs (siRNAs) mediate the sequence-specific targeting of the RNA-induced silencing complex to the matching protein coding RNA (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). RNAi gene knockdown has a high rate of sequence-specific off-target effects (which can be more dominant than the on-target effect) (Franceschini et al., 2014; Singh et al., 2015) and variable efficacy (Booker et al., 2011). To compensate for sequence-specific off-target effects in high throughput screens, multiple siRNAs (as many as 25) per target gene must be employed (Kampmann et al., 2015; Morgens et al., 2016) in combination with careful experimental design and post screen analysis (Franceschini et al., 2014; Kampmann et al., 2015). The need for a massively complex synthetic siRNA library reduces the cost-effectiveness of the approach compared to the other available techniques, therefore, RNAi will not be discussed further.

Both CRISPR-Cas9 and gene trap mutagenesis screens allow relatively easy (if time-consuming) subsequent identification of target genes through amplification and sequencing of the inserted plasmid (and the surrounding region; gene-trapping) from selected cells of interest. Use of the CRISPR/Cas9 system requires construction of a plasmid library and an equivalent lentiviral library to transfect the guide RNAs (and the Cas9 enzyme) into the cells (Shalem et al., 2014; T. Wang et al., 2014; Zhou et al., 2014), whereas gene trapping requires one plasmid, which randomly inserts into introns in the genome. The gene trapping mutational event occurs so infrequently that the vast majority of mutated cells only contain one mutation on one chromosome (heterozygosity). The Cas9 enzyme in theory will continue to cleave the DNA target site until it does not recognise any further sequences; in practise, zero, one or two (in diploid cells) copies of a target gene may become mutated (T. Wang et al., 2014; Zhou et al., 2014). Complete knockouts (KOs) are

often thought of as the ideal experimental model, as they will give the strongest phenotype. However, the majority of human diseases with a genetic component are the result of a complex interaction between numerous gene variants, with one faulty or completely non-functional copy of a given gene. Potentially, this makes a heterozygous mutant library more physiologically/pathologically relevant. Additionally, complete KOs of essential genes are lethal, completely removing these from the possible pool of mutants to be screened. Naturally, no one wants a drug that mimics the effect of a lethal gene KO, but interrogation of partial KOs of essential genes may lead to associated genes or gene products whose function can be safely modulated for therapeutic purposes. These advantages make gene trap mutagenesis a valuable tool for genomic interrogation.



**Figure 1.3 CRISPR- Cas9 causes double-stranded breaks in target DNA using a complementary guide RNA, which can be repaired by host non-homologous end joining (NHEJ), thus generating mutations.** *The Cas9 enzyme forms a ribonuclease complex with a guide RNA sequence that is complementary to the DNA target site adjacent to a protospacer adjacent motif (PAM) sequence. The guide RNA directs the endonuclease activity of the enzyme, which creates a double-stranded break within the target site. The host repair machinery then repairs the break through multiple possible mechanisms, including non-homologous end joining repair, which can cause insertions, deletions, or a complete repair (no mutation). All DNA repair outcomes described can generate functional protein, but any loss of functionality or functional knockout requires some alteration in the DNA sequence.*



**Figure 1.4 Polyadenylation-enhancer trap mutagenesis.** Trapped genes are identified through Rapid amplification of cDNA ends (RACE)-Polymerase chain reaction (PCR) using primers complementary to the gene trap vector e.g. (Mouse Pax2 exon) preceding the rest of the gene (E). The GTIV3 gene trap (GT) vector randomly inserts into the genome. If the insertion is within an intron, two mRNA transcripts are made from the trapped gene; the first formed from endogenous exons before the GT vector and part of GT vector itself, which encodes a fluorescent marker. The second transcript is initiated from a promoter within the GT vector, includes the second part of the GT vector (encoding a neomycin antibiotic resistance marker) as well as subsequent exon(s) (E).

## 1.9. Other infection-based genetic screens

Genetic screens can be used to interrogate the host-pathogen interface in a variety of ways; including the use of host-like environments in single organism (axenic) bacterial culture, infection of mammalian cell culture, exposure of mammalian cell culture to a bacterial effector, individual stimulation of PRRs and *in vivo* infection models. Mammalian genetic screens have identified regulators of key host response genes (Covarrubias et al., 2017), host genes interacting with bacterial effectors (Radin et al., 2014), host genes affecting cellular survival (Jeng et al., 2019), host genes affecting bacterial entry (Yeung et al., 2019), as well as host viral susceptibility and resistance genes *in vivo* (LaFleur et al., 2019). Bacterial genetic screens have identified bacterial fitness genes to survive inside macrophages (Llewellyn et al., 2011) or inside the host organism (Karlinsey et al., 2019) as well as bacterial response genes to acidified nitrite (Kim et al., 2003).

## 1.10. Previous work carried out as part of a Centre for Defence Enterprise (CDE) funded project identified a number of candidate host resistance genes

A gene trap library in HEK-293 cells was previously screened with infection parameters designed to kill the majority of wild type cells to find genes associated with broad range of host responses to intracellular infection. In order to find gene-host response associations that were applicable to a broad range of pathogens three different species of bacteria (STM, *Shigella sonnei*, *Listeria monocytogenes*) and one parasitic apicomplexan (*Toxoplasma gondii*) were chosen from the NIAID Category B Biodefense and Emerging pathogens list (NIH: National Institute of Allergy and Infectious Diseases, 2016) for screening purposes. A summary of infection and epidemiological characteristics of the diseases these organisms cause is presented in Table 1.2. These species were chosen as representatives of the wider array of intracellular human pathogens commonly encountered, or those with the potential for bioterrorist threat risk. Many genes were identified in the library screening; some have no known function, others could be readily identified through bioinformatics analysis, or had already been briefly characterised (results of screens summarised in Table 1.4). Based on the identities, functions and known interactions of these proteins, a potential innate response network was postulated (summarised in Figure 1.5), using bioinformatics and literature searches (listed in Table 1.5).

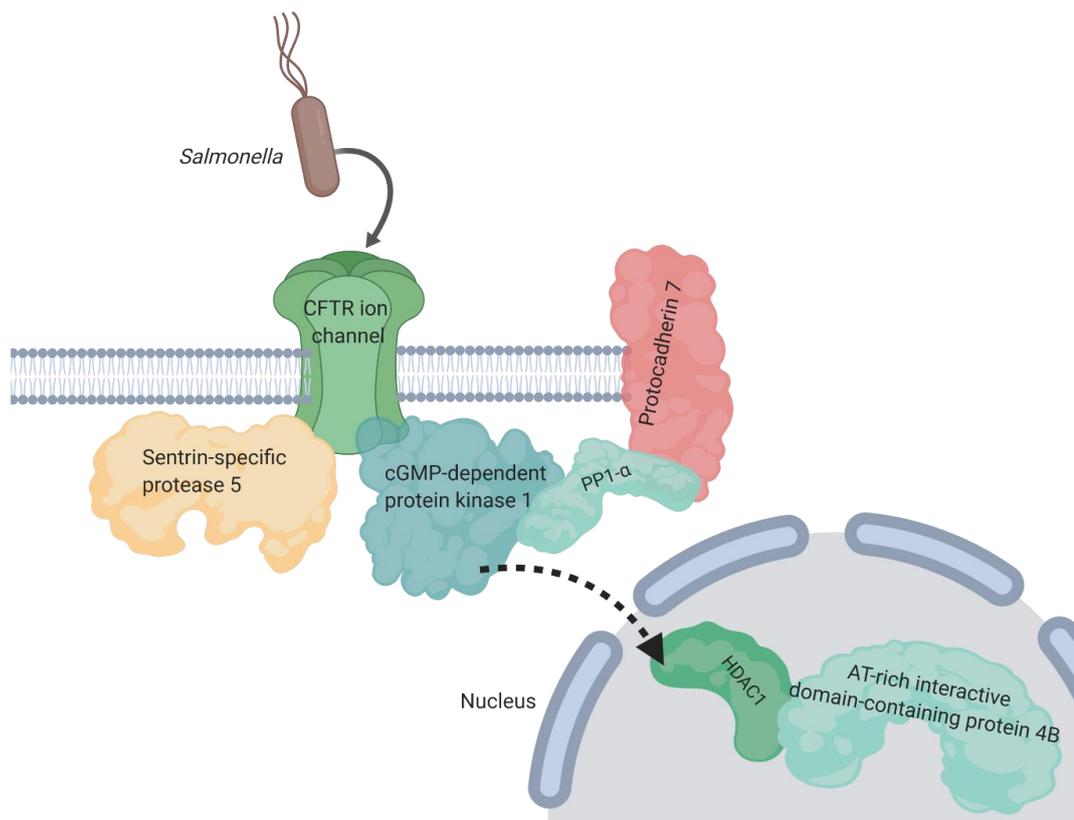
In addition to the gene trap screens, early validation work was performed in HEK-293 cells treated with CRISPR/Cas9 linear plasmids containing single guide RNA (sgRNA) sequences specific to some of the identified genes. The resulting mutants were verified by Sanger sequencing and phenotyped with STM, *Shigella sonnei* and *L. monocytogenes*. All three HEK-293 CRISPR mutants (targeting *SEN5*, *ARID4B* and *PCDH7* genes) tested, showed increased susceptibility to bacterial infection at 2-hours post addition of gentamicin (Table 1.4, #4).

**Table 1.4 Summary of infection screens and infections in CRISPR mutants done as part of the CDE project.** Screens #1, and #2 were performed by Dr Rasha Y. Mahmoud, Screen #3 was performed by Dr Stuart Woods, Screen #4 was performed by Dr Rasha Y. Mahmoud and Dr Wenqin Li. HEK-293GT#1 and #2 refers to two HEK-293 gene trap libraries which contain ~50,000 and ~250,000 mutational events, respectively. In this table *Salmonella*, *Listeria* and *Shigella* refer to *Salmonella enterica serotype Typhimurium strain SL1344*, *L. monocytogenes* and *Shigella sonnei*, respectively. TG3, 4 and 8 refer to 'clones' isolated from a HEK-293 Gene trap library. RACE-PCR refers to Rapid amplification of cDNA ends - Polymerase chain reaction.

Infection screens and CRISPR work	Cells used	Protocol	End-point	Result
#1: Survival screen	HEK-293 Gene trap library	Serially infected with STM 3x to select for survival	RACE-PCR and DNA sequencing from surviving host cells to identify genes	mutant cells containing mutations in <i>SENP5</i> , <i>ARID4B</i> , <i>PCDH7</i> or <i>PSMA1</i> genes isolated from this screen
#2: Fluorescence activated cell sorting (FACS) screen	HEK-293 GT#1 and HEK-293 GT#2	Cells infected with GFP-expressing STM and serially sorted 3x	Sorted into low, medium and high GFP-expression (proxy for <i>Salmonella</i> burden)	mutant cells containing mutations in the <i>PRKG1</i> gene isolated from this screen
#3: <i>T. gondii</i> growth assay	Isolated cells from HEK-293 Gene trap library	Isolated cells infected with luciferase expressing <i>T. gondii</i>	Bioluminescent total flux as a proxy for total parasite burden.	TG3 (intergenic insert), TG4 ( <i>PCDH7/LINC01170</i> ) and TG8 ( <i>ARID4B/SENP5</i> ) had significantly ( $P < 0.001$ ) lower parasite growth

Infection screens and CRISPR work	Cells used	Protocol	End-point	Result
#4: CRISPR mutant bacterial infections	Cas9/sgRNA targeted SENP5 HEK-293 mutant and CRISPR/Cas9 targeted ARID4B and PCDH7 HEK-293 mutants (GeneCopoeia)	Cells infected with STM, <i>Listeria</i> or <i>Shigella</i>	CFU counts of intracellular bacteria 4 hours post inoculation	More bacterial growth in the SENP5 mutant cell line for each species

## 1.10.1. Gene trap identified genes



**Figure 1.5** Postulated host response network based on positive hits from screens #1 and #2 and relationships identified in the literature. *SEN5*, *PCDH7* and *ARID4B* were positive hits in gene trap screen #1 and *PRKG1* came out of screen #2. Sentrin-specific protease 5 may be involved in deconjugation of Small Ubiquitin-Like Modifier-2/3 from Cystic fibrosis transmembrane conductance regulator (CFTR) (see related text) (Ahner et al., 2013) and cGMP-dependent protein kinase 1 beta (*PRKG1*) is able to phosphorylate CFTR in vitro (French et al., 1995). A potential interaction between cGK1 and PP-1A has been identified on STRING <https://string-db.org/> (Szkarczyk et al., 2017) and protocadherin 7 (protein product of *PCDH7*) inhibits Serine/threonine-protein phosphatase PP1-alpha catalytic subunit (protein product of *PPP1CA*) phosphatase activity of some substrates in vitro (Yoshida et al., 1999). cGK1 can phosphorylate isolated HDAC1 in vitro (Tsai and Seto, 2002) and AT-rich interactive domain-containing protein 4B (*ARID4B*) interacts with HDAC1 as part of the *mSIN3A* corepressor complex (in mice) as well as the human orthologue in affinity-capture studies (Hein et al., 2015; Huttlin et al., 2017; Joshi et al., 2013). HDAC1 and *ARID4B* are localised to the nucleus. Additionally, CFTR is suggested as receptor for *Salmonella enterica* serotype Typhi strain Ty2 based on in vitro and in vivo infection studies (Pier et al., 1998). This graphic includes a “schematic representation of cystic fibrosis transmembrane conductance regulator protein” by Lucarelli (2017).

**Table 1.5 Genes of Interest identified from HEK-293 gene trap library intracellular infection screen and now the focus of this PhD.** RNA expression and cellular location data (except for *CFTR* and *SEN5*) obtained from *The Human Protein Atlas* (2017). Cellular location for *CFTR* found in (Ameen and Apodaca, 2007; Milewski et al., 2001; Morris et al., 1994). Protein identities obtained from Uniprot <https://www.uniprot.org/> (The Uniprot Consortium, 2015).

Gene of Interest	Protein encoded	Cellular location
<i>CFTR</i>	Cystic fibrosis transmembrane conductance regulator	Apical plasma membrane
<i>ARID4B</i>	AT-rich interactive domain-containing protein 4B	Nucleoplasm and vesicles
<i>PCDH7</i>	Protocadherin-7	Plasma membrane and cell junctions
<i>PRKG1</i>	cGMP-dependent protein kinase 1	Vesicles and cytosol
<i>PSMA1</i>	Proteasome subunit alpha type-1	Nucleoplasm
<i>SEN5</i>	Sentrin-specific protease 5	See Section 1.10.1.6 text

#### 1.10.1.1. *CFTR*

Cystic fibrosis transmembrane conductance regulator (*CFTR*) is described here as it is a component of the proposed innate host response network (schematic in Figure 1.5) based on links in the literature to the genes identified in gene trap infection screens (**Table 1.5**). *CFTR* is an Adenosine Triphosphate (ATP)-dependent ion channel that allows the efflux of chloride and bicarbonate ions out of the apical side of the cell (Callebaut et al., 2017; Morris et al., 1994). Transcripts of *CFTR* are mainly found in epithelial cells, e.g. enterocytes of the colon and Paneth (epithelial) cells in the small intestine (found in *The Human Protein Atlas*, [www.proteinatlas.org](http://www.proteinatlas.org)) (Uhlén et al., 2015). The movement of chloride ions from the basolateral to the apical surface of the epithelium creates an osmotic gradient, promoting the movement of water in the same direction. Hence, *CFTR* has important roles in processes requiring effective regulation of viscosity; for example, mucociliary clearance in the respiratory tract (Tarran et al., 2005; Zhang et al., 2009) and effective production of pancreatic juices containing digestive enzymes (Park et al., 2010; Uc et al., 2012). Its mutation is the cause of the recessive, life-limiting condition, cystic fibrosis.

*CFTR* KO or non-functional  $\Delta F508$  *CFTR* phenotypes dramatically reduce *Salmonella enterica* serotype Typhi entry into primary human epithelial cells or cell lines *in vitro* as well as

reducing its entry in the gastrointestinal submucosa *in vivo* (Pier et al., 1998). The  $\Delta F508$  mutation is non-functional, and mostly remains in the endoplasmic reticulum before proteasomal degradation (Cheng et al., 1990; Du et al., 2005; Jensen et al., 1995; Pankow et al., 2019). Furthermore, cystic fibrosis genotypes correlated well with a lack of typhoid in typhoid endemic Indonesia (Vosse et al., 2005). This effect may be host-dependent, as CFTR associated entry of *S. enterica* ser. Typhi could not be replicated in a baby hamster kidney cell line (Bravo et al., 2011). Two other voltage-gated ion channel genes *KCNH5* and *KCNH7* in humans were associated with *Vibrio cholerae* infection in a cholera endemic area of Bangladesh (Karlsson et al., 2013).

CFTR KO and  $\Delta F508$  CFTR mutant cells are under increased oxidative stress due to uninterrupted tissue transglutaminase activation compared to cells with wild type CFTR (Luciani et al., 2009). CFTR has also been linked to the negative regulation of NF- $\kappa$ B (Vij et al., 2009). It would be interesting to see if the oxidative stress found in non-functional CFTR mutant cells has an impact on bacterial infection.

#### 1.10.1.2. *ARID4B*

The mouse orthologue of AT-rich interactive domain-containing protein 4B (*ARID4B*) was identified as a subunit of the mSIN3A corepressor complex, which has histone deacetylase (HDAC) activity (Fleischer et al., 2003). *ARID4B* interacts with HDAC1 as part of the mSIN3A corepressor complex (in mice) as well as the human orthologue in affinity-capture studies (Hein et al., 2015; Huttlin et al., 2017; Joshi et al., 2013). HDACs are enzymes that deacetylate histones, originally thought to allow histone methylation, subsequent condensation of the local chromatin and transcriptional repression of genes to occur. Now the picture is more nuanced, with HDACs (as part of large complexes like the SIN3A corepressor complex) being targeted to transcriptionally active as well as repressed regions of DNA (Kelly and Cowley, 2013). Other domains of *ARID4B* besides the one interacting with SIN3A corepressor complex have been characterised, one ( $R1\sigma$ ) with basal and active SUMOylation dependent transcriptional repression activity and another ( $R1\alpha$ ) with only basal transcriptional repression activity (Binda et al., 2006).

#### 1.10.1.3. *PCDH7*

*PCDH7* encodes the protein protocadherin-7, of which there are three splice variants 7a, 7b and 7c (S.-Y. Kim et al., 2011). Protocadherin 7 is a member of the cadherin family, which mediate cell adhesion. Indeed splice variants 7a and b demonstrate significant adhesion activity in transfected cells, particularly in the presence of calcium (Yoshida, 2003). E-Cadherin on the host cell surface is required for entry of *L. monocytogenes* into certain cell types (Sousa et al., 2007). Interestingly, siRNA knockdown of *PCDH7* in BGC-823 and MKN-45 (cell lines derived from Gastric Carcinomas) cells significantly reduced the expression of *E-Cadherin* (Chen et al., 2017). The presence of global changes in gene expression during cell carcinogenesis is well known (Sharma et al., 2010) and it remains to be seen whether the connection between *PCDH7* and *E-Cadherin* is relevant in the context of non-cancerous cells. Protocadherin-7 may have more direct functions in intracellular infection. It was upregulated in a DNA microarray screen using mouse embryonic fibroblast KO for Inhibitor of kappa B ( $I\kappa$ B) kinase  $\alpha$  ( $IKK\alpha$ ),  $\beta$  ( $IKK\beta$ ) or  $\gamma$  ( $IKK\gamma$ ) using  $TNF\alpha$  as the stimulus (Li et al., 2002). Protocadherin 7 associates with Serine/threonine-protein phosphatase PP1-alpha

catalytic subunit (PP-1A) (gene name: PPP1CA) in mammalian cells and inhibits PP-1A phosphatase activity of some substrates *in vitro* (Yoshida et al., 1999).

#### 1.10.1.4. *PRKG1*

The cGMP-dependent protein kinase 1 (cGK1) types 1 $\alpha$  and 1 $\beta$ , encoded by *PRKG1*, are soluble serine/threonine kinases that only differ in their N-terminal regions (Casteel et al., 2002; Ørstavik et al., 1997). Roles for cGK1 have been identified in sodium handling in the kidneys (Citterio et al., 2013), signalling in smooth muscle (Sawada et al., 2001; Surks et al., 1999; Wooldridge et al., 2004), inhibition of platelet aggregation (Antl et al., 2007) and upregulation of gene expression (Casteel et al., 2002; Gudi et al., 1997; Zhao et al., 2005). cGK1 $\beta$  is able to phosphorylate CFTR *in vitro* (French et al., 1995), though only cGK2 was found to phosphorylate CFTR in cells (Vaandrager et al., 1998, 1997). It was found that the co-localisation of cGK2 (with CFTR) to the membrane was a factor in the improved activity of cGK2 in cells. cGK1 can phosphorylate isolated HDAC1 *in vitro* (Tsai and Seto, 2002), indirectly connecting cGK1 with ARID4B. A potential interaction between cGK1 and PP-1A has been identified on Search Tool for Recurring Instances of Neighbouring Genes (STRING), <https://string-db.org/> (Szklarczyk et al., 2017).

#### 1.10.1.5. *PSMA1*

The proteasome subunit alpha type-1 (PSMA1) is a 20S core proteasome subunit (Fu et al., 1998). Proteasomal degradation is an integral part of proteostasis; the mechanism of generating and maintaining correct protein folding and the associated degradation of misfolded proteins (Díaz-Villanueva et al., 2015; Kaushik and Cuervo, 2015). Proteins can be sent for proteasomal degradation at many different points during their lifetime; from nascent polypeptide production, endoplasmic-reticulum (ER)-associated misfolded membrane or secretory proteins, mislocalised cytosolic hydrophobic proteins, to misfolding proteins due to cellular stress (Casson et al., 2016; Defenouillère and Fromont-Racine, 2017; Lindholm et al., 2017; Preston and Brodsky, 2017).

An infection-associated protein interaction network involving PSMA1 was identified in rainbow trout, through transcriptome and proteome sequencing and subsequent bioinformatics analysis (Long et al., 2015). This network also included multiple other proteasome subunits, strongly indicating that proteasome expression and hence activity is modified by the infection response. An older study found that LPS both interacted with and increased the activity of the proteasome (specifically subunits PSMA1 and Proteasome subunit beta type-4 (PSMB4)) (Qureshi et al., 2003).

#### 1.10.1.6. *SENPS5*

SUMOylation, a ubiquitin-like process of protein modification, has roles in meiosis (Nottke et al., 2017), mitosis (Mukhopadhyay and Dasso, 2017), regulation of metabolism (Kamynina and Stover, 2017), maintenance of genomic stability (Zilio et al., 2017), senescence (Princz and Tavernarakis, 2017) and development (Monribot-Villanueva et al., 2017); as well as in a multitude of other processes. To SUMOylate a protein, the mature form of a Small Ubiquitin-like Modifier (SUMO) is activated by an E1 activation enzyme, conjugated to a target protein by an E2 conjugation enzyme and in many cases, conjugation is stabilised by an E3 ligation enzyme (Pichler et al., 2017). Sentrin-specific

protease(s) (SENP)s turn SUMO proteins into their mature forms and deconjugate(s) them from their protein partners. SENP5 performs this function for SUMO-2 and -3 (Di Bacco et al., 2006; Di Bacco and Gill, 2006; Gong and Yeh, 2006). SENP5 cellular localisation is cell state dependent and can be found in the nucleolus (Di Bacco et al., 2006; Yun et al., 2008), the nucleus (Gong and Yeh, 2006), cytosol (Cheng et al., 2015) and in mitochondria (Zunino et al., 2009).

CFTR is SUMOylated with SUMO-1 as well as SUMO-2/3, which cannot be distinguished immunologically (Ahner et al., 2013). SUMOylation with SUMO-2/3 marks  $\Delta$ F508 CFTR for ubiquitination and subsequent proteolysis. Wild type CFTR also appears to be targeted by the same SUMOylation-mediated degradation pathway, as protein levels of CFTR increased in response to targeted reductions of  $\Delta$ F508 CFTR-SUMOylation. A large proportion of wild type CFTR protein is degraded before maturation, due to difficulty in folding. Cells co-transfected with CFTR, Ubc-9 (SUMO E2 conjugating enzyme) and SENP1 resulted in higher protein levels of CFTR than in CFTR and Ubc-9 transfected cells (Ahner et al., 2013). SENP1 can deconjugate SUMO-1, 2 and 3 from modified proteins, therefore SENP1 appears to (partially) override the signal for CFTR degradation. Thus, knocking out SENP5 would lead to lower CFTR protein levels by the same mechanism.

Interestingly, SUMOylation is itself associated with infection. Overexpression of host SUMO-2 (leading to increased levels of free SUMO-2 and likely increased levels of SUMOylation) decreased *Shigella flexneri* invasion and *L. monocytogenes* infection *in vitro* (Fritah et al., 2014; Ribet et al., 2012). Additionally, depletion of SUMO E1 activating enzyme 2 by RNA silencing caused an increase in *S. flexneri* invasion, indicating that *S. flexneri* invasion is inversely correlated to protein SUMOylation. Furthermore, both *L. monocytogenes* and *S. flexneri* reduce global protein SUMOylation in host cells *in vitro* by destabilising a key E3 ligation enzyme (Ribet et al., 2012; Sidik et al., 2015). The correlation between infection resistance and protein SUMOylation provides a potential explanation for the assumed infection survival advantage seen in the gene trap HEK-293 screen for trapped *SENP5*.

## 1.11. Aims and hypotheses

The overall objective of the work described in this thesis is to identify those innate responses to intracellular pathogens or host processes essential for pathogen survival that could potentially be disrupted by future pharmacological means without detriment to the host. A treatment that targets an intracellular aspect of the infection lifecycle but is dispensable for the host could prevent or reduce pathogen multiplication and potentially promote a more effective host response. This would reduce host morbidity/mortality while the adaptive immune response develops. Additionally, we hypothesise that such a therapeutic strategy would be much less susceptible to rapid bacterial acquisition of resistance – in contrast to conventional antibiotics.

We hypothesise that *in vitro* genetic screens can identify host genes that are functionally central to the success or failure of the bacterial invasion and replication processes – and that the protein products of such genes might become the future targets for novel therapeutic strategies to combat infection *in vivo*.

A further hypothesis is that infection phenotypes observed in the HEK-293 GT survival and Fluorescence-activated cell sorting (FACS) screens can be independently corroborated using CRISPR KO of the identified genes in the same genetic background.

For the purpose of achieving the above objective, HEK-293 cells were chosen as ‘non-professional immune cells,’ and U937 cells (a human monocyte-like cell line), differentiated into macrophages, were chosen to give insight into genes that might be important in ‘professional immune cells’. Two approaches, both of which made use of gene trap libraries were used to identify putative host infection resistance genes:

- (i) Firstly, prior to the commencement of this PhD, a gene trap library made in HEK-293 was used. This identified a number of putative genes that affect the growth and survival of intracellular bacteria by selecting for cells with disrupted genes that conferred host survival advantages or that conferred resistance to bacterial growth. Part of the work described in this thesis is a further investigation of these genes by construction of CRISPR/Cas9 mutated cells.
- (ii) Secondly, a gene trap library was constructed in U937 cells, which were then differentiated into macrophages before being selected for host survival during intracellular bacterial infection independently with *S. enterica* serotype Typhimurium (STM) SL1344 and *F. tularensis* subsp. *holarctica* Live Vaccine Strain (LVS). The differential overrepresented gene output from each screen was independently assessed for overrepresented gene ontology terms and shared protein interactions between screens were identified. One significantly overrepresented candidate gene, *SLC7A11* from the first U937 screen was independently assessed for its association with host infection resistance. The experimental work performed in this macrophage model as well as intended subsequent validation is summarised in Figure 1.2.

As macrophages share many intracellular immune response mechanisms with non-phagocytic cell types, one would expect several of the genes identified in the initial gene trap screens to appear in the gene trap screen with the macrophage-like cell line. In addition to those, genes idiosyncratic to macrophage (or phagocyte) specific functions are also expected.

# Chapter 2

## Materials and Methods

## 2. MATERIALS AND METHODS

### 2.1. Conditions for cell culture maintenance and archiving

HEK-293 cells (ECACC 85120602) were cultured in DMEM/F-12 with Gibco® GlutaMax™ (Life Technologies, Paisley, UK), 10% fetal calf serum (FCS) (Biosera, Kansas City, MO, USA), and 1% penicillin/streptomycin (Gibco, Paisley, UK). Cells were grown at 37°C in a humidified incubator with a 5% CO<sub>2</sub> atmosphere maintained.

HEK-293 cells were passaged every 2-3 days, or after 70% confluency was reached, by being washed and trypsinised with TrypLE™ Express (Gibco). Suspended cells were washed with media containing 10% FCS to reach a 10-fold greater volume and were centrifuged at 400 g for 5 minutes to obtain a cell pellet. The pellet was then resuspended in an appropriate volume of media and added to flasks or plates for continued growth and experimental use, respectively.

For archiving, adherent cells were frozen after reaching 100% confluency and split into 2 cryotubes per T25 flask in 500 µL freezing mix (90% FCS, 10% dimethyl sulfoxide (DMSO) (Merck & Co., Glasgow, UK)) per tube. Additionally, one half of a T25 flask was processed as above to obtain a cell pellet ready for DNA extraction.

U937 cells (ECACC 85011440) were cultured in Roswell Park Memorial Institute medium (Life Technologies, Paisley, UK) with 10% FCS, 1% L-glutamine (Life Technologies) and 1% penicillin/streptomycin (U937 media).

U937 cells were passaged by being centrifuged at 300 g for 5 minutes and resuspended in fresh media in a new flask every 2-3 days or when the media turned orange. Cells were counted before returning to the flask during this process to ensure cells were kept at an appropriate seeding concentration ( $2-9 \times 10^5$  cells/mL). U937 cells were frozen at a concentration of  $1.8 \times 10^6$  per cryotube in 500 µL of freezing mix.

### 2.2. Generation of CRISPR cell line and Gene trap library mutants

CRISPR and Gene trap (GT) library mutants were created by transfecting cells with purified linearised CRISPR plasmid construct, or pGTIV3 plasmid vector, respectively, using a Nucleofector™ 2b Device (Lonza, Slough, UK).

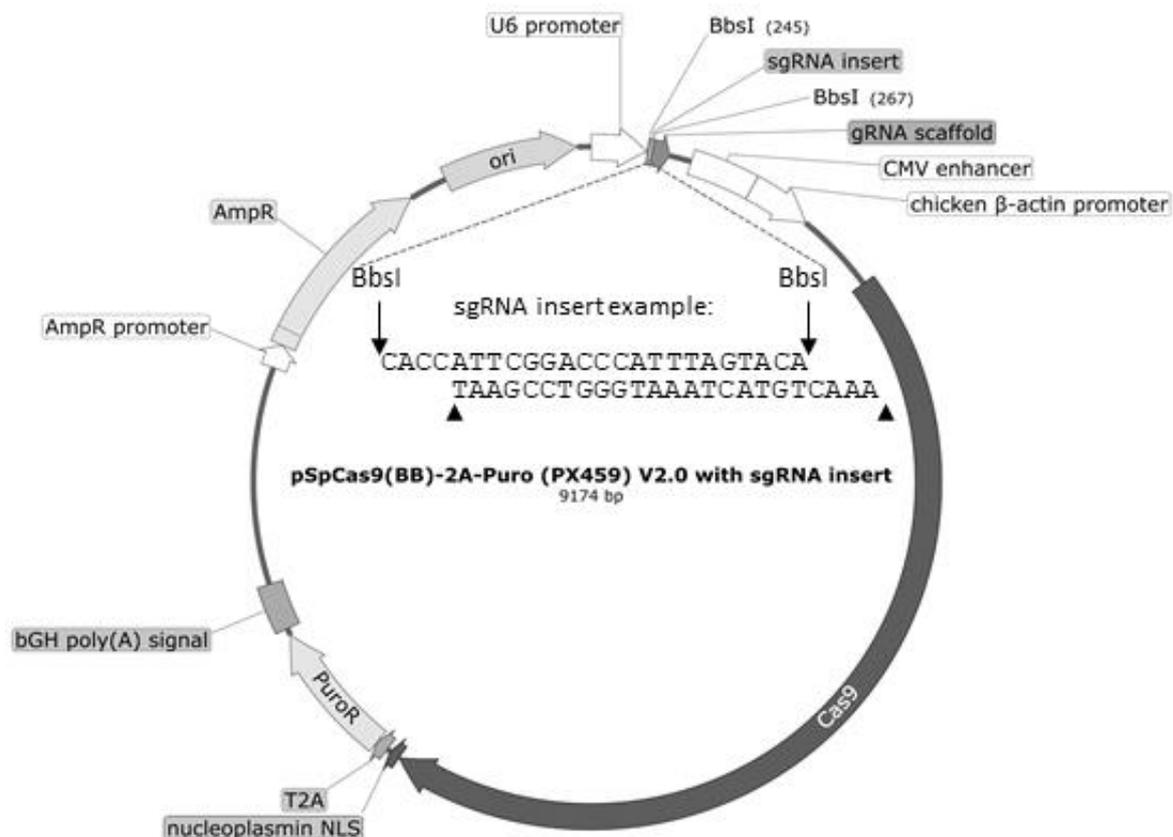
#### 2.2.1. Design or ordering of CRISPR constructs in order to knock out candidate genes

The following CRISPR plasmid constructs, purchased directly from GeneCopoeia, Rockville, Maryland, USA were used; HCP204954-CG01-1-B (*SEN5*), HCP267211-CG01-1-B (*PRKG1*), HCP200948-CG01-1-B (*CFTR*), HCP213001-CG01 (*ARID4B*), HCP212283-CG01 (*PCDH7*) and HCP262778-CG01-1-B (*PSMA1*). Restriction enzyme Sall (New England Bioscience (NEB), Hitchin, UK) was used to digest the GeneCopoeia CRISPR plasmids to create linear DNA suitable for stable transfection.

To generate U937 CRISPR KO cell lines, bespoke CRISPR constructs were created with sgRNA sequences designed in CHOPCHOP (<https://chopchop.cbu.uib.no/>) (Labun et al., 2019, 2016; Montague et al., 2014) and with additional bases added to bind to the plasmid overhangs created by

BbsI digestion of the CRISPR plasmid. A control sgRNA (dubbed 'miscullin') was designed by changing 3 bases in the *CUL1* sgRNA and checking for alternative alignments with NCBI nucleotide BLAST <http://www.ncbi.nlm.nih.gov/BLAST/> (McGinnis and Madden, 2004).

The CRISPR plasmid used in later stages was pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid #62988) (Cas9 plasmid), from the laboratory of Feng Zhang (Ran et al., 2013). Briefly, the top and bottom strand per pair of designed CRISPR sgRNA oligos (Merck & Co., Table 2.4) were mixed and boiled to melt the double stranded (ds) DNA, before cooling slowly to allow the intended top and bottom strands to anneal. The Cas9 plasmid was cut with BbsI restriction enzyme (Life Technologies) and purified with Qiaquick polymerase chain reaction (PCR) purification kit (Qiagen, Manchester, UK). The purified plasmid was ligated to ds oligo DNA using T4 DNA ligase (Promega, Southampton, UK). The ligated plasmids were transformed into chemically competent *Escherichia coli* K12 strain DH5 $\alpha$  bacteria (Thermo Fisher Scientific, Inchinnan, UK); colonies of which were grown overnight in liquid culture, 37°C shaking incubator overnight in Luria Bertani (LB) broth, Miller (Merck & Co.). Pelleted cultures were taken through standard miniprep protocols (New England Biolabs, Hitchin, UK) to extract the plasmid DNA. The extracted plasmids were sequenced (Source Bioscience, Nottingham, UK) to confirm correct sgRNA insert presence and sequenced before restriction cleavage with PvuI to linearise DNA and subsequent purification (Qiaquick PCR purification kit).



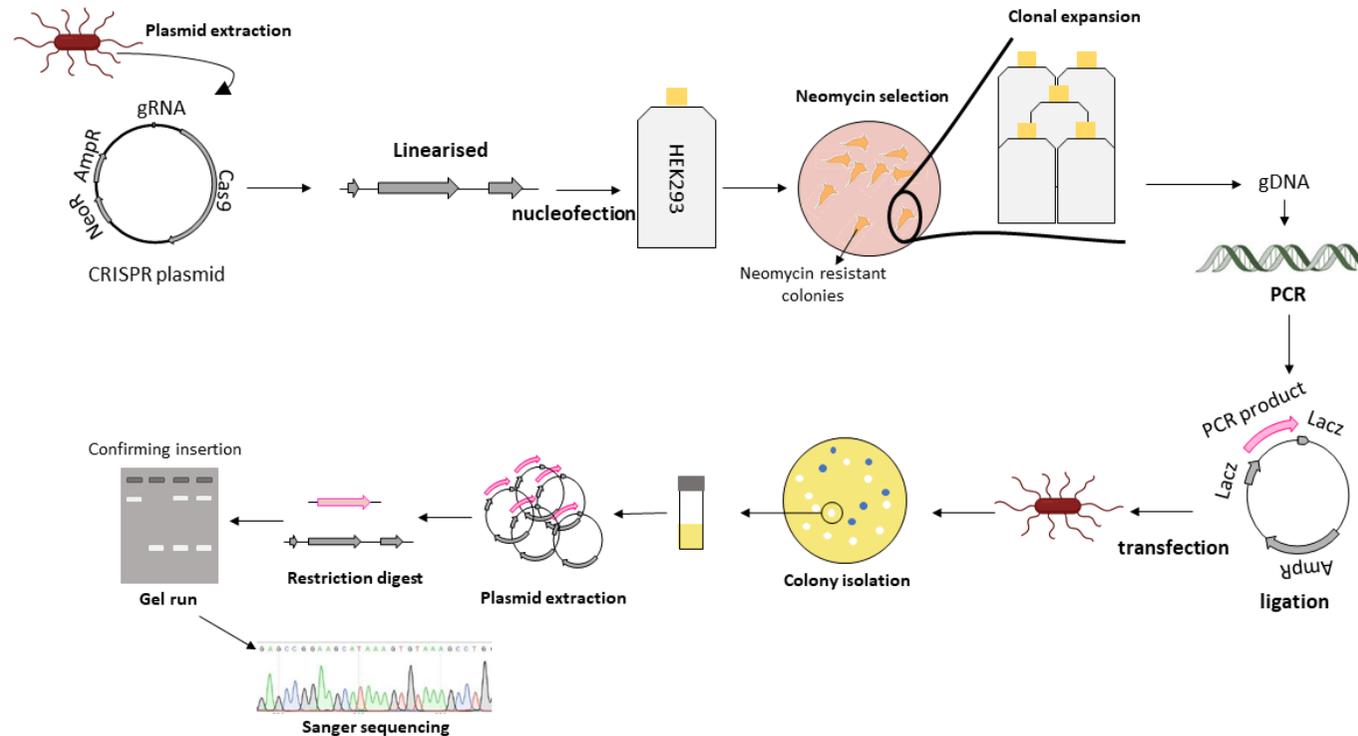
**Figure 2.1 pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid with sgRNA insert.** The plasmid contains an origin of replication (*ori*) to allow its replication within bacteria (e.g. *E. coli*), a U6 promoter to initiate transcription of an sgRNA insert connected to a gRNA scaffold in eukaryotic cells, as well as a human cytomegalovirus (CMV) immediate early enhancer and chicken  $\beta$ -actin promoter to initiate transcription of Cas9 endonuclease mRNA (derived from *Streptococcus pyogenes*). The Cas9 protein is expressed with a nucleoplasmin nuclear localisation sequence (NLS) and the translation is truncated part way through a T2A sequence designed to produce distinct proteins from a single transcript. Puromycin N-acetyltransferase (enables resistance to puromycin) is translated from the second half of the transcript – this transcript is completed with a bovine growth hormone (bGH) polyadenylation (poly(A)) signal. The plasmid also contains an ampicillin resistance (*AmpR*) promoter and gene, for retention of the plasmid in bacteria cultured with ampicillin. An example double stranded sgRNA insert sequence (*SLC7A11* top and bottom 1 sequences, **Table 2.4**) is presented to demonstrate the overhangs within the context of the *BbsI* restriction sites (each indicated by an arrow and arrow head) in the plasmid.

### 2.2.2. Transfection

Cells were left to recover overnight after linear plasmid nucleofection (programme Q-001 and solution V for HEK-293, or W-001 and solution C for U937) before undergoing selection in 500  $\mu\text{g}/\text{mL}$  (HEK-293) or 600  $\mu\text{g}/\text{mL}$  (U937) of geneticin (G418) (Roche, Welwyn Garden City, UK) for seven days or until parallel wild type control cells were completely dead. Laboratory constructed CRISPR plasmids required Puromycin selection, and in U937 cells a concentration of 0.3  $\mu\text{g}/\text{mL}$ .

To derive HEK-293 CRISPR mutant cell lines from CRISPR/Cas9-treated pooled mutant cells, the pooled population was diluted to 1,000 cells per well (in 3 mL) in a 6-well plate before serially diluting by 6-fold 5 times. Once distinct cell colonies became visible by eye, they were washed with

phosphate-buffered saline (PBS, 8 g sodium chloride, 1.16 g disodium hydrogen phosphate, 0.2 g potassium chloride, 0.2 g potassium phosphate monobasic, (all Merck & Co.) per litre of Millipore distilled water)) and picked with a 200  $\mu$ L pipette. The picked cells were transferred into prepared wells with fresh, warm media. To promote rapid growth, the clumps of cells generated by picking cells were washed with PBS and trypsinised with 500  $\mu$ L of TrypLE Express at 37°C. Washed cells were resuspended in 3 mL of media and replated – these separated cells were now able to grow faster than when constrained within a colony. CRISPR KO pools and lines targeting *SENP5*, *PRKG1*, *ARID4B*, *CFTR*, *SEMA3A* in a HEK-293 background were created and isolated by Dr Benjamin Pickard. CRISPR KO pools and lines targeting other genes in the same background were created and isolated with his help; or in the case of *ARID4B* Line K, performed by a Masters student, Robyn Maitland. The CRISPR KO *SLC7A11* Lines #1 and #2 in the U937 background were created by Masters student Jason Dick and Dr Benjamin Pickard. Furthermore, Jason Dick maintained these lines.



**Figure 2.2 Generation and genetic validation of CRISPR cell line knock outs.** *Genecopoeia*-designed plasmid extracted from *E. coli DH5α* and linearised before nucleofection into HEK-293 wild type cells. HEK-293 cells containing the plasmid with Neomycin were selected for, prior to isolation of resistant colonies for clonal expansion. The genomic DNA was extracted from isolated cell lines to confirm presence of mutation(s). The CRISPR Cas9 targeted region was amplified by PCR from the genomic DNA before ligating into the pGEM-T Easy vector. The plasmid containing the PCR was transfected into chemically competent *E. coli DH5α* and grown on LB agar containing IPTG, X-gal and ampicillin for bacterial colony isolation. Multiple white colonies (containing plasmids with inserts) were grown up to extract the plasmids, which were digested to check the insert size on a gel. Plasmids containing inserts (especially those with obvious deletions or insertions) were sent for Sanger sequencing in order to identify mutations.

### 2.2.3. Gene trap library construction

A high-complexity U937 gene trap library was created by pooling 2 separate sub-libraries. Each of these was constructed by five parallel nucleofections of linearised gene trap vector, pGTIV3, into 10<sup>6</sup> U937 cells, as above with the help of Dr Benjamin Pickard. These parallel sets of cells were pooled and grown in geneticin antibiotic-containing media (see above) to select for productive gene mutations. Serial dilutions of an aliquot of the pooled transfected cells were created and also placed under selection. By observing the lowest concentration of diluted cells which were still able to generate surviving cells, we were able to estimate mutation generation at the scale of 1,000s per sub-library.

### 2.3. Genomic DNA extraction from mutated cells

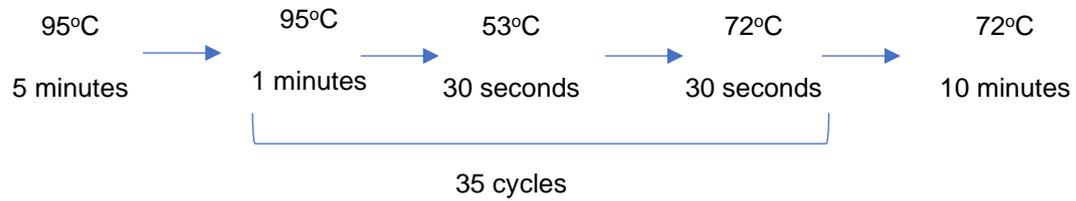
Genomic DNA was extracted from cell pellets (see above) using a Genomic DNA extraction kit (Bioline, London, UK) with the optional overnight proteinase K pre-lysis step. DNA was eluted from the columns in two 25 µL steps with preheated 70°C elution buffer (included in the kit). DNA was stored and aliquoted at -20°C until required.

### 2.4. Characterisation of CRISPR gene mutations using PCR, cloning and sequence analysis

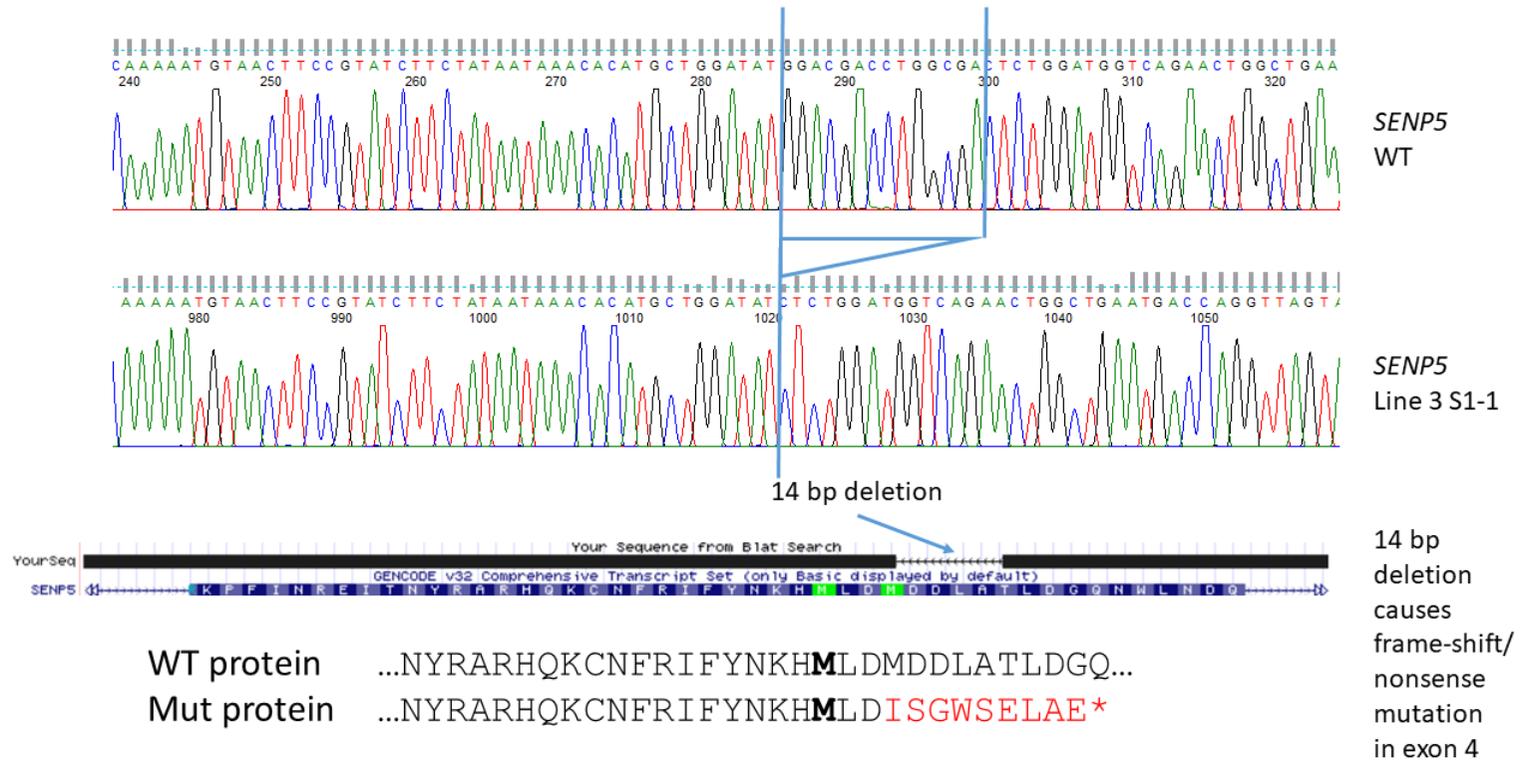
PCR was performed on a GeneAmp® PCR System 2400 thermocycler (Waltham, Massachusetts, USA) using MyTaq™ DNA Polymerase (Bioline) and Custom DNA oligos (primers, see Table 2.4) (Merck & Co.). The sequence to be amplified surrounded the locus targeted by the sgRNA within each CRISPR construct. PCR products were loaded onto a 1% agarose (Severn Biotech Ltd, Kidderminster, UK) gel (with Tris-Borate-Ethylenediaminetetraacetic Acid (EDTA) (89 mM Tris (pH 7.6), 89 mM boric acid, 2 mM EDTA) 1% ethidium bromide (all Merck & Co.)) which was run for 1.25 hours at 70 V to check product size and purity. Products were then purified directly, or from an agarose gel, using a PCR and gel purification kit (Bioline). Purified PCR products were ligated to pGEM® T Easy vectors with rapid ligation buffer and T7 ligase (all Promega) and the resulting plasmids were transformed into chemically competent *E. coli* K12 strain DH5α bacteria (Thermo Fisher Scientific, Inchinnan, UK). The transformed bacteria were streaked onto LB Lennox agar (Merck & Co.) plates with 100 µg/mL of ampicillin (Merck & Co.) with 50 µg/mL X-gal (Promega) and 20 µg/mL IPTG (Merck & Co.) spread over the surface. White (indicating successful insertion) colonies (and one blue colony for use as a negative digestion control) were picked, inoculated overnight in LB broth and then plasmid DNA was extracted with a miniprep plasmid extraction kit (Bioline), digested with EcoRI (NEB) and the digested products were run on a 1% agarose gel (as above). Plasmids containing inserts of the correct size (as determined by the digestion) from CRISPR treated cell lines (as well as two plasmids containing inserts from wild type cells) were sent off for sequencing (Eurofins, Edinburgh, UK). Six plasmids containing inserts from the CRISPR targeted locus were sequenced per fully characterised line.

Parental line sequences were compared with the human reference genome (December 2013, GRCh38/hg38 assembly) on the University of California Santa Cruz (UCSC) genome browser website <http://genome.ucsc.edu> (Kent et al., 2002) by entering the sequence into the website's Blat tool (Kent, 2002). Mutant sequences that passed the Eurofins quality check were entered in the Blat search to identify CRISPR-induced mutations. Any sequences containing potential deletions or insertions in the exon coding sequence (excluding any small nucleotide polymorphisms (SNPs) due to

myTaq polymerase inaccuracies or common population polymorphisms) were translated into the appropriate protein coding frame using the EMBL-EBI EMBOSS-Transeq tool <https://www.ebi.ac.uk/> (Li et al., 2015). Those protein sequences were then compared to the HEK-293 wild type protein sequence to predict the functional consequences/severity of the introduced changes.



**Figure 2.3 Schematic of the PCR programme used to amplify CRISPR target loci.**



**Figure 2.4** Process to identify the genetic mutation and predicted consequences to the amino acid sequence of a CRISPR KO cell line. The DNA sequences from the target locus are compared between the parent HEK-293 cell line (top electropherogram), the CRISPR UCSC genome browser. The UCSC genome browser also provides visual representation of the mutated region (YourSeq, thick black line) and a reference amino acid sequence (in blue). The reference amino acid sequence (wild type protein) is compared to the predicted amino acid sequence (Mut protein) for the mutated gDNA.

### 2.4.1. Western blot

Western blots were used to assess changes in protein expression as a result of CRISPR/Cas9 mutations as well as after infection. Western blots of ARID4B and SENP5 protein expression in HEK-293 wild type, CRISPR KO SEMA3A pool and CRISPR KO lines of *ARID4B* (Section 3.2) and *SENP5* (Section 3.3) respectively were performed by Masters student Robyn Maitland. The western blot of SLC7A11 protein expression in U937 wild type, CRISPR 'Miscullin' control and CRISPR KO *SLC7A11* lines (Section 5.4.1) was performed jointly by myself and Masters student Jason Dick.

Adherent cells were washed before lysing with Laemmli Sample Buffer (Sigma), scraping, and fine needle trituration to access all protein. U937 cells were washed and lysed in a similar way. Cell lysates were denatured by boiling for 2 minutes before use. Denatured total protein and a molecular marker (Prestained Molecular Weight Marker, SDS7B2, Merck & Co., Laemmli Sample Buffer, 8 M Urea solution) were loaded on Mini-PROTEAN® Precast Mini PAGE Gels (BioRad, Watford, UK) which were run for one hour at between 80-90V. The proteins were transferred to a nitrocellulose membrane with a Mini Trans-Blot Electrophoretic Transfer Cell (BioRad) before using the Pierce Fast Western Blot Kit, ECL Substrate (Life scientific) according to the manual instructions (unless stated) to stain with antibodies. The membrane (or parts) were bathed in Fast Western Antibody Diluent containing 1,000-fold diluted rabbit anti-human SENP5 polyclonal IgG (ab58420, abcam), rabbit anti-human ARID4B polyclonal IgG antibody (24499-1-AP, ProteinTech or HPA027333, Merck & Co.), SLC7A11 or  $\beta$ -actin (loading control, Invitrogen Beta Actin Ma5 – 15739) antibody for 30 minutes. To complete the process, blots were treated with a universal HRP-conjugated secondary antibody, washed, treated with a luminescent HRP substrate and exposed to film in the dark.

For some blots, the membrane was stripped to allow probing with anti-human  $\beta$ -actin antibody as a loading control. Membranes were bathed in stripping buffer (15 g glycine, 1 g sodium dodecyl sulfate (SDS), 10 mL TWEEN 20 (all Merck & Co.) in 800 mL of distilled water) two times for 5-10 minutes, washed twice with PBS for 10 minutes and finally washed twice in Tris-buffered saline (TBS), 0.1% TWEEN 20 (100 mL of TBS 10x (Thermo Fisher Scientific), 900 mL distilled water, 1 mL TWEEN 20) for 5 minutes. Subsequently the membrane was re-probed as above.

### 2.5. Final established differentiation protocol for U937 cells

The cells were seeded at a density of  $1.6 \times 10^4$  cells/  $\text{cm}^2$  and treated with 0.2  $\mu\text{M}$  PMA ab120297 (Abcam, Cambridge, UK) and 0.8  $\mu\text{M}$  1,25-Dihydroxyvitamin D3 (VD3) (Insight Biotechnology Limited, Wembley, UK) on day 0; before replacing with media without drugs on day 3 for a 3-day resting period finishing on day 6. Cells were used in desired experiments on day 6 as in the outline in Figure 2.5.



## 2.7. Bacterial culture conditions

*Salmonella enterica* serotype Typhimurium (henceforth, referred to as STM for brevity) strain SL1344 and *Shigella sonnei* strain 86 were each streaked on LB agar (with 0.01% Congo red in the case of *S. sonnei*) (Payne and Finkelstein, 1977) from glycerol stocks or streak plates and grown at 37°C overnight. Aliquots of LB were inoculated with a single colony each and grown overnight on a 37°C shaker at 180-250 rpm. Glycerol stocks of each bacterium were prepared with 25% glycerol (Merck & Co.) in LB and aliquoted into screw cap cryotubes (Starlabs, Milton Keynes, UK).

*F. tularensis* subsp. *holarctica* Live Vaccine Strain (LVS) lawns were prepared by spreading 100 µL of glycerol stock onto Blood glucose cysteine agar (BGCA) plates before incubating overnight at 37°C in a sealed box (work conducted within appropriate facilities at DSTL, Porton Down, England). Such work was performed with Helen Flick Smith of DSTL, Porton Down.

**Table 2.1. Bacterial strains and plasmids used.** *Green fluorescent protein is (GFP).*

Strain / plasmid	Genotype or description	Source or reference
<i>Francisella tularensis</i> subsp. <i>holarctica</i>	Live Vaccine Strain (LVS)	DSTL, Porton Down
<i>Salmonella enterica</i> serotype Typhimurium	SL1344 ( $\Delta$ hisG)	Sanger (BA337)
<i>Shigella sonnei</i>	Strain number 86	Gift from the Dr Jun Yu lab
<i>Escherichia coli</i> K12	DH5 $\alpha$ strain GFP	Addgene
<i>Escherichia coli</i> K12	DH5 $\alpha$ strain pGEM-T Easy mCherry (Ampicillin resistance)	Addgene
pGTIV3 plasmid vector		(Tsakiridis et al., 2009)

## 2.8. Bacterial infection assay

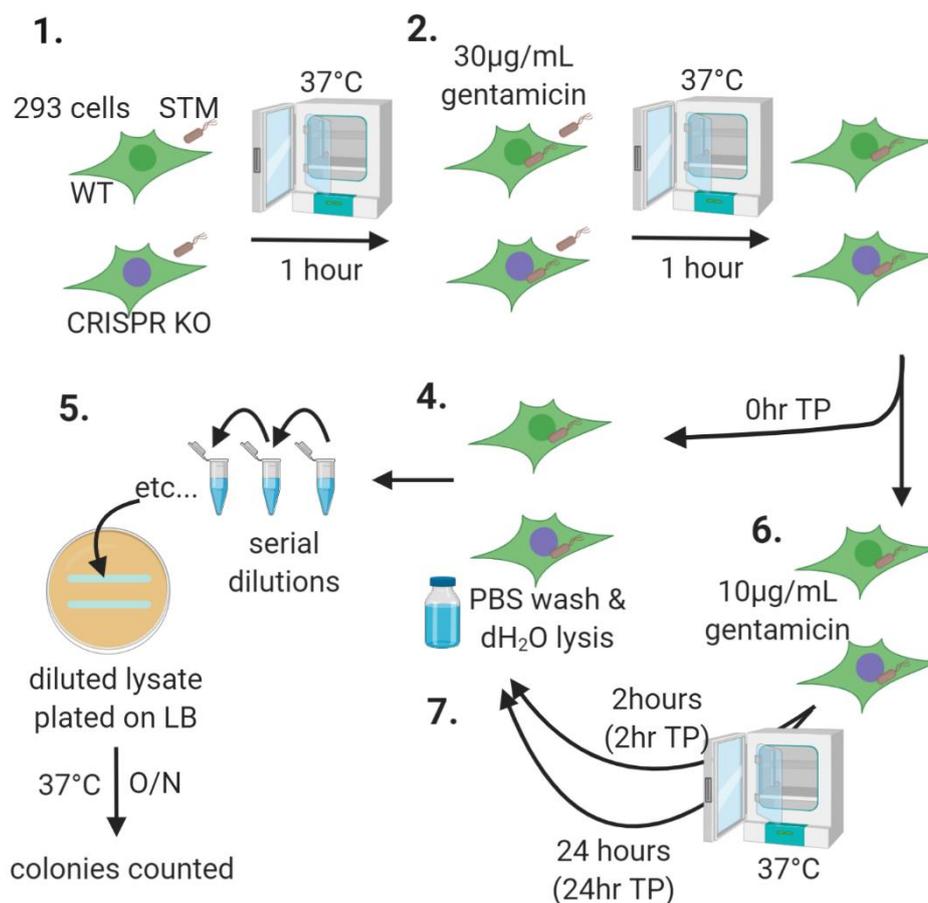
*S. sonnei* were streaked on Congo Red LB agar plates to identify virulent (Congo red positive) colonies (Payne and Finkelstein, 1977). Antibiotic-free HEK-293 cells were seeded at  $4.5 \times 10^5$ / well in poly-L-lysine pre-treated wells in 24-well plates. 100% confluent HEK-293 monolayers (parental or CRISPR mutant lines) inoculated the following day with a fresh 2-hour subculture of Congo red positive *S. sonnei* at a multiplicity of infection (MOI) of 10 (calculated using the equation in Figure 2.7); or 2 days later with a fresh 2-hour subculture of STM at a MOI of 70 (calculated using the equation in Figure 2.8) (summary in Figure 2.9). Infected cultures were incubated for 1 hour at 37°C, 5% CO<sub>2</sub> (inclusive of centrifugation at 896 x g for 10 minutes at room temperature, *S. sonnei*) to allow bacterial invasion. Inoculated medium was replaced with media containing 30 µg/mL gentamicin and cells were incubated for a further hour before treating with media containing 10 µg/mL gentamicin; or washing with PBS and lysing with 1 mL of distilled water for 5 minutes (0-hour time point). 100 µL (or of appropriate dilutions) of cell lysates were spread on agar plates to detect and quantify overnight bacterial growth. Remaining wells were washed and lysed, as above, at the 2-hour and 24-hour time points.

$$1.1026 \times (OD_{600}) + 0.250 = \frac{CFU \times 10^8}{mL}$$

**Figure 2.7 Equation to calculate MOI of *Shigella* sp. 2-hour subcultures.** An equation created by Dr Rasha Mahmoud to calculate the colony forming units (CFU) of *Shigella* sp. LB cultures after sub-culturing an overnight culture for 2 hours. OD stands for optical density (measured using a spectrophotometer, with sterile LB as a blank). CFU stands for colony forming units.

$$10.98 \times (OD_{600}) - 1.228 = \frac{CFU \times 10^8}{mL}$$

**Figure 2.8 Equation to calculate MOI of *Salmonella* sp. 2-hour subcultures.** An equation created to calculate the CFU of *Salmonella* sp. LB cultures after sub-culturing an overnight culture for 2 hours. OD stands for optical density (measured using a spectrophotometer, with sterile LB as a blank). CFU stands for colony forming units.



**Figure 2.9 Bacterial infection assay.** 1. HEK-293 (293) wild type (WT) and CRISPR KO mutant cell lines were infected with STM (as shown here) or *S. sonnei* before incubating for 1 hour at 37°C 5% CO<sub>2</sub>. 2. Infected cells were treated with 30 µg/mL gentamicin to kill extracellular bacteria and incubated for a further hour at 37°C 5% CO<sub>2</sub>. Intracellular bacteria were harvested from infected cells at the 0-hour time point (0hr TP) by 4. washing the cells with PBS and prompting cell lysis with distilled water (dH<sub>2</sub>O). 5. Host cell lysates were serially diluted in PBS, of which two lots of 100 µL were dispensed in straight lines onto LB agar for overnight (O/N) bacterial growth at 37°C. The bacterial colonies were then counted. 6. Prior to processing the 0-hour time point, 10 µg/mL gentamicin was added to cells retained for later time points before returning to 37°C until required. 7. The remaining infected cells were taken through steps (4) and (5) 2 hours (2hr TP) post gentamicin addition as well as 24 hours (24hr TP) post addition.

## 2.9. Lactate dehydrogenase (LDH) activity assay to determine the extent of cell death after infection

Cells were cultured in the previously stated culture media before inoculating with bacteria in DMEM high glucose, without phenol red (Life Technologies), 5% FCS and 1% L-glutamine without antibiotics (LDH media). Inoculation, 30 µg/mL gentamicin steps (for U937 cells) as well as the 10 µg/mL gentamicin step (for HEK-293 cells) were performed as described in Section 2.8, after which cells were incubated until required. The LDH assay was performed on 50 µL supernatant immediately after sampling according to manufacturer's instructions including maximum LDH controls with the absorbance measured at 490 nm. Maximum LDH controls were used as technical positive controls to

determine the maximum LDH produced from the cells present (performed by treating with the cell lysis agent provided in the kit). As *Salmonella* is predicted to have LDH activity based on homology (identified on Uniprot <https://www.uniprot.org/>) (McClelland et al., 2001; The Uniprot Consortium, 2019), 3 - 4 replicate maximum LDH controls were prepared for each MOI and time point tested. Some of the LDH activity assay experiments presented in Section 4.4 were performed with the help of Dr Ruth Holloway.

## 2.10. Immunofluorescence

For immunofluorescence, HEK-293 cells and undifferentiated U937 cells were seeded on poly-L-lysine treated coverslips in 6- or 24-well plates, and U937 cells were differentiated according to the protocol. Cells were washed with PBS before fixing with ice-cold paraformaldehyde (PFA) diluted in CSK buffer (5 M NaCl, 1 M sucrose, 0.5 M 1,4-Piperazinediethanesulfonic acid (PIPES) solution, pH 6.8) for 10-15 minutes. Cells were washed again with PBS before adding methanol as a second fixation step - incubating at -20°C for 20 minutes. Cells were washed again before blocking with antibody diluent (PBS/ 0.1% TWEEN 20/4% horse serum) for 30 minutes at room temperature. Cells were incubated with a primary antibody diluted in antibody diluent for 2 hours before being washed 3 times with PBS. Cells were then incubated with appropriate secondary antibodies diluted in antibody diluent for 1 hour in the dark before washing a further 3 times in PBS. Coverslips were added to microscope slides prepared with a drop of a ProLong™ Diamond Antifade Mountant with 4',6-diamidino-2-phenylindole (DAPI) (Thermo Fisher). Excess mountant was gently removed with pressure before securing the coverslip with superglue. Immunofluorescence experiments and images in Section 3.2 were prepared by Dr Benjamin Pickard or Masters student Robyn Maitland (both of who were coached through the bacterial infection assay in 2.8 by myself).

**Table 2.2 Antibody combinations used for immunofluorescence.** *All secondary antibodies were the Invitrogen brand from Thermo Fisher.*

Primary antibody	Fluorescent secondary antibody
Mouse anti-human $\alpha$ -tubulin monoclonal IgG antibody (Oncogene, CP06-100 UG)	Alexa Fluor 594 goat anti-mouse IgG or Alexa Fluor 488 goat anti-mouse IgG depending on experiment (both Invitrogen, Thermo Fisher)
Rabbit anti-human ARID4B polyclonal IgG antibody (24499-1-AP, ProteinTech)	Alexa Fluor 488 donkey anti-rabbit IgG (Invitrogen, Thermo Fisher)

## 2.11. Screening protocol for differentiated U937 GT library

### 2.11.1. STM U937 screen

The U937 GT library mix was seeded to obtain  $1 \times 10^7$  cells per sample condition and differentiated according to the protocol above (and fed with antibiotic free RPMI on day 3 post-PBS wash). Differentiated cells were then inoculated with a high MOI of 2,400 (calculated using the equation in Figure 2.8 and subsequent experiments) of STM from a 2.5-hr subculture or mock-infected (antibiotic-free RPMI only) on day 6. Infected cells were incubated for 1 hour at 37°C, 5% CO<sub>2</sub> to

allow bacterial invasion before replacing inoculum (or mock inoculum) with RPMI containing 30 µg/mL gentamicin. Cells were kept in the incubator until the RNA was extracted on day 8.

On day 8, the RNA was extracted with an RNeasy mini kit (Qiagen). Briefly, the media was removed, the cells washed with warm PBS before adding RLT buffer (Qiagen) and scraping with a cell scraper. The RLT buffer-cell suspension was centrifuged in 2 Qiashredder tubes per sample, the flow through collected and mixed with 1 volume of 70% ethanol prior to centrifuging in 2 RNeasy spin columns per sample 700 µL at a time according to the manual. The instructions were adhered to during subsequent steps unless stated. On-column DNase digestion was performed with the RNase free DNase kit according to the manual (all Qiagen). RNA was eluted in 3x 40 µL Invitrogen™ UltraPure™ DNase/RNase-Free Distilled Water (Life Technologies). All RNA was stored at -80°C in aliquots to avoid freeze-thawing until required. The RNA quality and concentration were assessed on a 2100 Bioanalyzer with a RNA 6000 Nano kit and ladder (all Agilent, Edinburgh, UK). The experiment was repeated with freshly defrosted U937 GT library cells 6 times, each time using STM infection as the selection pressure. For 4 of the experiments, the RNA extraction was performed jointly by Dr Benjamin Pickard and I.

### 2.11.2. *F. tularensis* U937 screen

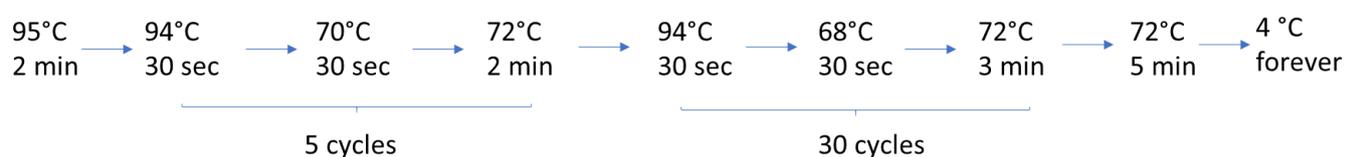
A second differentiated U937 GT library screen was performed with *F. tularensis* LVS as the selection agent at Porton Down, Defence Science and Technology Laboratory in collaboration with Helen Flick Smith. The experiment was repeated with the U937 GT library cells 3 times at passage 4 (Experiment B, C and D) and an additional time at passage 11 (Experiment A). The U937 GT library mix was differentiated as above in Section 2.11.1 (with 1 x 10<sup>7</sup> cells per sample condition) – bar the change to antibiotic, differentiation drug-free media and PBS wash which happened 1 day later (on Day 4). On Day 6, *F. tularensis* LVS from BGCA lawns were resuspended in antibiotic free RPMI to generate a measurement of optical density (OD) at 600 nm (OD<sub>600</sub>) of 1.677; 20 mL of which was diluted in a further 80 mL of antibiotic free RPMI to create the inoculum. The differentiated U937 GT cells were inoculated with 80 mL of the inoculum in antibiotic free RPMI. The MOI was calculated from colony forming unit (CFU) counts from the diluted inoculum one day later (data available in Table 4.2). After inoculation, the infection procedure followed that in Section 2.11.1. On day 8 the RNA was extracted as above in Section 2.11.1 bar the use of 4 Qiashredders per sample, 8 RNeasy mini spin columns per sample and eluted in 3x 40 µL of Invitrogen™ UltraPure™ DNase/RNase-Free Distilled Water. All RNA was stored at -80°C in aliquots to avoid freeze-thawing until required. The RNA quality and concentration were assessed on a 4200 TapeStation with RNA ScreenTape (all Agilent, Edinburgh, UK). The molecular processing described in Section 2.11.1 was performed at the University of Strathclyde, but prior to sample postage from DSTL, a sterility check was performed on all samples. BGCA plates were seeded with 10% of each RNA sample (4 µL), one plate was used per sample and incubated at 37°C. A positive *F. tularensis* BGCA plate and negative blank plate were included as bacterial growth controls. After 7 days plates were assessed for nil growth and results confirmed and witnessed by a second individual.

## 2.12. Molecular processing of RNA from gene trap library samples

cDNA was transcribed from a single (two, LVS screen) aliquot(s) of RNA using superscript IV (Life technologies, Paisley, UK) with the cDNA synthesis primer (Table 2.4, Merck & Co.) according to manufacturer's instructions, including the optional RNA degradation step afterwards with the RNase

H provided. The volume of RNA added to each reaction was normalised to achieve the same concentration in each sample.

cDNA products were initially amplified by PCR (Rapid amplification of cDNA ends (RACE)1) with Invitrogen™ Platinum™ SuperFi™ DNA Polymerase (Life technologies) with the programme in Figure 2.10, UPL (0.19  $\mu$ M), UPS (0.96  $\mu$ M) and SD5-P3 (0.38  $\mu$ M) primers (Table 2.4, Merck & Co.) with and without the SuperFi™ GC enhancer provided. RACE1 products (diluted by 500-fold) were amplified by PCR (RACE2) with the same reagents (including or excluding SuperFi™ GC enhancer) but nested primers (Nest1 and 3'RACE1, 0.5  $\mu$ M each, Table 2.4, Merck & Co.). RACE2 GC enhanced and unenhanced products were combined in a 1:1.4 ratio prior to purification with the Qiaquick PCR purification kit (Qiagen). The concentration of the purified RACE2 products were measured on a Nanodrop (Life Technologies) before sending 20  $\mu$ L (at the concentrations stated in Table 5.4) off to Glasgow Polyomics for DNA library preparation and next generation sequencing. The DNA fragment libraries were prepared with the NEB FS Ultra II DNA kit (size distributions of DNA fragment libraries in Table 2.3). The DNA fragment libraries were sequenced on a NextSeq500 (Illumina, Cambridge, UK) generating paired end sequences with a total read depth of 30,000,000 base pairs (bp) and an average read length of 75 bp.



**Figure 2.10 PCR programme used for RACE1 and RACE2.**

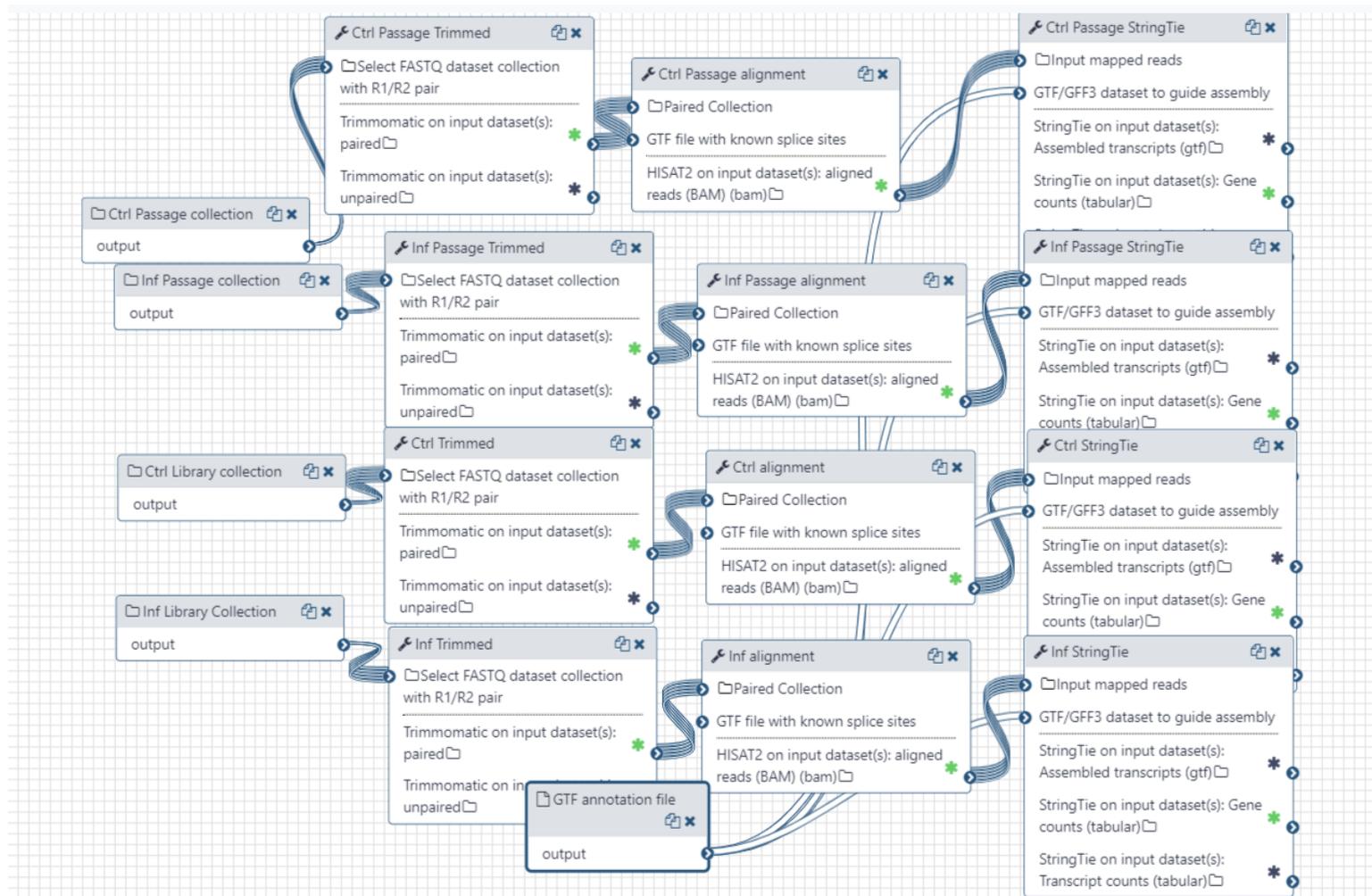
**Table 2.3 Size distribution of DNA fragment libraries created from each experimental sample.** *The PCR GT screen product mixture for each sample was sent to Glasgow Polyomics who fragmented the DNA enzymatically with the NEB FS Ultra II DNA kit to prepare a DNA library for each sample. The average fragment size for each sample is listed as well as the potential size range (based on an estimated +/- 100-150 bp from the average). P7 primer, barcode and adaptor 64 bp and universal primer is 58 bp.*

Sample name	Average Fragment size
D-control	331
E-control	337
F-control	329
G-control	339
H-control	335
I-control	336
D-Experiment	330
E-Experiment	320
F-Experiment	326
G-Experiment	333
H-Experiment	314
I-Experiment	331
Original 1	330
Original 2	328

### 2.13. Bioinformatics analysis of 'RNA-seq' data from gene trap screens

Glasgow Polyomics provided a MultiQC file containing summary data and figures from FastQC and FastQ screen reports of the whole dataset to indicate the raw data quality. The raw sequencing files were analysed on Galaxy Main at usegalaxy.org (Afgan et al., 2018); a summary diagram is presented in Figure 2.11 indicating those steps which could be automated in Galaxy. Some heuristic optimisation of quality processing was performed prior to processing samples in bulk, the details of this are described in Chapter 5 and Appendix C. In brief, reads were trimmed with the sliding window feature to an average Phred of 20 and filtered for a minimum length of 20 bp in Trimmomatic version 0.36.6 (Bolger et al., 2014). FastQC version 0.11.8 was used to assess the quality of raw and trimmed sequences in order to make decisions about the processing steps. The paired trimmed and filtered reads were aligned to the Human genome (b38, hg38) with HISAT2 version 2.1.0+galaxy4 (Kim et al., 2015), as a paired end dataset collection. Additionally, transcript assembly reporting was tailored for StringTie. The Bam output files were used to assemble and count transcripts, from which gene counts could be generated with StringTie version 1.3.4d (Pertea et al., 2015). A cDNA Human reference library, assembly GRCh38.p13 from Ensembl release 98 (Hunt et al., 2018) was used to guide transcript assembly; only transcripts which matched this were used (i.e. novel transcripts were discarded). Also, output files generated by StringTie were tailored for limma. The resulting gene count files were combined with Column Join version 0.0.3, Ensembl version numbers removed and genes annotated with entrez IDs, symbols and gene names with AnnotateMyIDs version 3.5.0.1. The resulting file was uploaded to Degust version 4.1.1 (<http://degust.erc.monash.edu/>) (David Powell, 2015) in a comma-separated values format to

perform differential gene analysis using Limma-voom version 3.40.6 (Law et al., 2014; Liu et al., 2015; Ritchie et al., 2015). The *de novo* transcriptome based bioinformatics analysis described in Section 5.2.5 was performed by Honours student Yiwen Ding under my co-supervision and training, due to the extended time required to optimise the process to fit it to our purposes.



**Figure 2.11 Overview of bioinformatics pipeline in Galaxy. Galaxy workflow (automated pipeline of tools where the output of one is fed into another) for LVS U937 GT screen used as an example. Following this workflow, the gene counts from each sample were combined with ColumnJoin, the Ensembl ID version numbers removed and AnnotateMyID was used to add additional information. Experiment A which used U937 GT library at passage 11 was processed separately (Passage) from the other Infected (Inf) and Control (Ctrl) samples. Collections represented here are lists of raw (untouched) dataset pairs (forward and reverse reads). The Galaxy workflow for the STM U937 GT screen was the same except with different labels.**

## 2.14. Independent validation of a gene hit (SLC7A11) from the STM U937 infection screen, through use of a CRISPR KO pool and pharmacological modulation of the protein

*SLC7A11* CRISPR KO pooled and wild type U937 cells were treated with the cocktail drug differentiation procedure (by Masters student Jason Dick under my guidance) described in Section 2.5 following the same timeline. On day 6, differentiated cells were inoculated with  $4.09 \times 10^7$  green fluorescent protein (GFP)-expressing STM /well in 6-well or 24-well plates. Sulfasalazine (5 or 10 ng/mL), where used, was added together with the typical 30 µg/mL gentamicin one hour later. Cells were washed with PBS, fixed with ice-cold PFA for 15 minutes at 24 hours-post inoculation before washing twice more with PBS and leaving in fresh PBS at 4°C until staining. Cells were stained with various antibody combinations as detailed in Table 2.2. The infection experiment was performed jointly by Master student Jason Dick and I. Jason performed the subsequent staining, whereas the images were acquired by both Jason and I.

### 2.14.1. Counting intracellular bacteria

Slides were 'blinded' to the observer (Dr Benjamin Pickard) with regard to treatment/genotype using tape to cover slide annotation. A 'scan' was performed at x40 oil-immersion objective magnification (x400 in total), moving from the far left of the microscope coverslip to the far right, selecting those cells with a clear and obvious flattened differentiation state in terms of size and structure. GFP-expressing bacteria or bacterial fragments were counted in these cells by scanning through all focal planes of the cells. Cells which were covered by others were excluded from the analysis.

**Table 2.4 Primers and CRISPR oligo sequences**

Name	Sequence
<b>Primers for CRISPR validation</b>	
SENP5 Forward	TGATGCTAAAGTTAAGCCCTTGA
SENP5 Reverse	CCAAAAGACAACCTGAGAAGCCA
PRKG1 F1	GACTTCTCATCCTCCCCTCG
PRKG1 R1	GTAAACTTTCCGCGGTCCTC
PCDH7 F3	GATCGGGTGAGGTGACTTTC
PCDH7 R3	CACCTGCAGCTCGAACAC
<b>Gene trap processing</b>	
cDNA synthesis primer	AAGCAGTGGTAACAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTV
UPL	CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT
UPS	CTAATACGACTCACTATAGGGC

Nest1	AAGCAGTGGTAACAACGCAGAGT
SD5-P3 (3'RACE7, 3pMS1-1)	CGCATCGCCTTCTATCGCCTTCTTGACG
3'RACE1 (3pMS1-2)	CAAGCGACGCCAACCTGCCATCACGAG
<b>CRISPR oligos for single guide RNA (sgRNA)</b>	
C5orf60 top sequence 1	CACCTGTAGAGGCGTAAGATAATC
C5orf60 bottom sequence 1	AAACGATTATCTTACGCCTCTACA
C5orf60 top sequence 2	CACCACAGGAAAGTGTAGATCTCA
C5orf60 bottom sequence 2	AAACTGAGATCTACACTTTCCTGT
Cul1 top sequence 1	CACCGTATCGAGCCAGCAACTCAG
Cul1 bottom sequence 1	AAACCTGAGTTGCTGGCTCGATAC
Cul1 top sequence 2	CACCTCCGGGTTGACGACATTGTG
Cul1 bottom sequence 2	AAACCACAATGTCGTCAACCCGGA
SLC7A11 top sequence 1	CACCATTCCGACCCATTTAGTACA
SLC7A11 top sequence 1	AAACTGTACTAAATGGGTCCGAAT
SLC7A11 top sequence 2	CACCAAGTATTACGCGGTTGCCAC
SLC7A11 bottom sequence 2	AAACGTGGCAACCGCGTAATACTT
LGALS12 top sequence 1	CACCTGGAAGGCGATATCTGGCCG
LGALS12 bottom sequence 1	AAACCGGCCAGATATCGCCTTCCA
LGALS12 top sequence 2	CACCCTGTCTCATGTGGACACGCT

LGALS12 bottom sequence 2	AAACAGCGTGTCCACATGAGACAG
SQLE top sequence 1	CACCATCCGAGAAGAGGGCGAACT
SQLE bottom sequence 1	AAACAGTTCGCCCTTCTCTCGGAT
SQLE top sequence 2	CACCCGCTGTCCGCCACCGAAACGG
SQLE bottom sequence 2	AAACCCGTTTCGGTGGCGACAGCG
PRKG1 top sequence 1	CACCTGTGGATTGTATGTACCCGG
PRKG1 bottom sequence 1	AAACCCGGGTACATACAATCCACA
PRKG1 top sequence 2	CACCGTACCCGGTGGAGTATGGCA
PRKG1 bottom sequence 2	AAACTGCCATACTCCACCGGGTAC
MANSC1 top sequence 1	CACCAGGGAGCTTGACTTACTT
MANSC1 bottom sequence 1	AAACAAGTGTAAAGTCAAGCTCCCT
MANSC1 top sequence 2	CACCATGAGTTACAGGATAATTAC
MANSC1 bottom sequence 2	AAACGTAATTATCCTGTAACAT
NCKAP1L top sequence 1	CACCCTGTTTCGGACATCTATGTT
NCKAP1L bottom sequence 1	AAACAACATAGATGTCCGAAACAG
NCKAP1L top sequence 2	CACCGCTGTTTCGGACATCTATGT
NCKAP1L bottom sequence 2	AAACACATAGATGTCCGAAACAGC
Control CUL1 top sequence 1	CACCGTATAGAGCAAGCAACTCAT
Control CUL1 top sequence	AAACATGAGTTGCTTGCTCTATAC

## 2.15. Statistical analysis

All stats were performed with Graphpad Prism 6.01, unless stated otherwise. In all cases where relevant, \* is  $p \leq 0.05$ , \*\* is  $p \leq 0.01$ , \*\*\* is  $p \leq 0.001$  and \*\*\*\* is  $p \leq 0.0001$ . Normality was assessed visually only due to the small sample sizes (Ghasemi and Zahediasl, 2012).

In Chapter 3, the correlation between CFU and OD was assessed by calculating a linear regression and goodness of fit. The LDH assay and bacterial burden assay outputs were analysed using two-way Analysis of variance (ANOVA) to determine if there was a statistically significant difference present. Multiple comparisons were made to wild type at the same time point with Dunnett's multiple comparisons test.

In Chapter 4, the flow cytometry data presented was analysed by one-way ANOVA and multiple comparisons were made with Tukey's multiple comparisons test between each treatment. The infection optimisation experiments were statistically analysed individually using two-way ANOVA on the intra-experimental replicates, after which post hoc comparisons were made between every condition using the Tukey test (using Graph Pad version 6.01). P values are only reported for comparisons between the same MOI across different time points and to MOI 0 at the same time point. The Cytotox96 reagent breakdown data was analysed using two-way ANOVA and multiple comparisons were made to the first use of the new assay buffer for LDH max and blank respectively, using Dunnett's multiple comparison test.

In Chapter 5, the intracellular bacterial burden in the infection phenotyping assay was processed in two ways – firstly it was presented on a logarithmic scale which was then statistically analysed with one-way ANOVA. No statistical difference was found, so post-hoc tests were not performed. Secondly, the count data was grouped into bins and is presented as a stacked bar chart.

In Appendix C, the output from the investigation of HISAT2 and StringTie were visually inspected, but no statistical analyses were made.

## Chapter 3

Creating and characterising cell line mutants to confirm host candidate gene roles in the infection process

### 3. CREATING AND CHARACTERISING CELL LINE MUTANTS TO CONFIRM HOST CANDIDATE GENE ROLES IN THE INFECTION PROCESS

In Chapter 1, a previous HEK-293 GT library screen using *Salmonella enterica* serotype Typhimurium SL1344 (STM) was described (a protocol is presented in Figure 3.11A). This identified infection-resistant cells and then isolated the mutated genes responsible for this phenotype (J Yu, personal communication, 3 October 2016). The mutated genes represented in the surviving population of STM-infected HEK-293 GT library cells included SUMO specific peptidase 5 (*SEN5*), AT-rich interaction domain 4B (*ARID4B*), protocadherin 7 (*PCDH7*) and proteasome 20S subunit alpha 1 (*PSMA1*). An additional selection procedure was carried out, using fluorescence-activated cell sorting (FACS) of high- or low-fluorescent HEK-293 cells after infection with GFP-expressing STM. This screen identified an additional mutated gene, protein kinase cGMP-dependent 1 (*PRKG1*). Together, these genes were positioned within a potential protein-protein interaction network and we hypothesised that this may represent a functional host defence network (details in Chapter 1). Large-scale genetic discovery screens are intentionally analysed in a more statistically liberal manner to reduce the quantity of missed hits (Type II errors). This more statistically liberal approach, as well as the great quantity of statistical tests performed, will naturally allow more false positives (Type I errors). Additionally, the imprecise nature of the gene trapping (or, rather, sequenced-based interpretation of specific insertional mutations) means that establishing a causal link between the identified gene and the phenotype (host survival under infection conditions) can be a complex task. The combination of these factors necessitates a second, independent set of experiments to confirm the candidate genes using targeted disruptions; thus, the decision was made to create KO cell lines using the CRISPR-Cas9 system. Creation and genetic or protein-based characterisation of CRISPR KO cell lines for four of these genes (*SEN5*, *PRKG1*, *PCDH7* and *ARID4B*) is described here as an independent means to validate the role of these genes in host cell protection against infection.

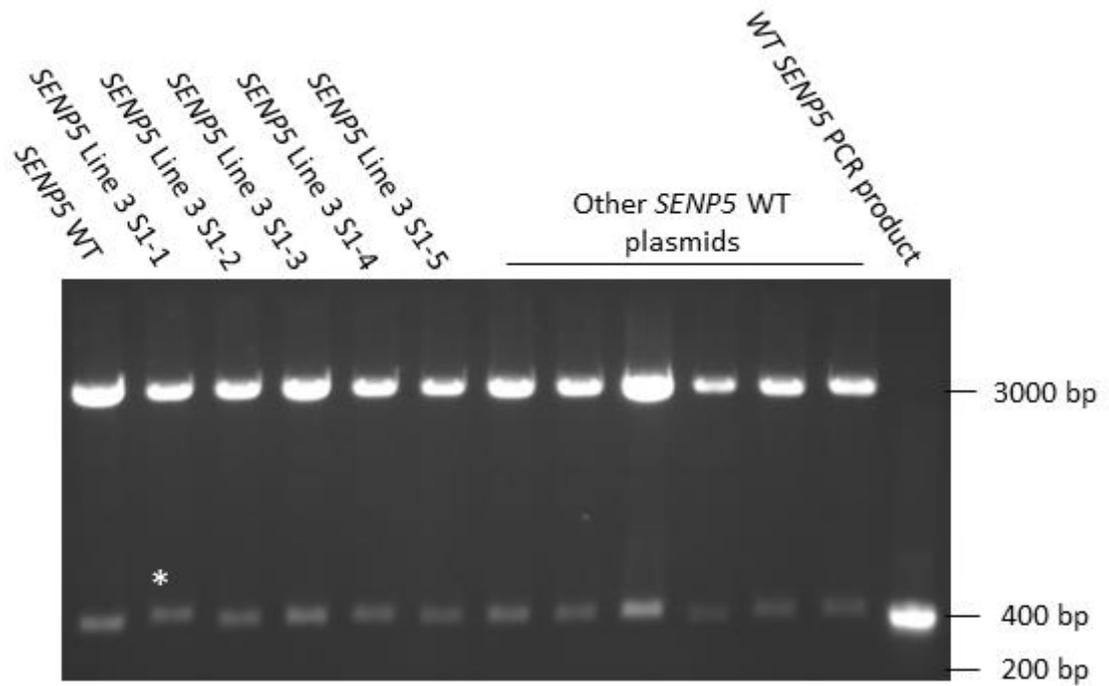
#### 3.1. Genetic validation of isolated CRISPR KO cell lines

CRISPR KO lines in the HEK-293 background were created as described in Materials and Methods with GeneCopoeia-designed plasmids (which could be transfected by nucleofection and remained consistent with previous work). Suitable CRISPR target sites were chosen by the company and made available on request after purchase. CRISPR induced mutation remedied by NHEJ (generating a frame-shift mutation) was most desired as it has the most profound effect on the protein expression and function. Amino acid deletions and insertions were also accepted. Briefly, the CRISPR-Cas9 linearised plasmid (which also contained the relevant sgRNA) was transfected into HEK-293 wild type cells, which were cultured under antibiotic selection as a pool, as well as serially diluted for colony isolation. Once a resistant colony became visible by eye, it was transferred by

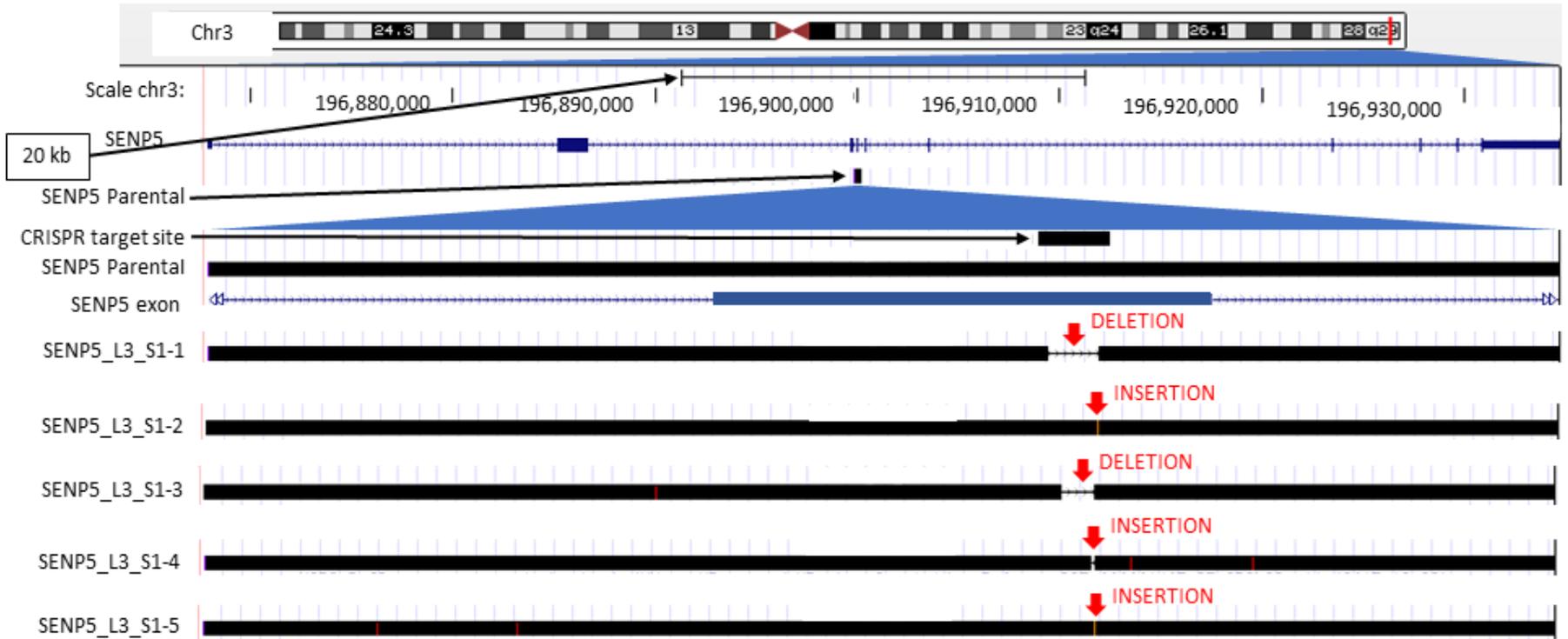
pipette to a fresh well. Once a given cell line had proliferated sufficiently, gDNA was extracted and the cell line frozen for future use.

### 3.1.1. Confirming CRISPR mutations in *SENP5* KO cell lines

To narrow down potential mutant cell lines of interest from the isolated cell lines, Sanger sequencing was employed to characterise the targeted locus. pGEM-T Easy plasmids containing PCR-amplified regions surrounding the relevant CRISPR target site were prepared as described in the Chapter 2. Briefly, the genomic DNA locus surrounding the CRISPR target site (20 bp) of each targeted gene was amplified by PCR using myTaq™ (a Taq polymerase) before ligating the PCR product to the thymine 3' overhangs of the pGEM-T Easy vector. A Taq polymerase was used (despite its error prone activity) to maintain the complementary adenine overhangs on the 3' ends of the products. Each ligated mixture was transformed into *E. coli* DH5 $\alpha$ . Once extracted from single colony-inoculated DH5 $\alpha$  liquid cultures, plasmids were digested with EcoRI (one of many restriction enzymes that can cut an insert from the pGEM-T Easy vector leaving it with an extra 20 bp compared to the original insert). Most CRISPR-derived insertion/deletion mutations would not be extensive enough to be distinguishable by eye as changes in band migration on an agarose gel; therefore, it was a surprise to find that one of the CRISPR mutant *SENP5* pGEM-T Easy EcoRI-digested plasmids was visibly different in size compared to the wild type PCR product (Figure 3.1). The visible difference in size indicated that imperfect NHEJ repair after CRISPR DNA damage had produced a relatively large deletion in the DNA sequence. Indeed, upon sequence comparison of the mutated sequences with the human reference genome, the plasmid inserts were found to have 3 deletions and 2 insertions (Figure 3.2) of varying sizes (details in **Table 3.1**), all located within the CRISPR target site of 20 bp. The presence of more than two versions of this locus was unexpected, as *Homo sapiens* is diploid. Nevertheless, this emphasised the importance of sequencing multiple plasmid inserts per cell line to be used.



**Figure 3.1 PCR products of some mutated *SENP5* lines are visibly different in size compared to wild type PCR product.** *EcoRI*-digested *pGEM-T Easy* plasmids with *SENP5* wild type (WT) and CRISPR mutant Line 3 PCR product inserts run on a 1% agarose gel. The *EcoRI* digested *pGEM-T Easy* vector is 2998 bp, the WT PCR product is 376 bp and the *EcoRI* digested inserts should be 396 bp. The band labelled with a \* shows deviation from the expected size and may represent a significant insertion mutation. Experiment performed once.



**Figure 3.2 HEK-293 CRISPR *SENP5* Line 3 contains multiple deletions and insertions at the CRISPR target site.** *The box at the top shows a schematic of chromosome 3 with a red line indicating the position of the *SENP5* gene. A scale bar in kilobases is presented above a representation of NCBI RefSeq version of the whole *SENP5* gene. Immediately below is a black square indicating the location of the amplified sequence from the parental line in relation to the gene. Underneath this is a zoomed in view of all full-length PCR products (thick black lines) surrounding the CRISPR target site (*CRISPR\_target\_seq*) from the parental line and *SENP5* Line 3. A representation of the reference *SENP5* amino acid sequence (given in single letter code) in blue for the protein coding region of the exon (4 of 9) is below each PCR product sequence. The PCR products from the mutant line are labelled in the following fashion; gene name\_L(line number)\_S (number of times sampled)-(sample number). Sample refers to the insert of a plasmid from a single bacterial colony transformed with the pGEM-T Easy vector ligated with PCR product from a CRISPR targeted region or wild type sequence. Vertical red lines in the PCR product indicate single nucleotide changes, orange lines indicate nucleotide insertions and breaks indicate nucleotide deletions. Arrows indicate the direction of the sequence. "hg38" is the UCSC Genome Browser assembly ID used in this analysis. Experiment performed once.*

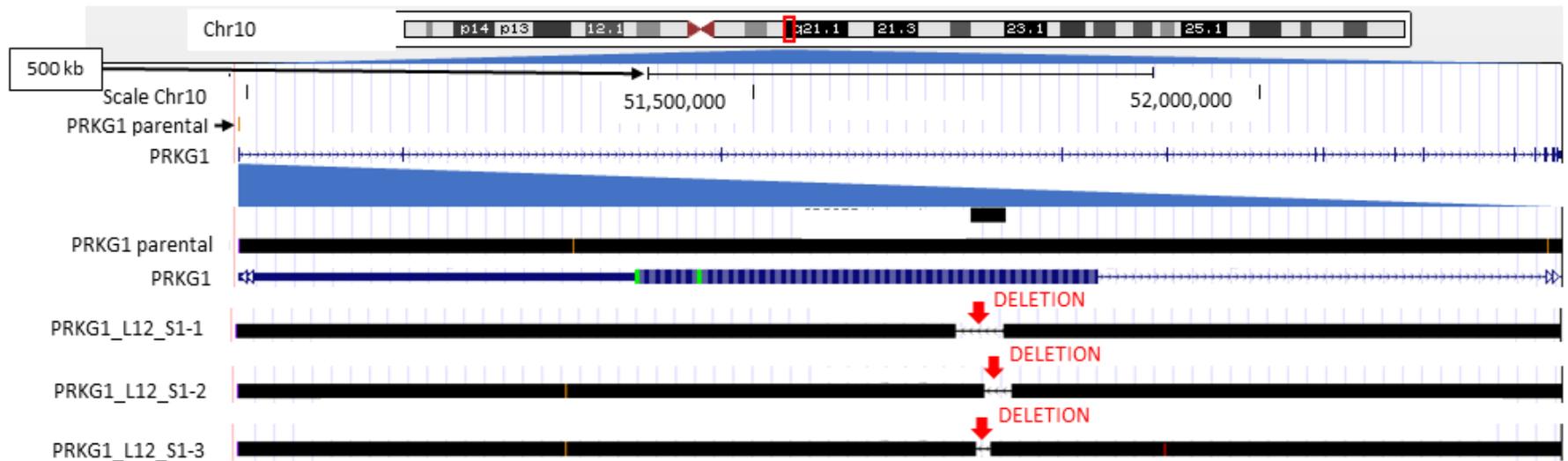
To demonstrate the strength of the consequences of the CRISPR mutations to the DNA sequence and resulting protein, the amino acid sequence of mutated *SENP5* was predicted as described in the Materials and Methods. The wild type amino acid sequence of *SENP5* is 755 amino acids long (Figure 3.3A) and the amino acids encoded by CRISPR target region are located more than three quarters of the way through this sequence. All plasmid inserts sequenced had mutations affecting the amino acid sequence (Figure 3.3B-F); including two frameshifts resulting in premature (nonsense) stop codons, an insertion resulting in a premature stop codon, a deletion and an insertion resulting in fewer or additional amino acids, respectively. A frameshift occurs when the number of nucleotides added or removed are not divisible by 3, causing subsequent nucleotides to code for different amino acids; until the altered reading frame results in a stop codon. The *SENP5* CRISPR KO cell line 3 was deemed to have a sufficient proportion of mutations predicted to be non-functional to take forward for functional analysis.

- A** MKKQRKILWRKGIHLAFSEKWNTGFGGFKFYFHQHLCLKAKLGRPVTWNRQLRHFQGRK  
 KALQIQKTWIKDEPLCAKTKFNVATQNVSTLSSKVKRDKAKHFISSSKTLRLQAELLSS  
 AKNSDHEYCREKNLLKAVTDFPNSALGQANGHRPRTDPQPSDFPMKFNGESQSPGESGTI  
 VVTLNHNKRGKFCYGCCQGPEHHRNGGPLIPKKFQLNQHRRIKLSPLMMYEKLSMIRFRYR  
 IILRSQHFRTKSKVCKLRKAQRSWVQKVTGDHQETRENGEGGSCSPFSPPEPKDPSCRHQF  
 YFPDMDSSAVVKGNTNSHVPDCHTKGSSFLGKELSLDEAFPQQNGSATNAWDQSSCSPKW  
 ECTELIHDIPLPEHRSNTMFISETEREIMTLGQENQTSVSDDRVKLSVSGADTSVSSVDG  
 PVSQKAVQENNSYQMEEDGSLKQSILSSELLDHPYCKSPLEAPLVC SGLKLENQVGGGKNS  
 QKASPVDDEQLSVCLSGFLDEVMMKYGSLVPLSEKEVLGRLKDVFNEDFSNRKPFINREIT  
 NYRARHQKCNFRI FYNKHMLD**MDDL**ATLDGQNWLNQVINMYGELIMDAVPDKVHFFNSFF  
 HRQLVTKGYNGVWRWTKKVDLFFKKSILLIPIHLEVHWSLITVTLNRIISFYDSQGIHFKF  
 CVENIRKYLLEAREKNRPEFLQGWQTAVTKCIPQQKNSDCGVFVLOYCKCLALEQPFQF  
 SQEDMPRVKRRIYKELCECLMD
- B** SENP5\_L3\_S1-1  
 RKPFINREITNYRARHQKCNFRI FYNKHMLDISGWSELAE\*
- C** SENP5\_L3\_S1-2  
 RKPFINREITNYRARHQKCNFRI FYNKHMLD**MDDL**AISRI\*
- D** SENP5\_L3\_S1-3  
 RKPFINREITNYRARHQKCNFRI FYNKHMLD**M**ALDQNWLNQ
- E** SENP5\_L3\_S1-4  
 RKPFINREITNYRARHQKCNFRI FYNKHMLD**MDDL**ALWMARTG\*
- F** SENP5\_L3\_S1-5  
 RKPFINREITNYRARHQKCNFRI FYNKHMLD**MDDL**AIWTLDGQNWLNQ

**Figure 3.3 HEK-293 SENP5 Line 3 contains multiple copies of the SENP5 gene with mutations predicted to create non-functional protein.** Predicted amino acid sequence of SENP5 based on the human reference genome, protein accession number NP\_689912.2, (isoform 1, encoded by transcript variant 1) from UCSC genome browser (A). The amino acid sequence encoded by the targeted exon is highlighted in grey and the specific region encoded by the DNA sequence targeted by the CRISPR sgRNA is in bold. Amino acid sequences from the target exon based on PCR product sequences from HEK-293 SENP5 Line 3 sample 1-1 (B), sample 1-2 (C), sample 1-3 (D) and sample 1-4 (E). Sample refers to the insert of a plasmid from a single bacterial colony transformed with the pGEM-T Easy vector ligated with PCR product from a CRISPR targeted region or wild type sequence. Underlined letters indicate positions where the amino acid sequence is different (due to an insertion, a deletion or a frame shift) and asterisks indicate premature (nonsense) stop codons. Experiment performed once.

### 3.1.2. Confirming CRISPR mutations in *PRKG1* KO cell lines

Five pGEM-T Easy plasmids containing *PRKG1* PCR products per cell line from four cell lines were prepared for sequencing (not shown). Two of the cell lines (11 and 12) had multiple non-functional mutations (not shown and Figure 3.4). Three PCR product samples are shown in Figure 3.4 of the five plasmids sent for sequencing as the remaining two were primer dimers. The PCR product samples represented in Figure 3.4 all have deletion mutations within or encompassing the CRISPR target region.



**Figure 3.4 HEK-293 CRISPR *PRKG1* Line 12 contains multiple deletions at the CRISPR target site.** The box at the top shows a schematic of chromosome 10 with a red line indicating the position of the gene. A scale bar in kilobases is presented above a representation of the whole *PRKG1* gene (the specific NCBI RefSeq version targeted). Immediately above is a vertical red line indicating the location of the amplified PCR sequences in relation to the gene. Underneath this is a zoomed in view of all full-length PCR products (thick black lines) surrounding the CRISPR target site (*CRISPR\_target\_seq*) from the parental line and *PRKG1* Line 12. A representation of the targeted *PRKG1* exon in blue is below each PCR product sequence. The PCR products from the mutant line are labelled in the following fashion; gene name\_L (line number)\_S (number of times sampled) - (sample number). Vertical red lines in the PCR product indicate single nucleotide changes, orange lines indicate nucleotide insertions and breaks indicate nucleotide deletions. Arrows indicate the direction of the sequence. “hg38” is the UCSC Genome Browser assembly ID used in this analysis. Experiment performed once.

The *PRKG1* protein sequence is 671 amino acids long (Figure 3.5A) and the amino acids encoded by the CRISPR target site are near the start of the sequence (encoded as part of the first exon). Any mutations causing frameshifts are thus likely to create severe or null mutations in this gene. The deletions in the nucleotide sequence for each of the PCR product samples for CRISPR KO *PRKG1* Line 12 have produced a frameshift, causing a different series of amino acids to follow those encoded by the CRISPR target site (Figure 3.5B-D). The altered reading frame continues past the end of the first exon, but as the splice site itself has not been disrupted, the mRNA and protein sequences continue into the next exon. The predicted protein sequences all terminate prematurely near the start of the second exon. The mutations found within the *PRKG1* CRISPR target site in *PRKG1* Line 11 also resulted in predicted short protein sequences (data not shown). When both lines were cultured for expansion purposes, *PRKG1* Line 12 grew more comparably to HEK-293 WT cells (little to no clustering) and was chosen for further experimental work.

- A** MSELEEDFAKILMLKEERIKELEKRLSEKEEEEIQELKRKLHKCQSVLPVPSTH  
 IGPRTTRAQGI**SAEPQTY**RSFHDLRQAFRKFTKSE~~RSKDLIKEA~~ILDNDFMKN  
 LELSQIQEIVDCMPVEYGKDSCI IKEGDVGSLSVYVMEDGKVEVTKEGVKLCF  
 MGPVKVFGELAILYNCTRATVKTLLVNVKLWAIDRQCFQTIMMRTGLIKHTEY  
 MEFLKSVPTFQSLPEEILSKLADVLEETHYENGEYIIRQGARGDTFFIISKGT  
 VNVTRSDSPSEDPVFLRTLGGKDWFGKALQGEDVRTANVIAAEAVTCLVIDR  
 DSFKHLIGGLDDVSNKAYEDAEAKAKYEAEAAFFANLKLSDFNIIDTLGVGGF  
 GRVELVQLKSEESKTFAMKILKRRHIVDTRQQEHIRSEKQIMQGAHSDFIVRL  
 YRTFKDSKYLYMLMEACLGGELWTILDRGSGFEDSTTRFYTACVVEAFAYLHS  
 KGIYRDLKPENLILDHRGYAKLVDFGFAKKIGFGKKTWTFCGTPEYVAPEII  
 LNKGHDISADYWSLGIILMYELLTGSPPFSGPDPMKTYNIIILRGIDMIEFPKKI  
 AKNAANLIKKLCRDNPSERLGNLKNGVKDIQKHKWFEGFNWEGLRKGLTTPPI  
 IPSVASPTDTSNFDSFPEDNDEPPPDDNSGWDIDF
- B** PRKG1\_L12\_S1-1  
 MSELEEDFAKILMLKEERIKELEKRLSEKEEEEIQELKRKLHKCQSVLPVPSTH  
 IGPRTTRAQGIIPRPPTGIPEVHQVRKVQGSYKGSYP\*
- C** PRKG1\_L12\_S1-2  
 MSELEEDFAKILMLKEERIKELEKRLSEKEEEEIQELKRKLHKCQSVLPVPSTH  
 IGPRTTRAQGI**SAEPQT**TSDRHSGSSPSPKGPRIL\*
- D** PRKG1\_L12\_S1-3  
 MSELEEDFAKILMLKEERIKELEKRLSEKEEEEIQELKRKLHKCQSVLPVPSTH  
 IGPRTTRAQGI**SAEPQ**VLRPPTGIPEVHQVRKVQGSYKGSYP\*

**Figure 3.5 HEK-293 PRKG1 Line 12 contains multiple non-functional copies of the PRKG1 protein.** Predicted amino acid sequence of PRKG1 based on the human reference genome, protein accession number NP\_001091982.1, (isoform 1, encoded by transcript variant 1) from UCSC genome browser (A). The amino acid sequence encoded by the targeted exon is highlighted in grey and the corresponding amino acid sequences predicted from the following PCR products are presented; HEK-293 PRKG1 Line 12 sample 1-1 (B), sample 1-2 (C) and sample 1-3 (D). Sample refers to the insert of a plasmid from a single bacterial colony transformed with the pGEM-T Easy vector ligated with PCR product from a CRISPR targeted region or wild type sequence. Where mutations in the DNA sequence cause a frameshift, predictions of amino acid sequences are presented based on changes to the mRNA transcript. Letters in bold indicate the DNA sequence homologous to the CRISPR sgRNA, underlined letters indicate positions where the amino acid sequence is different (due to frame shifts and deletions) and asterisks indicate premature (nonsense) stop codons. Experiment performed once.

### 3.1.3. Confirming CRISPR mutations in *PCDH7* KO cell lines

The transformations for the pGEM-T Easy plasmids containing *PCDH7* PCR product inserts from CRISPR *PCDH7* Lines 1-3 were of relatively low efficiency (3 surviving bacterial colonies each). Hence, the plasmid derived from each colony was checked for an insert and sent off for sequencing. Only the PCR product samples from CRISPR mutant *PCDH7* Line 2 had mutations (Figure 3.6, S1-1 and S1-2) in the CRISPR target region. The third PCR product sample for *PCDH7* Line 2 was a primer dimer. As only 2 PCR product samples had been sequenced for *PCDH7* Line 2, further plasmids containing PCR product samples for this line were generated and sequenced (Figure 3.6, S2-1, S2-2, S2-3 and S2-4). Overall, more than 60% of the PCR product samples from this cell line have mutations (2 insertions and 2 deletions).



**Figure 3.6 HEK-293 CRISPR PCDH7 Line 2 contains multiple deletions at the CRISPR target site.** *The box at the top shows a schematic of chromosome 4 with a red line indicating the position of the gene. A scale bar in kilobases is presented above a representation of NCBI RefSeq versions of the whole PCDH7 gene and the label (on the left) for the version targeted is highlighted in blue. Immediately above is a vertical red line indicating the location of the amplified PCR sequences in relation to the gene. Underneath this is a zoomed in view of all full-length PCR products (thick black lines) surrounding the CRISPR target site (CRISPR\_target\_seq) from the parental line and PCDH7 Line 2. A representation of the targeted PCDH7 exon in striped black is below each PCR product sequence. The PCR products from the mutant line are labelled in the following fashion; gene name\_L(line number)\_S (number of times sampled)-(sample number). Vertical red lines in the PCR product indicate single nucleotide changes, orange lines indicate nucleotide insertions and breaks indicate nucleotide deletions. PCR products S1-1 and S1-2 are represented with breaks in the line even though they have insertions; as they both have one adenine replaced with a series of thymines, cytosines and guanines. Arrows indicate the direction of the sequence. "hg38" is the UCSC Genome Browser assembly ID. Experiment performed once.*

Protocadherin 7 contains 1067 amino acids, with all but the last 11 encoded by the first exon (Figure 3.7A). CRISPR KO *PCDH7* Line 2 PCR product samples S1-1 and S1-2 have the same DNA mutation (an insertion resulting in a frameshift, see Table 3.1 for more detail) and predicted protein sequence (Figure 3.7B, C). Samples S2-1 and S2-3 (Figure 3.7D, F) have the protein sequence as wild type; whereas samples S2-2 and S2-4 have the same inserted amino acids, which replaced one of the amino acids encoded by the CRISPR target site (Figure 3.7E, G). CRISPR KO *PCDH7* line 2 was used to investigate the *PCDH7* KO infection phenotype (in further sections of this Chapter) identified in the HEK-293 screen. **Table 3.1** summarises the details of each mutation and its consequence for each cell line and gene of interest (including genes and cell lines that were not taken forward).

**A** MLRMRTAGWARGWCLGCCLLPLSLSLAAAKQLLRYRLAEEGPADVRIGNVASDLGIVTGSSEVT  
FSLESGSEYLKIDNLTGELSTSERRIDREKLPQCQMI FDENECFLDFEVSVIGPSQSWVDLFEQG  
VIVLDINDNTPTFFSPVLTLTVEENRPVGTLYLLPTATDRDFGRNGIERYELLQEPGGGGSGGES  
RRAGAADSAPYPGGGGNGASGGGSGGSKRRLDASEGGGGTNPGRSSVFELQVADTPDGEKQPQL  
IVKGALDREQRDSYELTLRVRDGGDPPRSSQAILRVLITDVNDNSPRFEKSVYEADLAENSAPGT  
PILQLRAADLDVGVNGQIEYVFGAATESVRRLRLDETSGLWSVLHRIDREEVNLQRFVTVMADR  
GQPPKTKATVVLNIKDENINVPSIEIRKIGRIPLKDGVANVAEDVLVDTPIALVQVSDRDQGEN  
GVVCTCTVVGDVFPQLKPASDTEGDQNKKKYFLHTSTPLDYEATREFNVVIVAVDSGSPSLSSNNS  
LIVKVGDTNDNPPMFGQSVVEVYFPENNI PGERVATVLIATDADSGKNAE IAYS LDSSVMGI FAID  
PDSGDILVNTVLDREQTDRYEFKVNADKGI PVLQGSTTVIVQVADKNDNDPKFMQDVFTFYVKE  
NLQPNSPVGMVTVMADKGRNAEMSLYIEENNI FSIENDTGTIYSTMSFDREHQTTYTFRVKAV  
DGGDPPRSATATVSLFVMDENDNAPTVTLPKNISYTLPPSSNVRTVVATVLIATDSDDGINADLN  
YSIVGGNPFKLFEDPTSGVSVLVGKLTQKHYGLHRLVVQVNDSGQPSQSTTTLVHVFVNESVSN  
ATAIDSQIARSLHIPLTQDIAGDPSYEISKQRLSIVIGVVAGIMTVILILIVVMARYCRSKNKN  
GYEAGKKDHEDFFTPQQHDKSKKPKDKKKNKSKQPLYSSIVTVEASKPNGQRYDSVNEKLSDSP  
SMGRYRSVNGGPGSPDLARHYKSSSPLPTVQLHPQSPTAGKKHQAVQDLPPANTFVGAGDNISIG  
SDHCSEYSCQTNNKYSKQMLHPYITVFG

**B** PCDH7 L2 S1-1

TATDRDATDRPQRHRALRAAPGARRRRQRRREPARRGGRQRPLPRGRRRERERRRRLGRLQAAAGR  
IRGRRRHQPRRPQQRVRAAGGGHPGRREAAAADREGGAGPRAARLLRADPASARRRRPASLIAGH  
PTGPHRRERQPPLEERVRGRLG\*

**C** PCDH7 L2 S1-2

TATDRDATDRPQRHRALRAAPGARRRRQRRREPARRGGRQRPLPRGRRRERERRRRLGRLQAAAGR  
IRGRRRHQPRRPQQRVRAAGGGHPGRREAAAADREGGAGPRAARLLRADPASARRRRPASLIAGH  
PTGPHRRERQPPLEERVRGRLG\*

**D** PCDH7 L2 S2-1

Same as reference protein sequence

**E** PCDH7 L2 S2-2

TATDRDATASSRNGIERYELLQEPGGGGSGGESRRAGAADSAPYPGGGGNGASGGGSGGSKRRLD  
ASEGGGGTNPGRSSVFELQVADTPDGEKQPQLIVKGALDREQRDSYELTLRVRDGGDPPRSSQA  
ILRVLITDVNDNSPRFEKSVYEADLAENS

**F** PCDH7 L2 S2-3

Same as reference protein sequence

**G** PCDH7 L2 S2-4

TATDRDATASSRNGIERYELLQEPGGGGSGGESRRAGAADSAPYPGGGGNGASGGGSGGSKRRLD  
ASEGGGGTNPGRSSVFELQVADTPDGEKQPQLIVKGALDREQRDSYELTLRVRDGGDPPRSSQA  
ILRVLITDVNDNSPRFEKSVYEADLAENS

**Figure 3.7 HEK-293 PCDH7 Line 2 contains multiple non-functional copies of protocadherin 7.** Predicted amino acid sequence of protocadherin 7 based on the human reference genome, protein accession number NP\_002580.2, (isoform *a*, encoded by transcript variant *a*) from UCSC genome browser (**A**). The amino acid sequence encoded by the targeted exon is highlighted in grey and the region encoded by the DNA sequence targeted by the CRISPR sgRNA is in bold. Amino acid sequences from the target exon based on PCR product sequences from HEK-293 PCDH7 Line 2 sample 1-1 (**B**), sample 1-2 (**C**), sample 2-1 (**D**), sample 2-2 (**E**), sample 2-3 (**F**) and sample 2-4 (**G**). Sample refers to the insert of a plasmid from a single bacterial colony transformed with the pGEM-T Easy vector ligated with PCR product from a CRISPR targeted region or wild type sequence. The sequences presented in B-G start at the amino acids encoded by the DNA CRISPR target site. Underlined letters indicate positions where the amino acid sequence is different to the wild type (due to an insertion or a frameshift) and asterisks indicate premature stop codons. Experiment performed once.

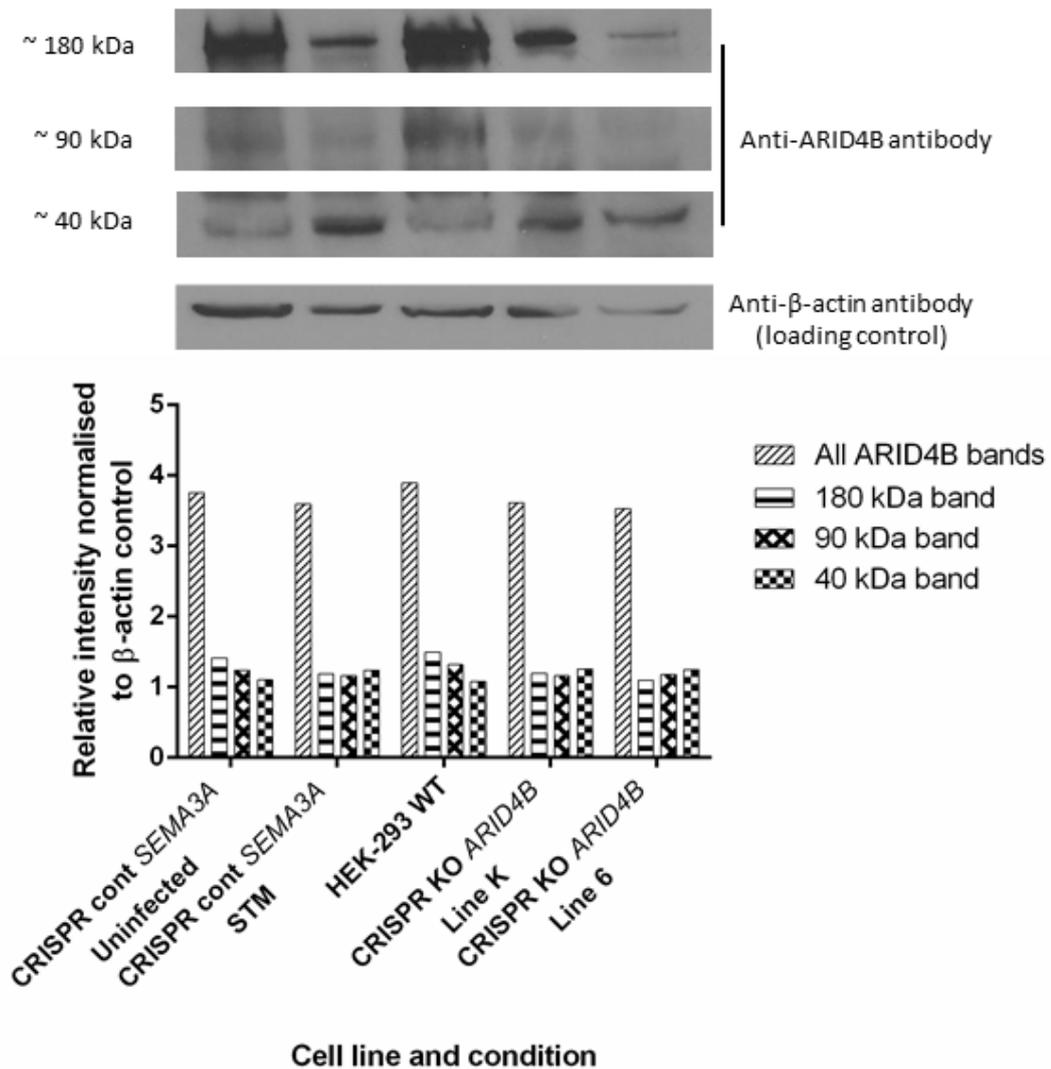
### 3.2. Failure to observe significant effects on protein expression in CRISPR *ARID4B* HEK-293 cell lines

A CRISPR Cas9 generated *ARID4B* mutant HEK-293 cell line (Line 6) was created in the same manner as the CRISPR KO cell lines in Sections 3.1.1, 3.1.2 and 3.1.3. However, several attempts to genetically characterise the line stalled due to inefficient target locus PCR amplification (complete lack of correct PCR product). Five different primer pairs were tested, but no product from the locus was generated. Annealing temperature gradients were tested, as well as different PCR machines, increasing the number of cycles and increasing the length of the final elongation. Some of this investigation was spent amplifying and optimising for a non-specific product of similar size.

Therefore, a different route was taken to confirm the KO status of the cell line – assessment of the *ARID4B* protein expression level (Figure 3.8) and cellular localisation (Figure 3.9) in Line 6 as well as a further cell line (Line K). A central nervous system gene thought not to be associated with infection (*SEMA3A*) was chosen to target using CRISPR for use as a CRISPR ‘sham’ control. No clonal cell lines were isolated for this control; the CRISPR-treated cells were instead used as a mixed pool of putative mutants. The CRISPR KO control *SEMA3A* was additionally infected with STM (according to the bacterial infection protocol in Section 2.8) after which total protein was extracted 24 hours post gentamicin maintenance addition to assess any differences in *ARID4B* protein expression levels to uninfected CRISPR KO control *SEMA3A* (Figure 3.8).

Multiple protein bands per cell line were anti-human *ARID4B* antibody positive and present at varying intensities (Figure 3.8, top). As the anti-human  $\beta$ -actin antibody loading control also produced bands of differing intensity, the intensity of each anti-human *ARID4B* antibody band was normalised to the respective loading control and plotted (Figure 3.8, bottom). Once normalised, the intensities of the bands for the potential KO Lines K, and 6, were comparable to CRISPR control *SEMA3A* and wild type. This suggests that on average

both Lines K and 6 express wild type levels of ARID4B protein. STM infection had no effect on ARID4B protein expression levels.

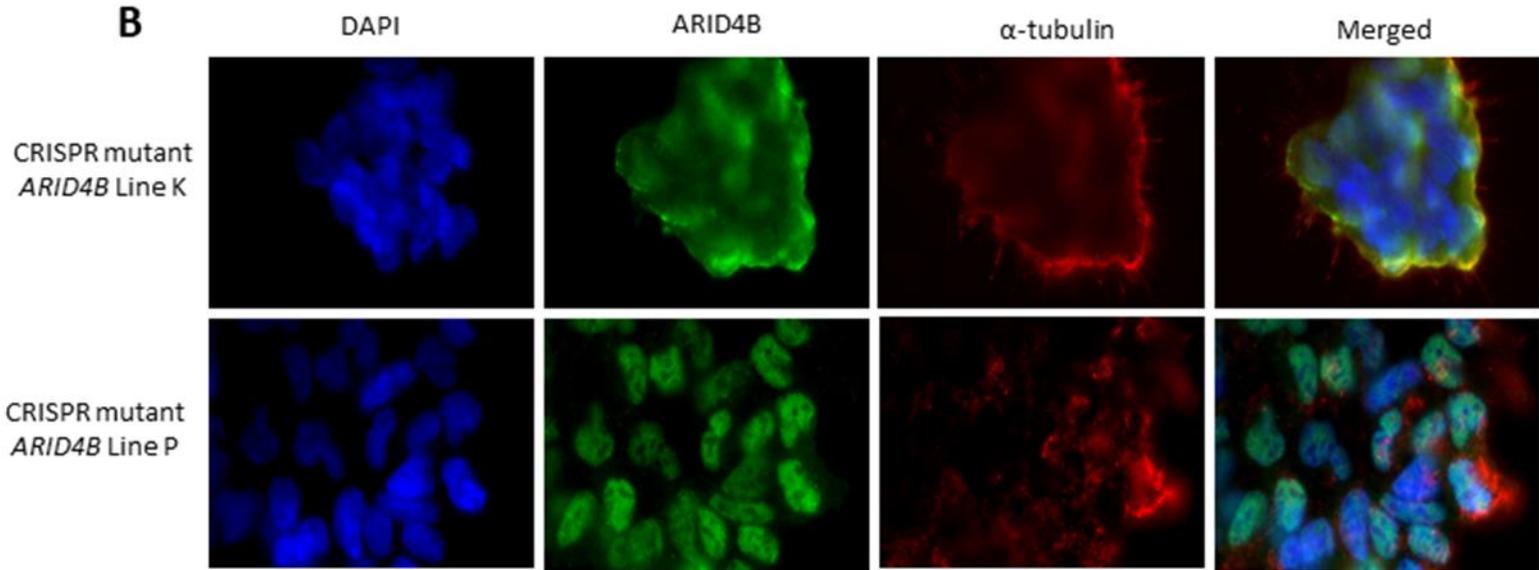
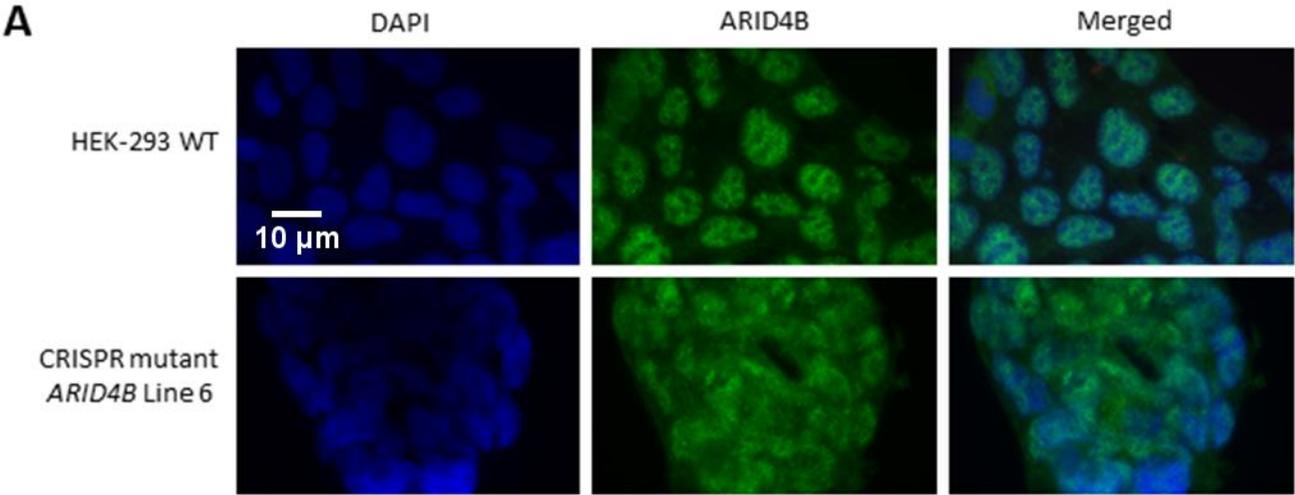


**Figure 3.8 HEK-293 CRISPR KO *ARID4B* Lines K and 6 have comparable *ARID4B* protein expression to CRISPR control (cont) *SEMA3A* and HEK-293 wild type (WT).** *ARID4B* protein expression in HEK-293 CRISPR KO *ARID4B* cell lines K, 6, CRISPR control (cont) *SEMA3A* (uninfected or STM-infected) and wild type (WT) stained with anti-*ARID4B* antibody on a western blot. Western blot band intensities were measured in ImageJ and plotted relative to the respective  $\beta$ -actin band. Experiment performed once.

The presence of *ARID4B* protein in Lines 6 and K may be the result of gene mutations that do not prevent protein production but still alter functionality, or simply a complete lack of amino acid changing (nonsynonymous) mutations. To test whether the protein in these CRISPR *ARID4B* cell lines might have altered protein localisation (an example of aberrant function), a visual assessment of the presence and localisation of the *ARID4B* protein across cultured cells of the cell line populations was performed using immunofluorescence microscopy. To facilitate this, HEK-293 wild type and potential KO *ARID4B* Line 6 were fixed and stained with the same anti-*ARID4B* antibody before staining with an Alexa Fluor 488

secondary antibody (Figure 3.9A). The antibody stained the protein well in wild type cells and was localised specifically to the nucleoplasm. In contrast the protein in potential KO *ARID4B* Line 6 is spread throughout the nucleus.

The presence and localisation of ARID4B was also tested in potential KO *ARID4B* Line K and Line P (a third CRISPR mutant cell line), which were stained with anti-ARID4B antibody (Alexa Fluor 488 secondary) and anti- $\alpha$ -tubulin antibody (Alexa Fluor 594 secondary) before imaging by microscopy (Figure 3.9B). Line P demonstrated wild type-like ARID4B localisation, but is presented to contrast with the abnormal ARID4B localisation and morphology seen in Line K. Potential KO *ARID4B* Line K cells appeared fused together in clusters, with little to no ARID4B staining in the nucleus, instead residing in dense regions in the cytoplasm. Presence and localisation of ARID4B was observed for 5 other potential KO *ARID4B* lines by immunofluorescence but none of these displayed any differences to wild type cells. Therefore, none of the CRISPR mutant *ARID4B* cell lines are KOs, but Lines K and 6 are sufficiently mutated to generate abnormal cellular localisation. However, these cell lines were dropped from further enquiry as it was felt that the mutations may not be strong enough to affect bacterial burden. The morphological and protein localisation changes described here are summarised in Table 3.1.

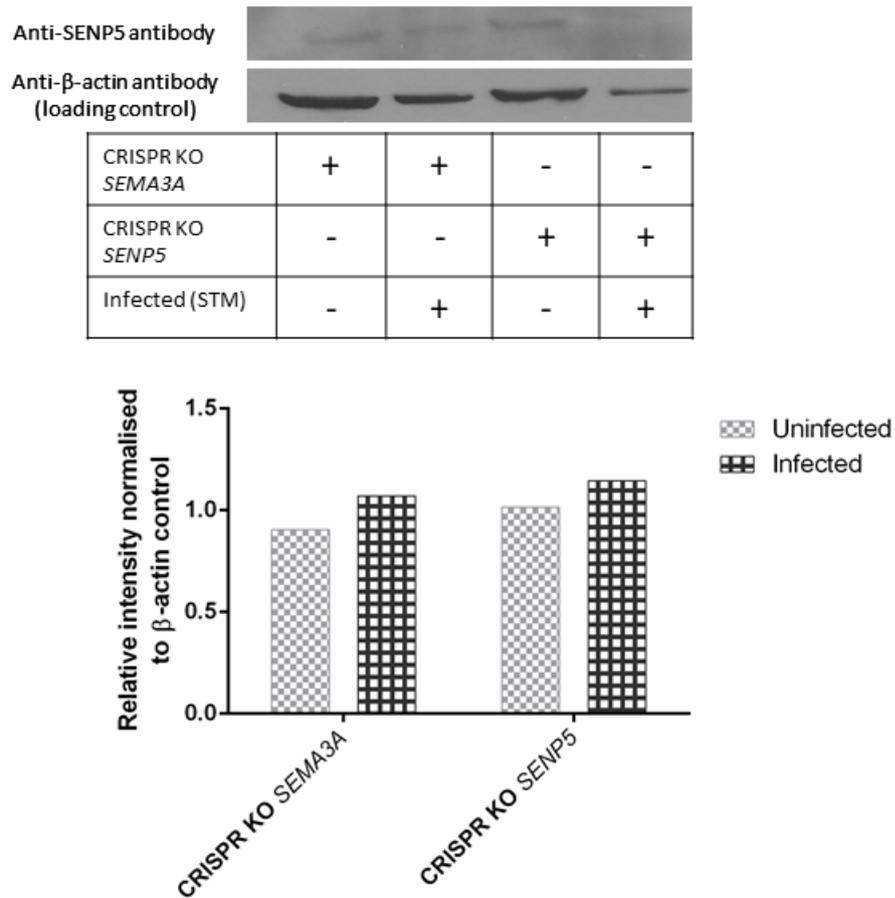


**Figure 3.9 ARID4B Lines 6 and K show different patterns of protein expression compared to wild type cells, which may indicate the presence of mutations.** HEK-293 wild type (WT), potential CRISPR KO ARID4B Line 6 **A**, or potential CRISPR KO ARID4B Lines K and P **B**, were stained with anti-human ARID4B antibody (with an Alexa Fluor 488 -green- fluorescent secondary antibody) and DAPI mountant to stain nuclei. Potential CRISPR KO ARID4B Lines K and P were additionally stained with anti- $\alpha$ -tubulin antibody (with an Alexa Fluor 594 -red- secondary antibody). All images were taken on an upright epifluorescence microscope at a total magnification of 400x. Experiment performed once.

### 3.3. Lack of altered SENP5 protein expression in CRISPR Cas9 *SENP5* targeted cell line

The SENP5 protein expression in CRISPR KO *SENP5* Line 3 was also characterised, in this case also adding STM infection as a variable. CRISPR control *SEMA3A* pool and CRISPR KO *SENP5* Line 3 were infected with STM at a MOI of 10 following the bacterial infection assay protocol in Chapter 2. Total protein samples (extracted from infected and mock-infected combinations of each at 24 hours post gentamicin maintenance addition) were loaded onto an SDS polyacrylamide gel electrophoresis (PAGE) gel before probing for SENP5 protein (Figure 3.10). The anti-SENP5 antibody stained a faint single band per lane between 80-90 kDa, which matches the theoretical molecular weight for SENP5 at 87 kDa. Staining for  $\beta$ -actin was used as a loading control, but the  $\beta$ -actin band intensity was much less in infected samples, as might be expected, given the likelihood of increased cell death. To adjust for the inconsistent  $\beta$ -actin band intensity, the band intensities were measured in ImageJ and the intensity of the SENP5 bands were plotted relative to the respective  $\beta$ -actin band. Based on the relative band intensities, SENP5 protein expression does not change after infection or CRISPR mutation. The presence of protein despite predicted protein truncation or mRNA nonsense mediated decay may be the result of multiple scenarios which are described in Section 3.5.4.

Due to the failures to observe distinct reductions in the protein expression in SENP5 or ARID4B in the respective CRISPR Cas9 cell lines, none were taken forward for phenotypic infection characterisation.



**Figure 3.10 CRISPR KO *SENP5* Line 3 has similar SENP5 protein expression to CRISPR control *SEMA3A* regardless of STM-infected status.** *SENP5* and  $\beta$ -actin loading control protein from STM-infected (cross-hatched) or uninfected (squares) HEK-293 CRISPR control *SEMA3A* pool and CRISPR KO *SENP5* Line 3 stained with anti-*SENP5* antibody or anti- $\beta$ -actin antibody on a western blot. Total protein used. The band stained with the anti-*SENP5* antibody is between 80-90 kDa. Western blot band intensities were measured in ImageJ and plotted relative to the respective  $\beta$ -actin band. Experiment performed once.

**Table 3.1 List of CRISPR cell lines created, isolated and genetically characterised for the genes of interest from the HEK-293 screen.** *Neomycin resistant CRISPR Cas9 treated isolated cell lines from a HEK-293 background. \*Refers to the antibiotic selected E. coli transformed with pGEM-T Easy plasmids containing PCR products from the relevant cell line listed. SENP5 Line 3, PRKG1 Line 12 and PCDH7 Line 2 were taken forward. PRKG1 Line 11 S1-3 – insertion may be due to Taq polymerase, which has a low reported insertion rate (McInerney et al., 2014; Potapov and Ong, 2017).*

Gene targeted	Cell line	Bacterial colony*	Mutation summary
SENP5	Line 3	S1-1	14 nucleotides deleted, frameshift, premature stop codon
		S1-2	15 nucleotides inserted including stop codon
		S1-3	9 nucleotides deleted, no frameshift
		S1-4	Missing one base, frameshift, premature stop codon and 2 point mutations
		S1-5	6 nucleotides inserted, no frameshift
PRKG1	Line 10	S1-1	No mutation
		S1-2	2 single nucleotide changes
		S1-3	No mutation
	Line 11	S1-1	31 nucleotides deleted, frameshift, premature stop codon
		S1-2	30 nucleotides deleted, no frameshift
		S1-3	1 nucleotide inserted towards end of sequence, outside of target site
		S1-4	16 nucleotides deleted, frameshift, premature stop codon
	Line 12	S1-1	28 nucleotides deleted, frameshift, premature stop codon
		S1-2	16 nucleotides deleted, frameshift, premature stop codon

Gene targeted	Cell line	Bacterial colony*	Mutation summary
<i>PRKG1</i> (continued)	Line 12 (continued)	S1-3	8 nucleotides deleted, frameshift, premature stop codon
	Line 13	S1-1	2 nucleotide changes
		S1-2	1 nucleotide change
		S1-3	2 nucleotide changes
		S1-4	1 nucleotide insertion near the end of the sequence
		S1-5	4 nucleotide changes
<i>PCDH7</i>	Line 1	S1-1	No mutations
		S1-2	3 nucleotide changes
		S1-3	No mutations
	Line 2	S1-1	8 nucleotides inserted in place of 1 nucleotide, frameshift, premature stop codon
		S1-2	8 nucleotides inserted in place of 1 nucleotide, frameshift, premature stop codon
		S2-1	No mutation
		S2-2	13 nucleotides inserted; 4 nucleotides replaced, 2 amino acids replaced with 5 others, no frameshift
		S2-3	No mutation
		S2-4	13 nucleotides inserted, 4 nucleotides replaced, 2 amino acids replaced with 5 others, no frameshift
	Line 3	S1-1	No mutations
		S1-2	2 nucleotide changes
		S1-3	No mutations
<i>PSMA1</i>	Line 1	S1-1	2 nucleotide changes
		S1-2	2 nucleotide changes

Gene targeted	Cell line	Bacterial colony*	Mutation summary
<i>PSMA1</i> (continued)	Line 1 (continued)	S1-3	2 nucleotide changes
	Line 2	S1-1	No mutations
		S1-2	No mutations
		S1-3	6 nucleotide changes
	Pool	S1-1	2 nucleotide changes
		S1-2	3 nucleotide changes
		S1-3	5 nucleotide changes
		S1-4	3 nucleotide changes
		S1-5	2 nucleotide deletions, resulting in a frame shift and a premature stop codon.
		S1-6	2 nucleotide changes
<i>ARID4B</i>	Line 1	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)
	Line 4	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)
	Line 5	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)
	Line 6	N/A	ARID4B protein found throughout the nucleus, comparable protein expression to wild type
	Line K	N/A	ARID4B protein in cytoplasm, comparable protein expression to wild type (nucleoplasm specificity)
	Line N	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)
	Line P	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)
	Line R	N/A	Wild type-like localisation of ARID4B protein (nucleoplasm specificity)

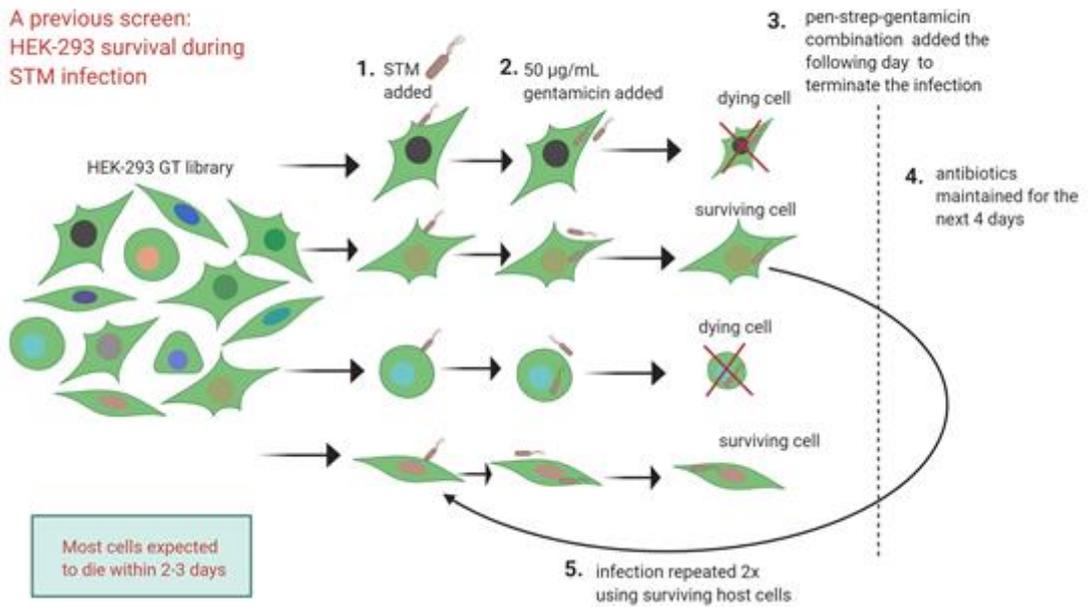
### 3.4. Phenotypic characterisation of *PCDH7* and *PRKG1* CRISPR KO cell line infection by *Salmonella* and *Shigella*

Assays to measure infection resistance through host cell resistance to death (to replicate the original screen) or bacterial burden (as an alternative readout of bacterial resistance) were performed. Bacterial burden can provide additional indications of why a mutation might promote increased or decreased host survival, such as antibacterial activity, reduced bacterial replication or restricted access to the intracellular space.

A timeline of the bacterial infection assay and other procedures used in this chapter is shown in Figure 3.11B. The CRISPR KO 'sham' control described in Section 3.2 is used in infection experiments throughout this section.

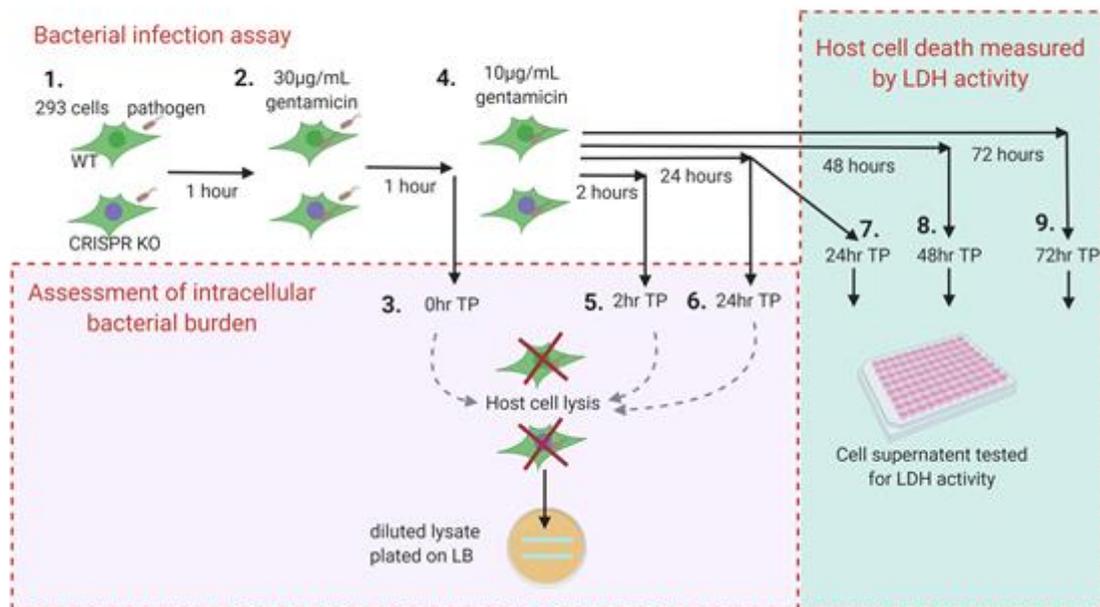
**A**

A previous screen:  
HEK-293 survival during  
STM infection



**B**

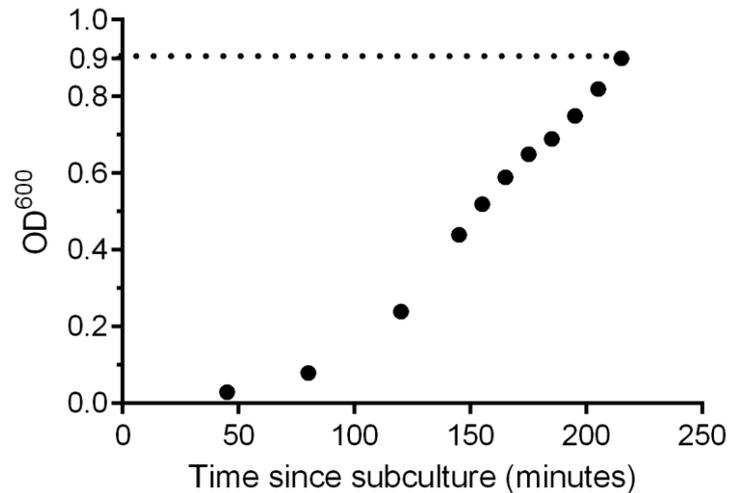
Bacterial infection assay



**Figure 3.11 Comparison of the protocols that were used to isolate, and subsequently validate, infection-resistant genes in HEK-293 cells.** Summary of the previous HEK-293 GT library selection screen using STM as the infectious agent (A) and of the bacterial infection assay used to compare bacterial burden in HEK-293 CRISPR KO cell lines with wild type and sham control cells (B). **A, 1.** HEK-293 GT library cells were inoculated with STM, before **2.** supplementing with 50 µg/mL gentamicin one hour later. **3.** A penicillin, streptomycin, gentamicin combination added a day later to begin the termination of the experiment. **4.** Antibiotic supplementation maintained in culture for the next 4 days. Most HEK-293 GT library cells are expected to die within 2-3 days of STM inoculation. **5.** Upon elimination of remaining bacteria, surviving cells were reinfected 2 more times. **B, 1.** HEK-293 CRISPR KO cells and wild type (WT) cells were inoculated with an intracellular pathogen (e.g. STM). **2.** An hour later, 30 µg/mL gentamicin was added to kill extracellular bacteria. **3.** To assess bacterial burden, after a further hour the designated 0-hour time point was processed: the host cells were lysed, and diluted lysate was plated to generate CFU data. **4.** A maintenance concentration of gentamicin (10 µg/mL) was added to the remaining cells. **5.** After 2 hours the 2-hour time point was processed. **6.** The 24-hour time point was processed 24-hours after addition of gentamicin maintenance media. **7, 8, 9.** In an alternative assay, measuring host cell death, LDH activity was quantified in cell supernatant 24, 48 and 72 hours after addition of gentamicin maintenance media. Further details on the bacterial infection assay, bacterial burden and LDH activity assay are in Chapter 2.

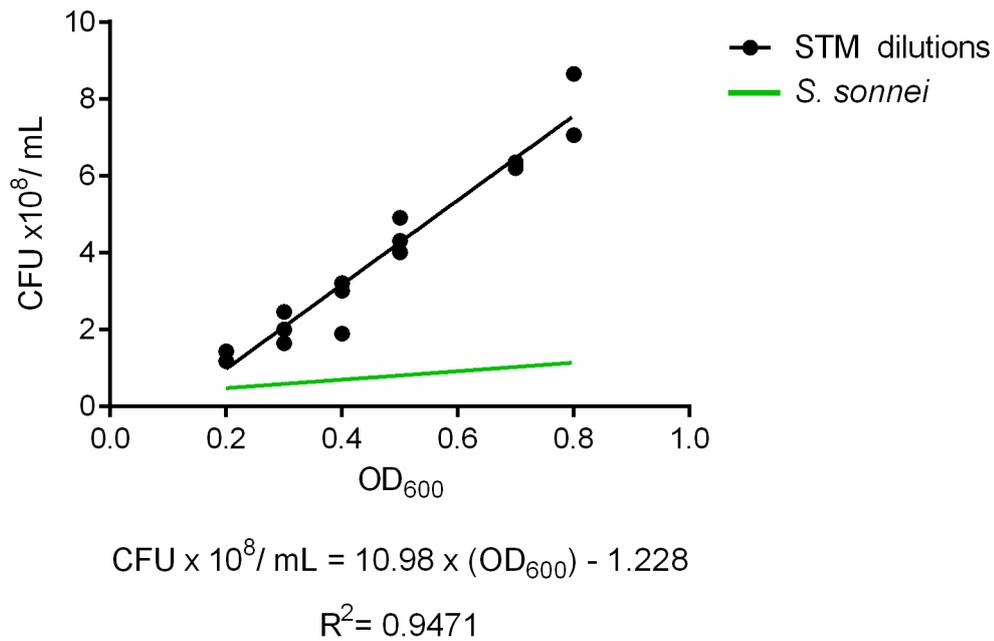
#### 3.4.1. Determining accurate multiplicity of infection for *S. enterica* ser. Typhimurium: correlation of optical density measurements with colony forming units

In preparation for the candidate gene validation study, it was necessary to optimise infection parameters. Estimates of bacterial number in a liquid culture can be made by diluting the culture and plating it on agar to generate CFU, which may be counted and extrapolated to give the original bacterial number. Estimating concentration from the OD is a much quicker alternative but requires an empirically derived correlation between OD and CFU. Such a correlation has previously been determined in the laboratory for *S. sonnei* by Dr Rasha Mahmoud, but STM is a different size and thus requires its own specific equation. Two STM subcultures (diluted 1:100 from overnight cultures) were grown at 37°C over a period of 215 minutes until a final OD<sub>600</sub> reading of 0.9 was reached. A third replicate experiment was performed on a different day. In parallel, to determine the phase of growth at an OD<sub>600</sub> of 0.9, a growth curve of OD<sub>600</sub> values against time was created from one of the subcultures (Figure 3.12). The growth curve shows that at an OD<sub>600</sub> reading of 0.9, the subculture is still in the exponential growth phase (or potentially on the verge of entering stationary phase), meaning that the ratio of living to dead bacteria will be similar to that at OD<sub>600</sub> 0.3 and 0.6. OD<sub>600</sub> readings of 0.3 and 0.6 are important as the original screen used a STM subculture at OD<sub>600</sub> 0.3 and experiments in other chapters were performed at OD<sub>600</sub> 0.6.



**Figure 3.12 STM growth in subculture, as measured by OD<sub>600</sub> over time: an OD<sub>600</sub> of 0.9 is still in exponential phase.** STM subculture in LB was prepared from a 1:100 dilution of overnight culture before shaking at 37°C until the culture reached an OD<sub>600</sub> of 0.9. The OD of the subculture was measured periodically over the course of 215 minutes to estimate the stage of growth of the culture at an OD<sub>600</sub> of 0.9. Experiment performed once and each point is representative of a measurement.

Subcultures at OD<sub>600</sub> 0.9 were diluted with LB to obtain the range of OD<sub>600</sub> readings required within  $\pm 0.02$  and serially diluted in PBS prior to plating on LB agar to generate CFU counts. The concentration of the LB-diluted subculture at each OD<sub>600</sub> reading was calculated from the CFU counts and plotted against OD<sub>600</sub> to generate a set of data points through which a line of best fit was drawn (Figure 3.13). The resulting equation for this line (shown below the figure) was therefore available for use in applying an accurate MOI in all subsequent infection experiments. A line representing the OD<sub>600</sub> to CFU correlation for *S. sonnei* (equation in Chapter 2, Figure 2.7) is also shown in Figure 3.13 for comparison. The two lines are very different to each other, with different gradients and intercepts, thus demonstrating that it was important to generate an equation for *Salmonella*.



**Figure 3.13 OD: CFU correlations for STM and *S. sonnei*.** STM subculture in LB was prepared from 1:100 dilution of overnight culture before shaking at 37°C until the culture reached an OD<sub>600</sub> of 0.9. The bacterial culture was diluted in LB to the desired OD<sub>600</sub> values (LB diluted-culture) before preparing 10-fold serial dilutions in PBS of the LB diluted-culture to put onto agar. Each black data point represents one biological replicate (experiment performed three times). A linear regression and goodness of fit was calculated in Graphpad Prism 6.01. A (green) line representing the *S. sonnei* equation (previously empirically determined) between OD<sub>600</sub> and CFU is shown for comparison.

#### 3.4.2. Survival of CRISPR *PCDH7* and *PRKG1* KO cell lines during *S. enterica* ser. Typhimurium infection as a means to validate the survival phenotype originally identified in the HEK-293 gene trap survival screen

Armed with an equation to determine the bacterial concentration of a STM subculture, a method to compare the host cell death of CRISPR KO lines to wild type HEK-293 cells under infection conditions was sought. The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay measures metabolic activity but is often used as a rough assessment of host cell survival (Cummings et al., 2004). The use of Annexin V in flow cytometry is used as a direct measure of apoptosis but does not measure other kinds of cell death. STM-infected epithelial cells are putatively thought to undergo pyroptosis as well as epithelial cell extrusion (Gudipaty and Rosenblatt, 2017; Knodler et al., 2010; Rauch et al., 2017). HEK-293 cells are not epithelial cells, but STM infection data for these cells is lacking. Therefore, the method chosen was required to be quantitative, relatively quick to use, and a direct measure of cell death. The lactate dehydrogenase (LDH) assay matched these criteria. The LDH assay relies on the release of host LDH into the culture medium as a result

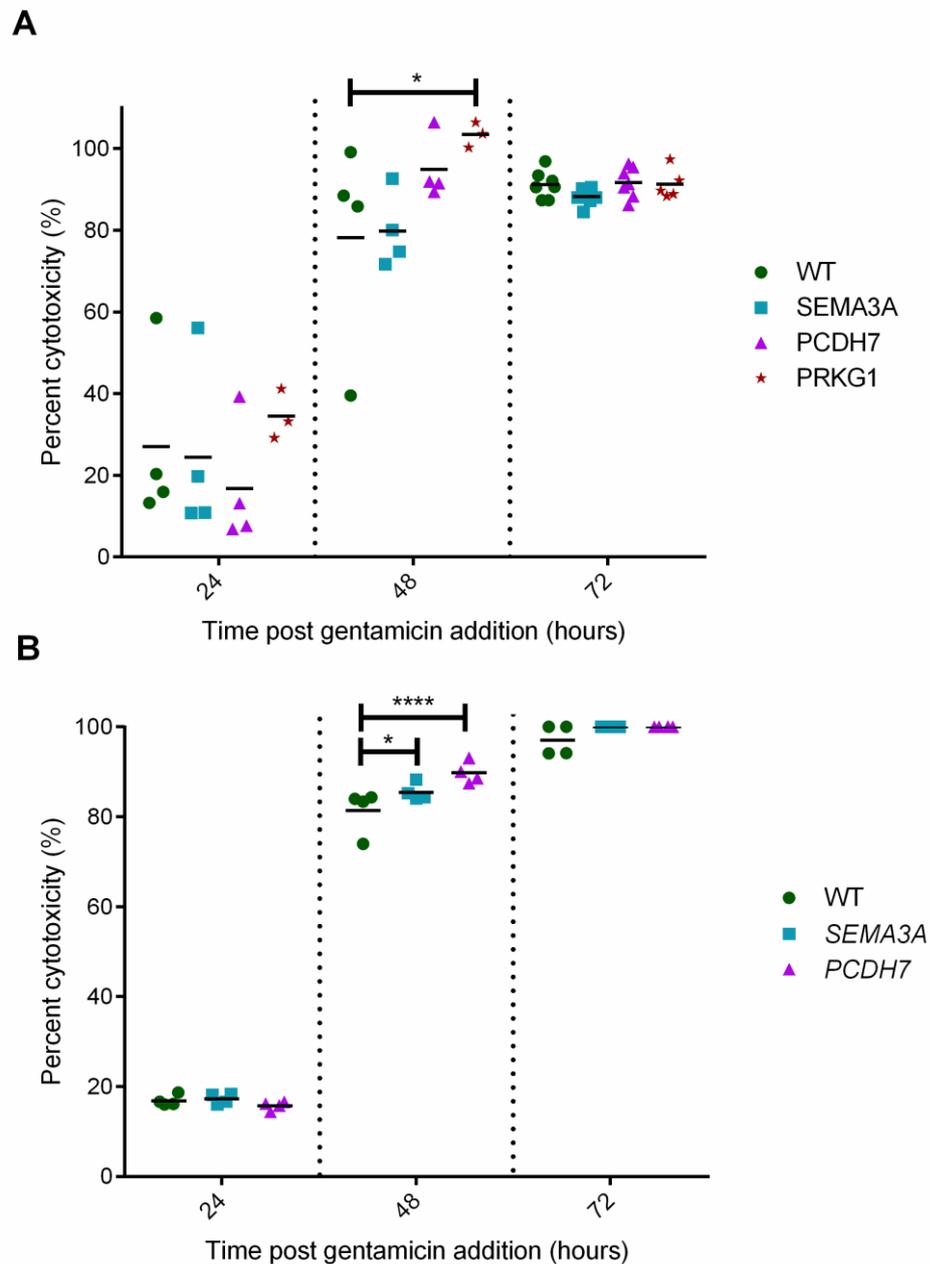
of plasma membrane permeability (characteristic of necrosis, necroptosis and pyroptosis). The enzyme activity is then measured through the addition of a synthetic substrate, which generates a coloured product in proportion to LDH released/cell death.

HEK-293 CRISPR KO *PCDH7* Line 2 and *PRKG1* Line 12, together with control cells, were inoculated with STM at a MOI of 10 (the details of the procedure are outlined in Chapter 2) to emulate the conditions of the original HEK-293 infection survival screen (described in Chapter 1). The infection in that survival screen lasted for 3 days, thus, LDH activity was measured here over the course of 72 hours (summary in Figure 3.11) as the enzyme has a half-life of 9 hours in culture medium (Riss et al., 2004). The individual measurements for experimental samples were converted into percent cytotoxicity using the equation (Figure 3.14) provided by the manufacturers. The denominator in this equation is the 'Max LDH release' which is the average of measurements from 3 wells of (wild type) cells (also infected with STM) treated with lysis buffer.

$$\frac{\text{Experimental LDH release} - \text{blank}}{\text{Max LDH release} - \text{blank}} \times 100 = \text{percent cytotoxicity (\%)}$$

**Figure 3.14 Equation to determine percent cytotoxicity provided by the Promega Cytotox96 Non-radioactive cytotoxicity assay manual.** Each of the components in the fraction are  $OD_{490}$  values determined from experimental well supernatant (Experimental LDH release), supernatant from lysis buffer treated wells (Max LDH release) or media only control (blank).

The infection and LDH assay were performed twice for *PCDH7* Line 2 (Figure 3.15A, B); but only performed once for *PRKG1* Line 12 (Figure 3.15A) due to difficulties in achieving adequate cell growth prior to seeding. *PRKG1* Line 12 has a significantly higher percentage cytotoxicity at 48 hours post gentamicin than wild type (Figure 3.15A) suggesting that it has a reduced survival advantage, though the cytotoxicity becomes comparable to wild type at 72 hours post gentamicin. It is likely that percentages higher than 100 in Figure 3.15 are the result of inadequate lysis of the 'Max LDH release' wells (which were inoculated with a comparable number of bacteria). In the first experiment (Figure 3.15A) the LDH activity nearly (or actually) reaches 100% at 48 hours for all cell lines (including wild type and CRISPR control *SEMA3A* pool) with the LDH activity decreasing slightly at 72 hours. In the second experiment (Figure 3.15B) the LDH activity reaches 80-90% at 48 hours, peaking at 72 hours. Surprisingly, the LDH activity in CRISPR KO *PCDH7* Line 2 (in Figure 3.15B) significantly surpasses that found in wild type supernatant, suggesting an increased susceptibility to infection. This pattern is mirrored in Figure 3.15A but does not reach significance. The increased LDH activity (as a measure of cell death) seen in these two cell lines is unexpected, because it was assumed that the increased survival of the *PCDH7* mutant observed in the original screen would correlate with decreased cell death here. LDH activity is also slightly (but significantly) increased in CRISPR KO *SEMA3A* cells at 48 hours compared to the wild type, in contrast to the lack of difference found in the initial experiment (Figure 3.15A).



**Figure 3.15 Infected HEK-293 CRISPR KO cell lines show increased cytotoxicity at 48 hours post gentamicin maintenance.** HEK-293 wild type (WT), CRISPR KO SEMA3A pool (the CRISPR control), CRISPR KO PCDH7 Line 2 and in **A** only, PRKG1 Line 12 were inoculated with STM at a MOI of 10. At 24-, 48- and 72- hours post gentamicin maintenance addition supernatant samples were taken to measure lactate dehydrogenase activity with the cytotoxicity assay as a proxy for host cell death. Percent cytotoxicity was calculated (according to the CytoTox96® Non-Radioactive Cytotoxicity Assay manual instructions) using the equation in Figure 3.14. Experiment performed twice (presented each in **A** and **B**). The data was analysed by two-way Analysis of variance (ANOVA) and multiple comparisons were made to wild type with Dunnett's multiple comparisons test; \*\*\*\* refers to  $p \leq 0.0001$  and \*\* to  $p \leq 0.01$ .

### 3.4.3. Assessment of the intracellular bacterial burden of CRISPR KO cell lines during infection with intracellular pathogens.

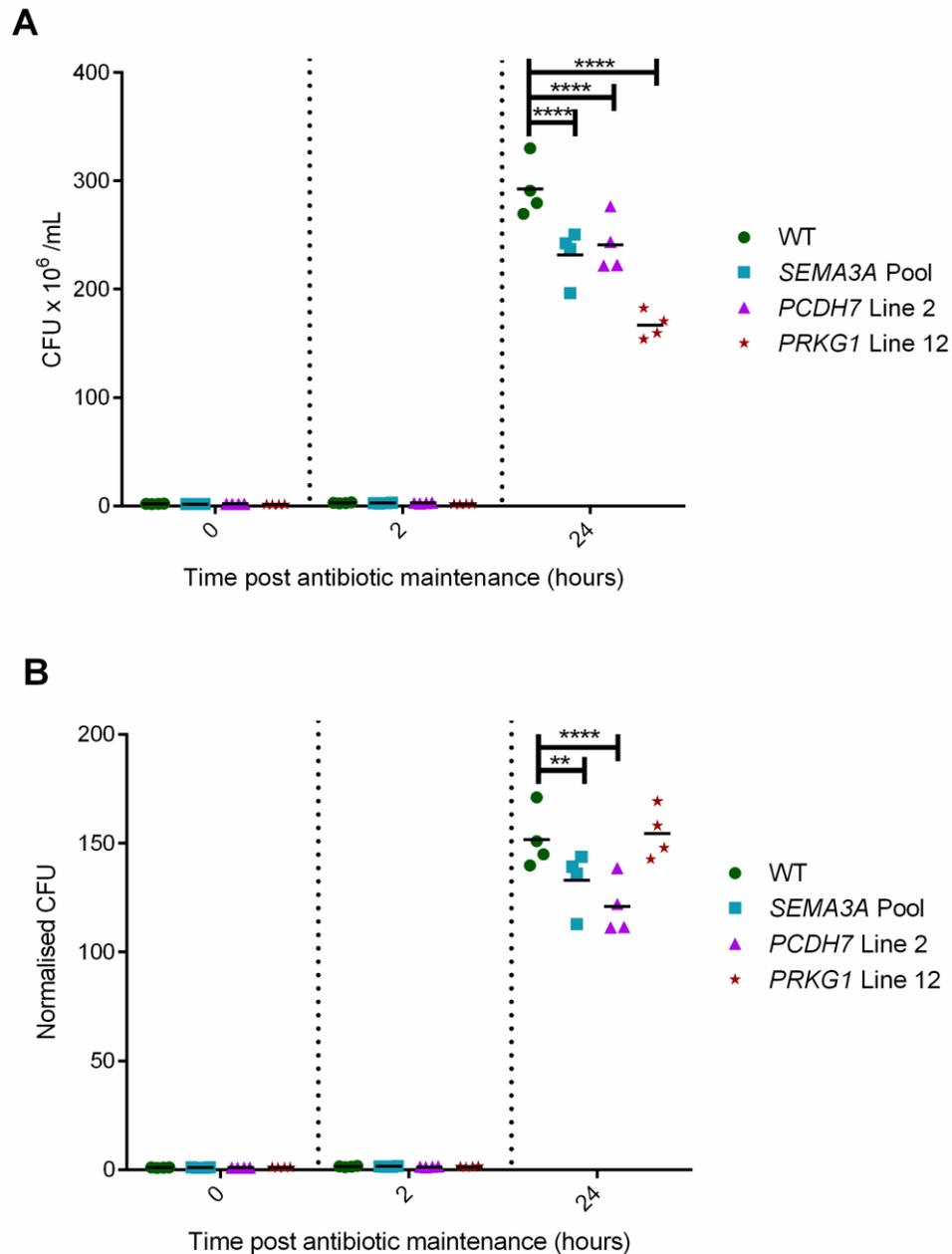
Improved host survival during infection may not be a simple resistance to cell death upon intracellular entry of the pathogen but may result from a more subtle suppression of intracellular proliferation or bactericidal activity. Indeed, phenotypes in this second category are of most interest. Simply promoting host cell survival regardless of pathogen infection may preserve a suitable replicative niche, thus allowing greater overall pathogen burden for the host organism. To better understand the underlying features of the host infection phenotypes found, intracellular bacterial burden was calculated over infection time courses with 4 bacterial species. These species were chosen from 4 orders within the class of Gammaproteobacteria to promote identification of potential 'universal' host defence processes, which would be of particular interest as therapeutic targets.

#### 3.4.3.1. *S. enterica* ser. Typhimurium infections

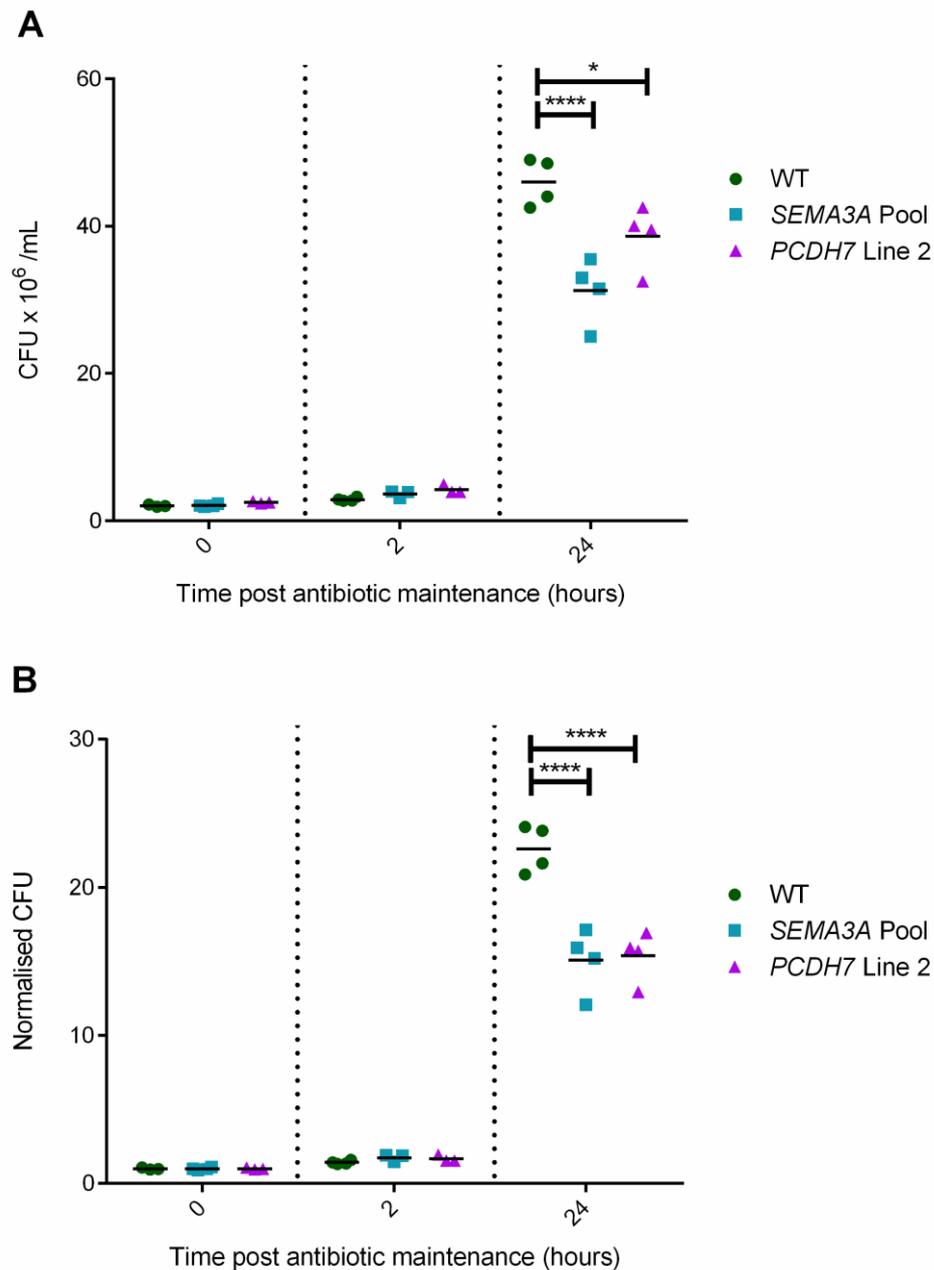
HEK-293 CRISPR KO *PCDH7* Line 2, *PRKG1* Line 12 and appropriate controls were inoculated with STM at a MOI of 70 before 1-hour treatment with a concentration of gentamicin sufficient to kill extracellular bacteria. A higher MOI was chosen than in the infection survival screen and cytotoxicity assay to produce a more pronounced effect. Thereafter a lower maintenance concentration of gentamicin was used to avoid bacterial growth (from previously lysed host cells) overtaking the culture media and obscuring the measurement of the intracellular bacterial burden. Immediately after adding the gentamicin maintenance concentration, the 0-hour time point sample was processed (timeline in Figure 3.11B). The intracellular bacterial burden was measured by plating the serially diluted lysate from infected wells and enumerating the colony forming units (CFU). In infection experiment 1, the intracellular STM burden in *PRKG1* Line 12 cells was slightly lower at 0- and 2- hours post gentamicin maintenance than in wild type or *SEMA3A* pool control cells, though this did not attain significance (Figure 3.16A). The difference in intracellular burden increased by the 24-hour time point, gaining significance compared to the burden in wild type cells. CRISPR KO *PCDH7* Line 2 cells also had significantly fewer intracellular STM by 24-hours post gentamicin maintenance than wild type cells (Figure 3.16A). Concerningly, the CRISPR 'sham' control KO *SEMA3A* had a significantly lower bacterial burden than the wild type control (Figure 3.16), the potential causes of this are discussed in Section 3.5.12 in the Discussion. Interestingly, when the bacterial burden was normalised per cell line to 0-hour levels (Figure 3.16B), the difference between wild type and CRISPR KO *PRKG1* Line 12 disappears; indicating that the mutation may be most important in controlling STM infection prior to the 0-hour time point. Use of a '0-hour' time point to normalise subsequent bacterial burden has precedent in the study of *Listeria monocytogenes* (another facultative intracellular bacterium) and STM infection (Drecktrah et al., 2006; Drevets et al., 1994).

The decrease in bacterial burden in CRISPR KO *PCDH7* Line 2 compared to wild type was replicated in the second infection experiment with HEK-293 wild type, CRISPR KO *SEMA3A* pool and *PCDH7* Line 2 cells (Figure 3.17). CRISPR KO *PRKG1* Line 12 was not included in the second experiment due to insufficient cell growth prior to seeding. Without further replicate experiments, the differences seen in CRISPR KO *PCDH7* Line 2 and *PRKG1* Line 12

cells are inconclusive, though promising. These experiments, and future work exploring these phenotypes, are discussed in Section 3.5.



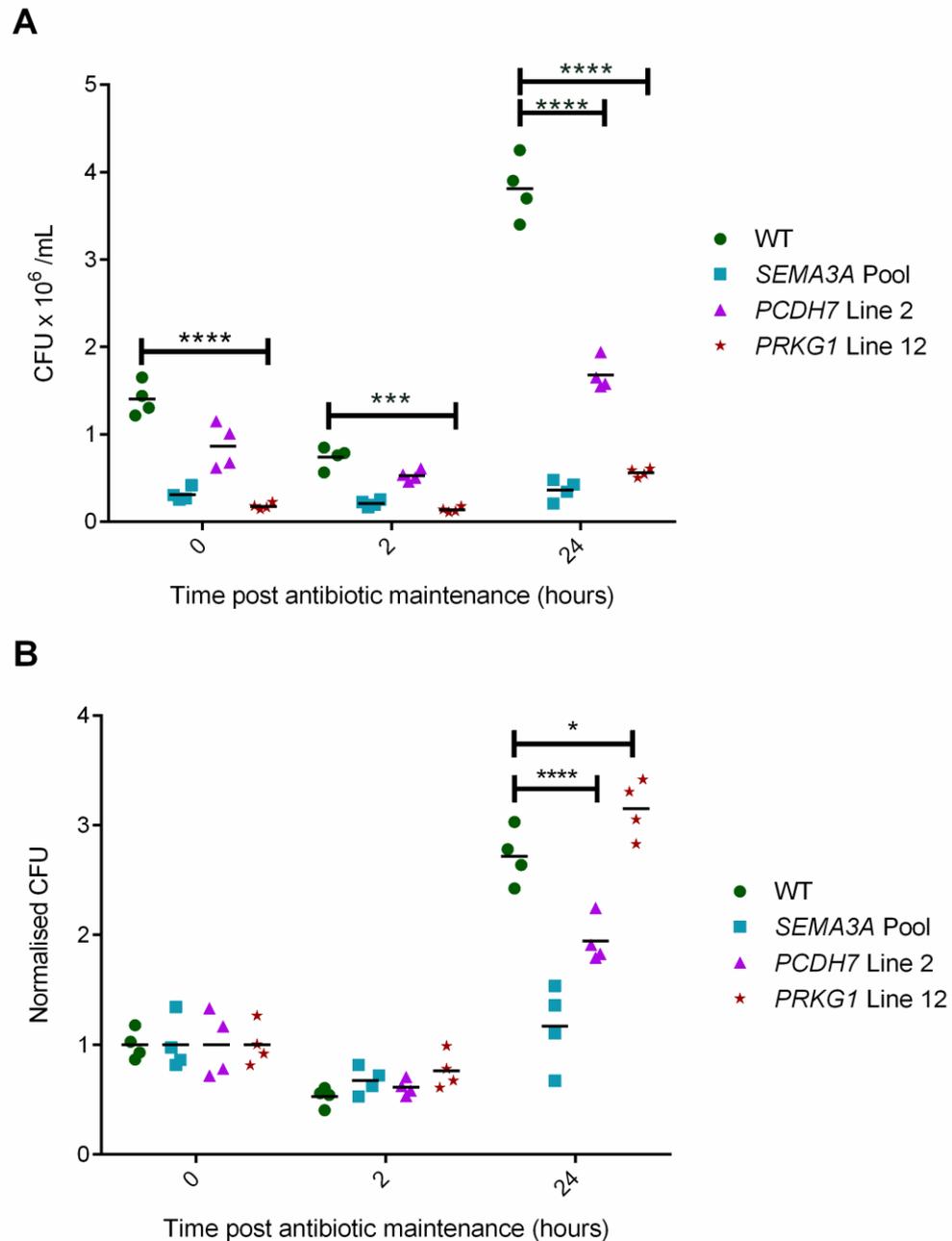
**Figure 3.16 Infection experiment 1 - HEK-293 CRISPR KO *PRKG1* Line 12 has significantly reduced intracellular STM burden at 24 hours; a difference that disappears when normalised to 0-hour burdens.** *Salmonella enterica serotype Typhimurium SL1344* intracellular burden of HEK-293 CRISPR KO cell lines at 0-hours, 2-hours and 24-hours post gentamicin maintenance as determined by CFU (protocol in Figure 3.11B). HEK-293 cells inoculated with a MOI of 70 before treating with a bacterial killing and maintenance concentration of gentamicin at consecutive hourly intervals (more detailed methodology found in Chapter 2). Data are presented as CFU x 10<sup>6</sup>/mL (A) or normalised to the average 0-hour CFU per cell line (B). The data were analysed by two-way ANOVA and multiple comparisons were made to wild type with Dunnett's multiple comparisons test; \*\*\*\* refers to  $p \leq 0.0001$  and \*\* to  $p \leq 0.01$ . This is experiment 1 of 2, the data points are representative of intra-experiment replicates and the bars show the mean.



**Figure 3.17 Infection experiment 2 - HEK-293 CRISPR KO *PCDH7* Line 2 has significantly reduced intracellular STM burden at 24 hours; a difference which is retained when normalised to 0-hour burdens.** *Salmonella enterica serotype Typhimurium SL1344* intracellular burden of HEK-293 CRISPR KO cell lines at 0-hours, 2-hours and 24-hours post gentamicin maintenance (protocol in Figure 3.11B). HEK-293 cells inoculated with a MOI of 70 before treating with a bacterial killing and maintenance concentration of gentamicin at consecutive hourly intervals. The data were analysed by two-way ANOVA and multiple comparisons were made to wild type with Dunnett's multiple comparisons test; \*\*\*\* refers to  $p \leq 0.0001$  and \* to  $p \leq 0.05$ . This is experiment 2 of 2, the data points are representative of intra-experiment replicates and the bars show the mean.

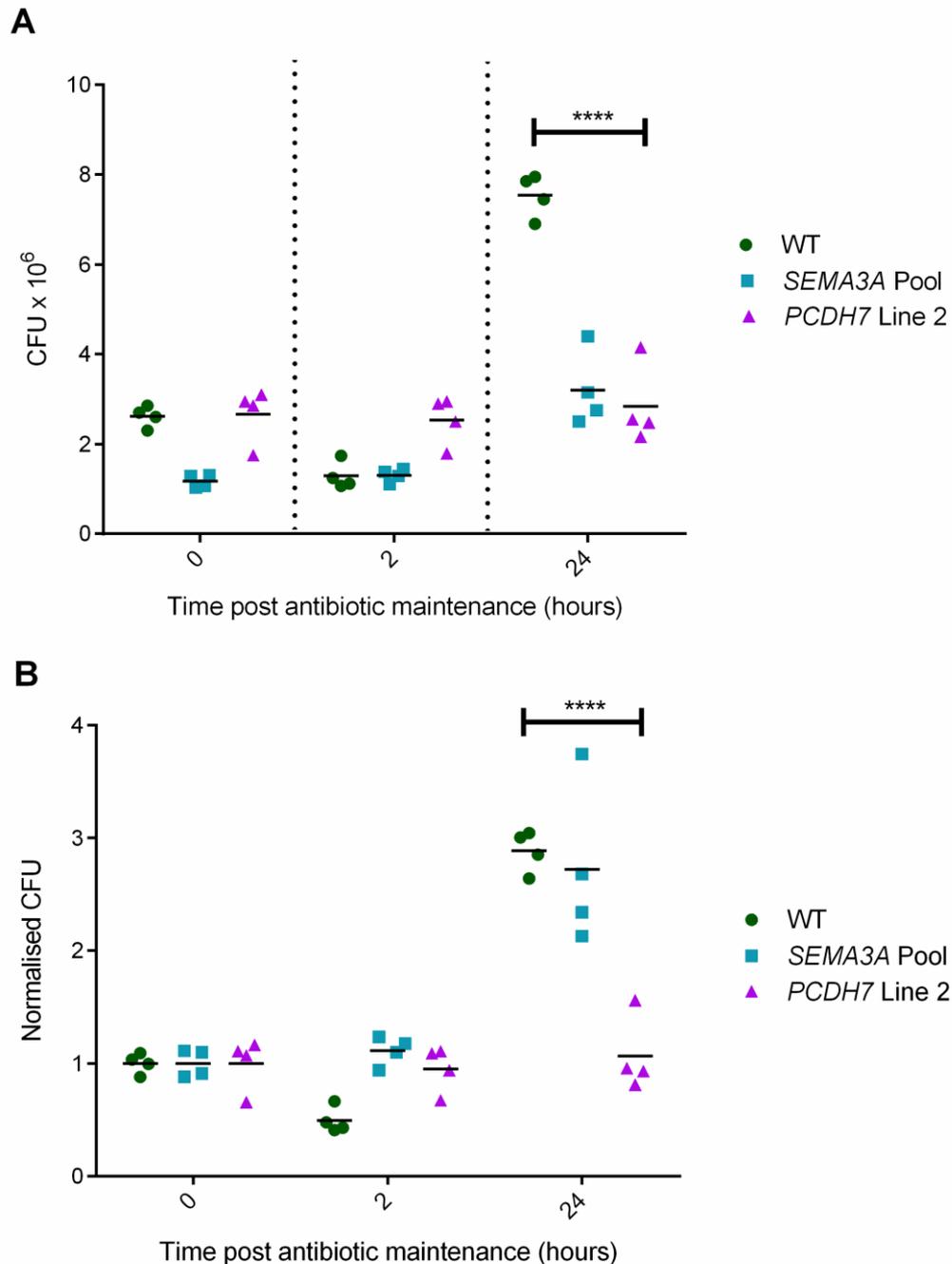
#### 3.4.3.2. *Shigella sonnei* infections

To assess the breadth of the effect of the CRISPR-induced candidate gene mutations in response to infection, bacterial burden was also measured when cells were inoculated with *S. sonnei*. In infection experiment 1, HEK-293 wild type, CRISPR KO *SEMA3A* pool, *PCDH7* Line 2 and *PRKG1* Line 12 were inoculated with *S. sonnei* at a MOI of 10 (Figure 3.18), as described in Chapter 2. The intracellular *S. sonnei* burden was significantly less in CRISPR KO *PRKG1* Line 12 than in wild type cells at all time-points (Figure 3.18A). At 24 hours post gentamicin maintenance (see Figure 3.11B), CRISPR KO *PCDH7* Line 2 also had significantly less intracellular bacteria than wild type cells (Figure 3.18A). However, the CRISPR KO *SEMA3A* pool proved problematical as a CRISPR control in *S. sonnei* infection; in fact, it had the lowest bacterial burden of any of the cell lines (except for CRISPR KO *PRKG1* Line 12) at all time-points. When the CFU were normalised to the 0-hour data per cell line, the bacterial burden in CRISPR KO *PCDH7* Line 2 remained significantly lower than in wild type cells at 24 hours (Figure 3.18B). This suggests that a large part of the effect of the *PCDH7* mutation on the infection occurs later, perhaps during bacterial multiplication. In contrast, the burden in CRISPR KO *PRKG1* Line 12 became significantly higher than in wild type cells at 24-hours.



**Figure 3.18 Infection experiment 1 - HEK-293 CRISPR KO *PRKG1* Line 12 and *PCDH7* Line 2 have significantly lower intracellular *S. sonnei* burdens than wild type (WT) at 24 hours; a difference that is inverted for HEK-293 CRISPR KO *PRKG1* Line 12 when normalised to 0-hour burdens. *S. sonnei* strain 86 intracellular burden of HEK-293 CRISPR KO cell lines at 0-hours, 2-hours and 24-hours post gentamicin maintenance. HEK-293 cells inoculated with a MOI of 10 before treating with a bacterial killing and maintenance concentration of gentamicin at consecutive hourly intervals (more detailed methodology found in Chapter 2, timeline in (Figure 3.11B)). The data were analysed by two-way ANOVA and comparisons were made to wild type with Dunnett's multiple comparisons test; \*\*\*\* refers to  $p \leq 0.0001$ , \*\*\* to  $p \leq 0.001$  and \* to  $p \leq 0.05$ . This is experiment 1 of 2, the data points are representative of intra-experiment replicates and the bars show the mean.**

In the second experiment (Figure 3.19A), a significantly lower bacterial burden was once again seen for CRISPR KO *PCDH7* Line 2 cells at 24-hours compared to the wild type burden. The bacterial burden in CRISPR KO *PCDH7* Line 2 cells remained low compared to wild type cells at 24 hours when the CFU were normalised to the CFU at 0 hours per cell line (Figure 3.19B). CRISPR KO *SEMA3A* demonstrated the same pattern of (low) bacterial burden in the second experiment (Figure 3.19A) as that seen in the first (Figure 3.18A). CRISPR KO *PRKG1* Line 12 cells were not used in this replicate experiment due to insufficient cell growth prior to seeding. The differences found in bacterial burden for CRISPR KO *PCDH7* Line 2 and *PRKG1* Line 12 are consistent with the patterns found in STM infection.



**Figure 3.19 Infection experiment 2 - HEK-293 CRISPR KO *PCDH7* Line 2 has significantly reduced intracellular *S. sonnei* burden at 24 hours compared to wild type (WT); a difference which is retained when normalised to 0-hour burdens. *S. sonnei* strain 86 intracellular burden of HEK-293 CRISPR KO cell lines at 0-hours, 2-hours and 24-hours post gentamicin maintenance. HEK-293 cells inoculated with a MOI of 10 before treating with a bacterial killing and maintenance concentration of gentamicin at consecutive hourly intervals (Figure 3.11B). The data were analysed by two-way ANOVA and comparisons were made to wild type with Dunnett's multiple comparisons test; \*\*\*\* refers to  $p \leq 0.0001$  and \* to  $p \leq 0.05$ . This is experiment 2 of 2, the data points are representative of intra-experiment replicates and the bars show the mean.**

## 3.5. Discussion

### 3.5.1. Overview

This project started with a selection of gene hits identified from two previous HEK-293 GT library infection screens that aimed to identify host resistance genes and pathways. The intention was to identify key host biological processes critical for infection that might be targets for drugs which would limit pathogen invasion or replication. In this chapter we used cell lines mutated by CRISPR to test the validity of these genes in an independent fashion.

### 3.5.2. Summary of the genetic characterisation of CRISPR KO cell lines

The CRISPR/Cas9 system was used to target 5 of the genes (*SENP5*, *PRKG1*, *PCDH7*, *ARID4B*, *PSMA1*) hypothesised to comprise a host infection response network to generate a pool of mutated cells for each respective gene. Multiple lines were isolated and for each of which a subset was tested for CRISPR generated mutations through a variety of methods. At least one cell line harbouring CRISPR generated mutations predicted to produce no or non-functional protein were found for 2 of the genes (*PRKG1*, *PCDH7*) and 2 cell lines with altered protein localisation were identified for a third (*ARID4B*). Genetic characterisation of CRISPR KO *SENP5* Line 3 indicated that the protein was likely to be absent or non-functional, but a Western blot demonstrated comparable protein expression to the CRISPR control KO *SEMA3A* pool.

CRISPR generated mutations (2 of 6 plasmids tested) were found in a pool of CRISPR *PSMA1* targeted cells, of which one was a nonsense mutation. *CFTR* CRISPR lines were also made but not characterised due to time constraints.

The mutations found in CRISPR KO *SENP5* Line 3 and *PCDH7* Line 2 were a combination of nonsense mutations (resulting from a premature stop codon) and moderately sized amino acid insertions/deletions. Nonsense mutations produce abnormally short transcripts, which can be identified as abnormal and degraded by the cell (in a process known as nonsense mediated decay). If not degraded, the transcript will produce a truncated protein, which will likely be non-functional. Small- to moderately- sized amino acid deletions or insertions, however, vary substantially in effect depending on the mutation location, effect on folding and effect of any charge changes on function.

CRISPR KO *PRGK1* Line 12 was genetically less characterised, as initial cloning isolated 2 primer dimers as well as copies of the genuine target region. Based on this limited evidence CRISPR KO *PRGK1* Line 12 could be as genetically corrupted as *PCDH7* Line 2, or the Cas9 enzyme could have been more efficient. Nevertheless, these lines were taken forward for phenotypic analysis of the mutated genes in infection.

### 3.5.3. Are CRISPR KO HEK-293 cell lines heterogeneous?

As somatic cells in *Homo sapiens* are diploid, one might initially expect up to two versions to present in a truly clonal CRISPR mutated isolate; however, HEK-293 (the cell line

background) is aneuploid (Lin et al., 2014; Stepanenko and Dmitrenko, 2015). Thus, one possibility is simply that there are several copies of these genes per cell. Indeed, interrogation of the 293 Variant Viewer (<http://www.hek293genome.org/index.php>) (Lin et al., 2014) revealed that *PRKG1* accession NM\_001098512, *SENP5* accession NM\_152699 and *PCDH7* accession NM\_002589 have copy numbers of 2.98, 3.16 and 1.85, respectively in the HEK-293 cell isolates investigated (Lin et al., 2014). The absence of integers means the copy numbers are averages, such that the cell population copy numbers varies from cell to cell. Five different copies of the CRISPR target region for *SENP5* and *PCDH7* were found in single isolated CRISPR treated lines; suggesting that either the variation in copy number was wider than expected, or the isolated lines were heterogeneous. A possible cause of mutant cell line heterogeneity may be that the CRISPR/Cas9 system takes longer to mutate the target sequences than the processes of antibiotic selection and cell division, leaving antibiotic resistant colonies of cells that are initially homogeneous, but become more heterogeneous as time passes due to continual CRISPR-mediated mutation.

Others have reported results of mutations generated with CRISPR/Cas9 that indicate the presence of more than two versions of a given target locus in ostensibly clonal cell lines (Koch-Edelmann et al., 2017; Napier et al., 2016; B. Pickard 2018, personal communication, 27 July). Koch-Edelmann et al., 2017 described a CRISPR treated and clonally isolated cell line which had (at least) four different copies of the targeted locus. They used the HeLa cell line, which is also aneuploid and is described as genetically stable (Adey et al., 2013). Adey et al., 2013 have sequenced the HeLa genome and identified the copy number for different genomic regions. The copy number reported for chromosome 5 region 70,612,810-83,952,451 (where the gene mutated by Koch-Edelmann et al., 2017 (*CERT*) is located) is three (Adey et al., 2013). Analysis of the CRISPR induced mutations described in Napier et al., 2016 are limited, as they only published the electropherograms of the CRISPR target locus using the wild type electropherogram as a template. The noise in these sequences was ascribed to the presence of biallelic mutations and this is believable for two of the three targeted genes for which electropherograms are presented. In contrast, the third is much noisier, suggesting either contamination or a similar phenomenon to that encountered here. Some authors rely on an indirect technique known as the T7 endonuclease assay to confirm CRISPR-induced mutations (Grobarczyk et al., 2015; Zhou et al., 2014) and note that while mosaic colonies are possible, this can be remedied through later re-isolation.

#### 3.5.4. Were the 'mutants' functionally mutated?

Similar levels of *SENP5* protein were found in CRISPR KO *SEMA3A* and CRISPR KO *SENP5* cell line samples. This was disappointing because the genetic characterisation had indicated the presence of harmful (nonsense) mutations. This issue might be explained by a combination of factors. The *SENP5* alleles presented as S1-3 and S1-5 (Section 3.1.1) may have produced protein that could fold sufficiently well such that ubiquitination and proteasomal degradation were avoided, and the epitope recognised by the antibody was still present. Alternatively, wild type alleles may have been present that did not show up in the PCR sampling process. Either of these possibilities could explain the discrepancy between the genetic characterisation and the intensity of protein expression if combined with additional considerations. Firstly, wild type versions may have been expressed at a higher,

compensatory level in heterozygote cells. Secondly, the PCR sampling process used for mutation characterisation did not fully represent the ratio of nonsense mutations to mild mutations or perhaps the presence of many wild type alleles. Thirdly, severely mutated cells in the CRISPR KO pool were compromised and outcompeted through greater proliferation by milder mutants between the genetic characterisation and measurement of protein expression. Given the resources available, there was a limited ability to analyse the infection phenotypes generated by KO of *SENP5* or *ARID4B*.

The presence of *SENP5* or *ARID4B* protein (whether mutated or not) in the respective CRISPR KO cell lines suggest that the mutations are not sufficiently penetrant, or at a high enough percentage, to effectively interrogate the respective KO infection phenotypes.

### 3.5.5. Summary of the phenotypic characterisation of CRISPR KO cell lines

The LDH activity and bacterial burden results from this chapter are summarised in **Table 3.2**. One CRISPR KO line (*PCDH7* Line 3) has consistently (across 2 replicate infection experiments) reduced STM and *S. sonnei* burdens at 24 hours post gentamicin maintenance addition. Another CRISPR KO line (*PRKG1* Line 12) had reduced STM and *S. sonnei* burdens at this time point, though only one experiment for each pathogen was performed.

**Table 3.2 Summary of infection results using isolated and genetically characterised CRISPR KO cell lines.** LDH activity or bacterial burden increased ( $\uparrow$ ), decreased ( $\downarrow$ ) or stayed the same (=); where not statistically significant but suggestive, the indicated direction has a \*. Time points for which a comparison is not applicable are listed as NA. Experiments were performed using CRISPR KO PRKG1 Line 12 once only.

Pathogen	Assay	Time point (hours)	Gene of interest	
			PCDH7	PRKG1
STM	LDH activity	24	=	$\uparrow^*$
		48	$\uparrow$	$\uparrow^*$
		72	=	=
	Bacterial burden	0	=	$\downarrow^*$
		2	=	$\downarrow^*$
		24	$\downarrow$	$\downarrow$
	Normalised bacterial burden	0	NA	NA
		2	=	=
		24	$\downarrow$	=
<i>S. sonnei</i>	Bacterial burden	0	=	$\downarrow$
		2	=	$\downarrow$
		24	$\downarrow$	$\downarrow$
	Normalised bacterial burden	0	NA	NA
		2	=	=
		24	$\downarrow$	$\uparrow$

The reduced bacterial burden observed in CRISPR KO line *PRKG1* Line 12 was unexpected as *PRKG1* was identified on the basis of high bacterial burden/GFP expression in the FACS analysis of gene trapped cells (Chapter 1, J Yu 2016 personal communication, 3 October). One possible theory to explain this apparently reversed effect is the presence of large numbers of internal, but non-culturable (dormant) or compromised, GFP-expressing STM bacteria within GT mutant *PRKG1* cells that drove selection under the original FACS screen conditions. Under the conditions of host cell lysis assay, these bacteria would fail to grow, therefore not contributing to countable CFU and generating the appearance of infection resistant host cells. Similar dormant bacteria are known as ‘persisters’ and have been observed in BMDM and fibroblast infection (Helaine et al., 2014, 2010; López-Montero et al., 2016). In contrast, ‘compromised’ propidium iodide positive *E. coli* populations (nevertheless expressing GFP) may be generated by treatment with moderate concentrations of ethanol (Lehtinen et al., 2004). It may be that in CRISPR KO *PRKG1* cells some mechanism is contributing to the transformation of growing bacteria into similar compromised or persister populations.

An alternative theory is that clustered *PRKG1* GT mutants present in the FACS screen appeared to have an artificially high GFP signal due to signals from multiple STM-GFP infected cells in the same droplet. Before sorting, cell populations are strained as standard practise to remove as many clusters as possible. Additionally, small doublet populations can normally be gated out and thus would not be sorted; however, host cell size can depend on

the number of intracellular bacteria (Westermann and Vogel, 2018), making doublet gating potentially impossible. Furthermore, if samples are particularly clumpy, distinguishing doublets from single cells can be difficult. During generation and phenotyping of CRISPR KO *PRKG1* cells, it was found that these were particularly prone to clustering or clumping; therefore, there is a strong possibility that this was the cause for the high GFP signal in the initial screen.

### 3.5.6. Validated gene *PCDH7*

*PCDH7* is a protocadherin cell-cell adhesion molecule. It is a membrane protein localised to the plasma membrane and cell junctions. Both protocadherins and cadherins are linked to the infectivity of pathogenic organisms and viruses in the literature (Devaux et al., 2019; Franklin-Murray et al., 2020; Jangra et al., 2018; Kochi et al., 2019). This is due to their action either as sites of entry for pathogens, or as components of tight junctions that maintain barriers (gut, blood-brain). *PCDH7* membrane localisation means that it is unlikely to directly influence the intracellular bacterial burden at later time points. However, CRISPR KO *PCDH7* Line 2 had lower bacterial burden at 24-hours post gentamicin maintenance addition (Figure 3.11). To explain this later effect, indirect mechanisms such as participation in signalling pathways may influence burden. *PCDH7* is not a well characterised member of the protocadherins but other cadherin superfamily members have well-documented roles in signalling (e.g. Cadherin-1 in the *wnt* pathway) see van Roy and Berx (2008) for review. Interestingly, the identification of a second cadherin gene is described in Chapter 5.

### 3.5.7. Initial invasion protection may affect burden later

The 0-hour time point described here is approximately 2-hours post inoculation (Figure 3.11) and 1.5-3 hours post inoculation is frequently used to compare bacterial invasion (Edwards and Massey, 2011; Mahmoud et al., 2016a, 2016b; J. Wu et al., 2014). Indeed, it has been noted that the amount of intracellular STM replication is trivial in the first 6 hours after uptake/ invasion in BMDMs (Helaine et al., 2010). This is mirrored in the almost imperceptible changes observed in bacterial burden between 0- and 2- hour time points. The reduction in *S. sonnei* burden in CRISPR KO *PRKG1* Line 12 compared to the wild type burden at 0-hours post gentamicin maintenance most likely indicates reduced invasion. That this reduced burden is maintained at 2- and 24-hours suggests that the *S. sonnei* in CRISPR KO *PRKG1* Line 12 failed to catch up to the wild type burden. As both *PRKG1* isoforms are intracellular proteins, this would require interaction or modification of another protein. *PRKG1* is a serine/threonine protein kinase, so lack of phosphorylation (due to non-functional *PRKG1*) on a protein involved in phagocytosis, or bacterial invasion may generate such an effect. This analysis fits better with the theory about artificially high GFP signals from *PRKG1* mutant doublets (described in Section 3.5.1) than the theory about persister- or compromised- GFP<sup>+</sup> intracellular bacterial populations.

### 3.5.8. Cross-species similarities and differences

The consistent patterns seen in CRISPR KO *PRKG1* Line 12 and *PCDH7* Line 2 between STM and *S. sonnei* infections suggest host processes or essential features for pathogen survival common to both bacteria. *Salmonella* and *Shigella* share many molecular features of

infection, such as the activation of host cell Ras-related C3 botulinum toxin substrate 1 (RAC1), resulting in lamellipodia/membrane ruffling, thus, promoting bacterial entry (Bulgin et al., 2010; Criss et al., 2001; Friebel et al., 2001; Handa et al., 2007; Ohya et al., 2005; Patel and Galán, 2006; Weigele et al., 2017). The conservation of effects across species is an encouraging sign for the future generation of a common therapeutic drug.

### 3.5.9. Critique of CFU to OD

The primary aim of the CFU to OD<sub>600</sub> correlation generated here was to know the viable bacterial concentration at an approximate OD<sub>600</sub> of 0.6 ( $\pm 0.5$ ). This information is required to quickly calculate the volume of bacterial culture required to produce a desired MOI. The CFU to OD<sub>600</sub> correction presented here was determined from dilutions of culture at a starting OD<sub>600</sub> of 0.9. The correlation between OD<sub>600</sub> and CFU should be approximately linear during the log phase of growth in the absence of antibiotics if the OD measurements are taken within the linear range of the spectrophotometer. Ideally, the bacterial culture would have been sampled at a starting OD<sub>600</sub> of 0.6, as bacteria (including STM) are known to change shape during the transition from log to stationary phase with a corresponding change in the OD<sub>600</sub>-bacterial concentration relationship (Hawkins et al., 2019; Stevenson et al., 2016). Stevenson et al., (2016) state that the change in OD<sub>600</sub>-concentration gradient accelerates as the diameter of the organism nears the wavelength of light used. Stevenson et al., (2016) found that the change in gradient for an *E. coli* culture transitioning between mid-log phase and early stationary phase was quite small. STM rods during stationary phase are 3.5-5 times longer than the wavelength (600 nm) used to measure the OD (González-Pérez et al., 2019) and the reduction in rod length between phases is reported to be ~20% of the total length (Hawkins et al., 2019). An equation and the associated CFU to OD<sub>600</sub> values generated from STM stationary culture in another laboratory is provided (Table 3.3) for comparison (González-Pérez et al., 2019). The differences in path length of the spectrophotometer used will change the OD<sub>600</sub>-concentration as well as the difference in growth phase, nevertheless, both equations generate CFU of the same order of magnitude at the same OD values. Together, these factors indicate that a change in the OD<sub>600</sub> - concentration relationship due to changes in shape between an OD<sub>600</sub> of 0.6 and 0.9 is small enough that the equation developed here (Figure 3.13) will suffice.

**Table 3.3 Side by side comparison of predicted STM numbers from an initial OD reading is largely the same for my equation and that produced by Gonzalez et al. Comparison of STM CFU generated from an equation developed by (González-Pérez et al., 2019) ( $Y = (6.895 \times 10^8) X - 0.319 \times 10^6$ ) and an equivalent equation developed here (presented in Figure 2.8 and Figure 3.13) across a range of OD<sub>600</sub> values.**

OD <sub>600</sub>	González-Pérez et al. (2019)	My STM equation
	CFU /mL	CFU /mL
0.3	2.07E+08	2.07E+08
0.4	2.75E+08	3.16E+08
0.5	3.44E+08	4.26E+08
0.6	4.13E+08	5.36E+08
0.7	4.82E+08	6.46E+08
0.8	5.51E+08	7.56E+08
0.9	6.20E+08	8.65E+08
1	6.89E+08	9.75E+08

### 3.5.10. Critique of the LDH assay

The original method of selection for 4 of the 5 candidate genes was based on infection survival; therefore, the increase in LDH, as a proxy for increased cytotoxicity, with STM infection observed in CRISPR KO *SENP5* Line 3 and *PCDH7* Line 2 was unexpected. One explanation may be that LDH activity is only a proxy for certain types of cell death that include membrane permeability, as the LDH needs to move into the media to generate a change in the concentration of coloured formazan product. Plasma membrane permeabilisation, and thus LDH diffusion, occur during necroptosis and pyroptosis, but not apoptosis. Pyroptosis appears to be the dominant form of *Salmonella* generated epithelial cell death *in vivo* (Knodler et al., 2010; Rauch et al., 2017), but uninfected epithelial cells within the infected host demonstrate characteristics of apoptosis (Knodler et al., 2010). Furthermore, low levels of apoptosis were found in colon cross-sections from attenuated STM-infected mice; whereas mice infected with wild type STM (with increased replicative capacity) had increased colonic epithelial extrusion and inflammation (Hefele et al., 2018).

Thus, some protective effects against host cell death (e.g. apoptosis) would be 'invisible' to the assay. An alternative explanation for conflicting results could be that an increase in STM death, rather than host death, is contributing to the assay readout: a STM protein (L-lactate dehydrogenase) is predicted to have LDH activity based on homology (identified on Uniprot <https://www.uniprot.org/>) (McClelland et al., 2001; The Uniprot Consortium, 2019). Countering this, several authors have used LDH activity as a proxy to measure host cell death during STM infection (Shuai-Cheng et al., 2016; Volf et al., 2010; L. Wang et al., 2019; Wu et al., 2018). Considering the alternatives available (see Section 3.4.2), this option was the most appropriate; given further time, more method optimisation could have been performed as well as further investigation of the phenotype.

### 3.5.11. Critique of the bacterial burden assay

The final protocol (see Chapter 2) was arrived at through collaboration with colleagues at DSTL, Porton Down and the need to adjust for the aberrant growth of some mutant cell lines (e.g. CRISPR KO *PRKG1* Lines). A baseline time point (0-hours post gentamicin maintenance addition) was chosen for the purposes of normalisation, though in reality, this is approximately 2-hours post inoculation (Figure 3.11).

The gentamicin concentration was reduced to a maintenance concentration after 1 hour to reduce the chances of gentamicin exposure to intracellular bacteria. Additionally, host cells were only washed once immediately prior to lysis to avoid washing off loosely adhered cells. Both of these factors may result in greater numbers of extracellular bacteria (due to less thorough washing and antibacterial activity) compared to some literature protocols (Knodler et al., 2010; Mahmoud et al., 2016b; J. Wu et al., 2014). The use of these tools to promote a suitable assay for assessment of differences in intracellular bacterial growth is a balancing act as disadvantages are likely for each extreme. The caveats described for the infection protocol used here are more likely to obscure invasion resistant host phenotypes due to a background level of (extracellular) bacterial growth. Additionally, the CFU readout is averaged across all cells per well, which would obscure potential variation in phenotype inherent in likely heterogeneous CRISPR KO lines. Despite the reduced sensitivity to detect

invasion resistant phenotypes, a potential invasion resistant host cell line has been identified (CRISPR KO *PRKG1* Line 12), indicating that the real effect on burden may be greater.

### 3.5.12. Problems with the CRISPR sham (*SEMA3A* KO) control

The CRISPR KO *SEMA3A* pool produced lower levels of bacterial burden than HEK-293 wild type cells in many STM infections and all *S. sonnei* infections performed. Slightly reduced differences in the phenotype under study between CRISPR controls and CRISPR KOs compared to wild type cells may be expected; but the consistent extremely low bacterial burden seen in *S. sonnei* infection is unusual for a control. The gene *SEMA3A* was chosen because of its known 'irrelevant' neuronal function. However, after subsequent literature research, it was found that wild type *SEMA3A* promotes apoptosis in multiple cell types (Birger et al., 2018; B. Huang et al., 2018; Rienks et al., 2017; Tian et al., 2018; Wehner et al., 2016) whereas its deletion causes cell death in others (Birger et al., 2018). *SEMA3A* was not expected to be expressed in HEK-293 cell lines, but low levels of the transcript have been reported in HEK-293 (<https://www.proteinatlas.org/ENSG00000075213-SEMA3A>, available from v20.0.proteinatlas.org)(Thul et al., 2017). Perhaps the deletion of *SEMA3A* makes HEK-293 cells more susceptible to cell death. If greater numbers of dead host cells were lost during media changes or washes, this could result in a concomitant greater loss of bacteria, artificially reducing the CFU count. Some low-level monolayer loss was observed in some wells prior to host cell lysis, but it seems unlikely that this is the cause for the difference in *S. sonnei* bacterial burden.

Perhaps *SEMA3A* is serendipitously involved in the promotion of cellular *S. sonnei* infection. Interestingly, further literature research unearthed a receptor for the agonist ligand *sema3a* in mice, *plexin-A4*, that is involved in bacteria-induced cytokine responses through *Rac1* activation (Wen et al., 2010). *RAC1*, a small GTPase, is activated by *Salmonella* and *Shigella* effector proteins (e.g. Guanine nucleotide exchange factor *SopE* (*SopE*) in *Salmonella* infection and *IpgB1* in *Shigella* infection) which in turn causes lamellipodia/membrane ruffling, thus, promoting bacterial entry (Bulgin et al., 2010; Criss et al., 2001; Friebel et al., 2001; Ohya et al., 2005; Patel and Galán, 2006; Rudolph et al., 1999; Weigele et al., 2017). Thus, perhaps the lack of *SEMA3A* protein in the HEK-293 CRISPR control reduces the activation of *RAC1*, resulting in reduced membrane ruffling and pathogen entry, which would explain the reduced CFU.

This method of control was chosen as all aspects of the intervention are controlled for – ectopic Cas9 and antibiotic resistance gene expression, as well as Cas9 targeting activity. Another potential cause of the unusual CRISPR KO sham control responses is activation of the p53 pathway in response to Cas9 overexpression or sgRNA directed activity, a connection that has recently been explored by multiple groups (Enache et al., 2020; Haapaniemi et al., 2018; Ihry et al., 2018; Schirotti et al., 2019). Indeed, *Shigella flexneri* instigates p53 degradation, an activity essential for maintenance of the intracellular epithelial niche through the prevention of host cell death (Bergounioux et al., 2012). In contrast, STM infection activates p53 (Wu et al., 2010), which promotes STM proliferation *in vitro* (T. Huang et al., 2018) and increases host susceptibility to the infection *in vivo* (Khan et al., 2012). The difference in pathogen responses to p53 suggest that p53 activation is

either not a major factor, or simply one of multiple factors, affecting the phenotype of the CRISPR KO 'sham' control. Another CRISPR control instead lacking a functional sgRNA responded comparably to wild type HEK-293 cells in STM and *Shigella* infections (data not shown) – indeed, Enache et al., (2020) notes that undirected Cas9 nuclease activity may induce less p53 activation. Unfortunately, this CRISPR control demonstrated other persistent undesirable behaviours.

### 3.5.13. Future work

#### 3.5.13.1. Creation of a better CRISPR control

Infection experiments involving *F. tularensis* Schu S4 (a Biosafety level 3 pathogen) were planned by collaborators at DSTL, Porton Down with promising CRISPR KO cell lines (such as *PRKG1* Line 12 and *PCDH7* Line 2); but the CRISPR sham control *SEMA3A* KO performed poorly during infection with STM and *S. sonnei*. A new alternative (and effective) CRISPR control population will be required before high cost experiments involving Biosafety level 3 pathogens can be performed. A potential solution could be to create new CRISPR KO lines as well as a control through stable sgRNA expression but transient Cas9 expression (or even electroporation of the Cas9 ribonuclease) – e.g. Liang et al., (2015) and Hultquist et al., (2016).

#### 3.5.13.2. Clarification and further empirical analysis of infection phenotypes

The genetic complexity of the CRISPR generated 'KO' lines may have made analysis of the effect of KO mutation on the infection phenotype difficult but combining these results with the following suggested approaches might clarify the interpretation of the mutant phenotypes.

More clonal CRISPR mutant populations could potentially be re-isolated from the CRISPR KO *PCDH7* and *PRKG1* lines described here to heighten functional changes. A variety of isolation strategies are commonly used, some of which are purported to avoid contamination of isolates with other lines (such as limiting dilution, or use of clonal rings, discs or low-melting agarose). Limiting dilution has previously been reported to cause changes in gene expression (Casden and Behar, 2019) which could affect phenotyping assays. Casden and Behar (2019) suggest co-plating highly diluted CRISPR treated cells containing an antibiotic selection marker with wild type cells lacking the selection marker to help CRISPR mutants of interest grow. This method requires a selection marker to which recipient cells are highly susceptible. Unfortunately, a small number of HEK-293 wild type cells used as selection controls were able to outlast G418 antibiotic selection at an optimised concentration.

Analysis of host cell death by LDH activity alone has produced interesting, though inconclusive results; therefore, additional methods of interrogating the host cell death phenotype of the CRISPR KO lines during infection is suggested. A cell death marker could be used to assess CRISPR KO cell lines infected with GFP expressing STM at different time points and visualised by microscopy. The number of dead cells as a proportion of total cells (both an average number pre-infection and the number of cells at the time point) could be enumerated and the bystander vs infected population death could be inferred. A less labour-intensive option is also possible: HEK-293 wild type cells (internal control) could be

co-plated with CRISPR KO cell lines before infecting with STM, before staining with anti-Cas9 antibody (which would stain CRISPR KO cells) to visually identify a difference in population survival at multiple time points. A more quantitative (though also more expensive) method would be to infect HEK-293 CRISPR KO cell lines as well as appropriate controls with GFP expressing STM before measuring apoptosis (annexin-V) and membrane-permeabilising forms of cell death on a flow cytometer at different time points, as a means to dissect out the possible forms of death.

It has previously been noted that the CRISPR KO *PRKG1* cell lines had a slower proliferation rate and the cells are prone to clustering. As the output of the bacterial burden assay relies on an averaged bacterial burden per well, assumptions must be made about the number of host cells present. Substantial effort has been made to control for this – including the use of confluent monolayers during the assay; nevertheless an alternative method of counting intracellular bacteria could provide additional evidence to support the identified phenotype.

The genetic validation of these gene hits have been performed with a single sgRNA sequence per gene; but the transfection of further sgRNAs to target different sites within the same genes into HEK-293 cells would greatly enhance the evidence for gene-phenotype relationships. To understand the applicability of these KO(s) in other (more infection relevant) cell types, KO(s) should be generated in well characterised epithelial lines derived from the small intestine or colon for use in infection studies. To understand the connection between mutation and phenotype the underlying molecular mechanisms must be characterised. To determine if the infection phenotype has relevance at a system or organismal level, *in vivo* characterisation of the KO(s) in infection must be performed.

#### 3.5.13.3. Generation of new CRISPR KO *SENP5* and *ARID4B* cell lines for validation of infection phenotypes

Transfection of HEK-293 cells with plasmids encoding Cas9 and sgRNAs targeting *SENP5* or *ARID4B* generated genetic mutations in the respective genes but failed to generate complete KO cell lines. Creation of more complete KO cell lines should be performed in order to validate the infection phenotype identified in the HEK-293 GT infection survival screen described in Chapter 1. Creation of these KO cell lines should incorporate multiple sgRNAs targeting different exons per gene (as described above) to enhance the effect of the mutations.

#### 3.5.14. Conclusion and rationale for work described in the next chapter

The aims of the current chapter were to generate independent KO mutants for genes (*SENP5*, *PCDH7*, *PSMA1*, *ARID4B* and *PRKG1*) identified from the HEK-293 screens, as well as for a further gene (*CFTR*) identified through bioinformatics analysis; and use those to validate the infection phenotype. In Chapter 1, a host infection response network composed of the above gene products was proposed based on the screen results and loose connections from the literature. A further hypothesis was that infection phenotypes observed in the HEK-293 GT survival and FACS screens could be independently corroborated using CRISPR KOs of the identified genes in the same genetic background.

Despite the complexity of the analysis, the data presented here indicates that mutating *PCDH7* promotes HEK-293 cell resistance to infection by both STM and *S. sonnei*.

As a plasma membrane protein, PCDH7 might be expected to exert an effect on bacterial adhesion or entry; indeed, other cadherin superfamily proteins function in this manner (Everman et al., 2019; Kochi et al., 2019; Li et al., 2016; Pizarro-Cerdá and Cossart, 2006; Sousa et al., 2007; Watters and Palmenberg, 2018). However, the bacterial burden data presented here indicate that the PCDH7 KO may be most important at time points later than 2 hours post gentamicin maintenance addition, thus perhaps it is involved in a signalling pathway that has a downstream effect. In contrast, CRISPR KO *PRKG1* lines appear to proliferate at a slower rate and Line 12 showed a bacterial burden phenotype counter to that expected based on the initial screen. The difference in burden of culturable bacterial cells found in Line 12 compared to the wild type host is sufficiently great so as to compensate for reduced host proliferation. The contrast between the original GFP-screen and CFU count-CRISPR phenotype may also be explained; either that GFP<sup>+</sup> non-culturable bacterial populations made up the difference, or that the prone-to-cluster nature of the *PRKG1* KO mutant resulted in their appearance in the high GFP gate. The combination of these technical and biological complications, together with the lack of biological replicates, means that further evidence is required for confirmation of infection resistance through *PRKG1* mechanisms.

Due to technical and time constraints, suitable CRISPR KO cell lines for *ARID4B*, *SENP5*, *PSMA1* and *CFTR* were not identified, so the respective infection phenotypes indicated in the HEK-293 screens still await independent empirical investigation. Amplification of the loci around the respective CRISPR target sites for *ARID4B* and *CFTR* failed to produce genuine amplification products (data not shown). The cloning and sequencing for the CRISPR targeted loci of *SENP5* and *PSMA1* indicated that the cell isolates chosen did not have sufficient mutation penetrance (Section 3.3, data not shown). Work to investigate further cell isolates for mutations was de-prioritised due to time constraints on the project.

In both of the initial screens (high GFP/STM-containing host cells selected by FACS, and infection survival), the identification of novel genes of interest relied on successive rounds to remove background effects. Without an uninfected control population taken through the same processes, it is impossible to know if selection pressures (such as proliferation rate) in addition to the infection were responsible for some of the identified genes. These screens could have been more physiologically relevant if primary cells (or a cell line) of the same type and from the same tissue or organ as those normally infected *in vivo* were used. For example, HEK-293 cells do not express many of the PRRs found in colonic epithelial cells that could recognise *Salmonella*, therefore the signalling pathways initiated in HEK-293 cells upon infection would likely be quite different. Additionally, HEK-293 cells were not characterised prior to immortalisation, making cell type specific observations impossible. Indeed, evidence supporting renal progenitor (Ashokkumar et al., 2006; Cusick et al., 2010), neuronal (Shaw et al., 2002) and adrenal gland (Lin et al., 2014) origins have been found (Stepanenko and Dmitrenko, 2015). Furthermore, few studies of *Salmonella* infection have used HEK-293. These factors make troubleshooting difficult.

With these confounding issues in mind, a cell line more relevant to intracellular bacterial infection was adopted for use in two future infection screens described in Chapter 5. These screens also allowed a more objective analysis due to the parallel use of an uninfected control population, as well as the quantitative assessment of the full set of library gene mutations within each population.

In conclusion, the hypothesised host infection response network described in Chapter 1 remains to be fully tested, but the results described in this chapter are promising.

## Chapter 4

The development of a macrophage-like gene trap library and its screening with two species of bacteria

## 4. CONSTRUCTION, DIFFERENTIATION, AND SCREENING OF A GENE TRAP LIBRARY IN U937 MACROPHAGE-LIKE CELLS

The previous chapters identified and validated a number of genes that affect the growth of intracellular bacteria in HEK-293 cells. The screening method employed iterative rounds of infection and recovery (using penicillin/streptomycin antibiotic treatment) of those surviving cells containing gene traps (GTs) which conferred resistance to bacterial infection. This approach had strengths, such as the power to substantially purify genuine resistance mutants, but it also had limitations. The primary limitation was the use of a manual RACE-PCR/gene cloning protocol that could only isolate and identify a few mutated genes in total – and it was not possible to estimate the how strong the survival advantage conferred by each mutation was (in comparison to other mutations). A second limitation was the choice of a human embryonic kidney cell line, HEK-293, as a model for infection. While technically amenable to genetic manipulation, this line may not represent the most immunologically relevant infection model. While intracellular bacteria are often capable of infecting a multitude of host cells, there may be greater, and more relevant, biological information to be gained from the screening of professional immune cells such as macrophages. STM is known to infect epithelial cells (Rauch et al., 2017), fibroblasts (López-Montero et al., 2016), macrophages (using them as a replicative pathogenic niche) and dendritic cells (Kiama et al., 2006) - a range of cellular targets that has been recently reviewed by Hume et al., (2017). STM is a widespread human-infective pathogen and has potential as a bioterror agent. *F. tularensis* Schu S4, another potential and highly pathogenic bioterror threat which can cause a lethal pulmonary infection (Adalja et al., 2015), is capable of infecting a range of non-phagocytic (e.g. epithelial) cells and phagocytic cells including dendritic cells and macrophages (macrophages are also the preferred replicative niche) (Celli and Zahrt, 2013; Hall et al., 2008, 2007). *F. tularensis* LVS (a biosafety level (BSL) 2 pathogen) is commonly used as a model for *F. tularensis* Schu S4 (BSL 3).

The studies described in this chapter aim to identify genes that affect the growth and, or, survival of these two species of bacteria in macrophages. To facilitate this, a neoplastic, histiocytic cell line, U937, was chosen as a cellular model. Although these cells are cultured and archived as proliferating monocytes, previous studies have found that they can be effectively differentiated into ‘macrophage-like’ cells *in vitro* (Chanput et al., 2015). However, the exact protocol for differentiation varies from publication to publication, as does the extent and phenotype of the cells produced. Gene trapping of a cell line rather than primary immune cells, such as human blood monocyte-derived macrophages, was pursued for two reasons. Firstly, non-proliferating human monocytes or monocyte-derived macrophages (MDMs) would not be amenable to making a multi-use GT library. This is an important consideration because multiple screens with 2 different pathogens taking place over 2 geographical sites were planned – a single library would be ideal to provide consistency and comparability across the screens. Secondly, the effective use of transient RNAi knock-down screens as an alternative is extremely labour- and time-intensive, as well

as requiring the use of advanced computational tools to remove off-target, sequence-specific effects from the analysis (Québatte and Dehio, 2017). The use of U937 cells solves these problems – our strategy was to transfect the monocyte cell line in suspension culture with the GT construct, select and amplify that library, and then later differentiate identical aliquots into ‘macrophage-like’ cells for multiple rounds of analysis. It was also originally envisaged that we could ‘retrodifferentiate’ U937 macrophage-like cells in order to multiply surviving cells after an infection screen – this would permit iterative rounds of selection. This final aim proved not to be possible, however.

The specific practical aims of the work described in this chapter are to: (i) refine a method for differentiating U937 cells into macrophages, (ii) construct a GT library in U937 cells, (iii) differentiate the library into macrophages and infect these, separately, with 2 bacterial pathogens to select resistant cells and (iv) determine which GT mutations are over- or underrepresented in the population of cells post-infection using next generation sequencing.

## 4.1. Creation of a differentiation protocol for the U937 cell line to generate macrophage-like adherent cells

U937 cells are a human monocyte derived-cell line that can be differentiated into cells with macrophage-like characteristics with appropriate drug treatments. A variety of differentiation drugs, drug concentrations and timelines have been documented in the literature, but the characterisation of the outcomes are varied and difficult to compare. Additionally, a large proportion of the literature focuses on THP-1 cells (another ‘monocyte-like’ cell line derived from monocytic leukaemia) (Chanput et al., 2014); so it is not clear how well these protocols would translate to U937 cells. To create the most physiologically relevant cellular model for screening purposes, it was necessary to perfect the treatment process such that large numbers of macrophage-like cells could be generated from the library prior to screening.

In pursuit of a suitable differentiated protocol, three different strategies were assessed (see Appendix A for Strategies #1 and #2, Strategy #3 is described in Section 4.1.1). From a technical standpoint, it was decided that Strategies #1 and #2 were unsuitable for use in infection discovery screens, due to the low number of cells with morphologies of interest.

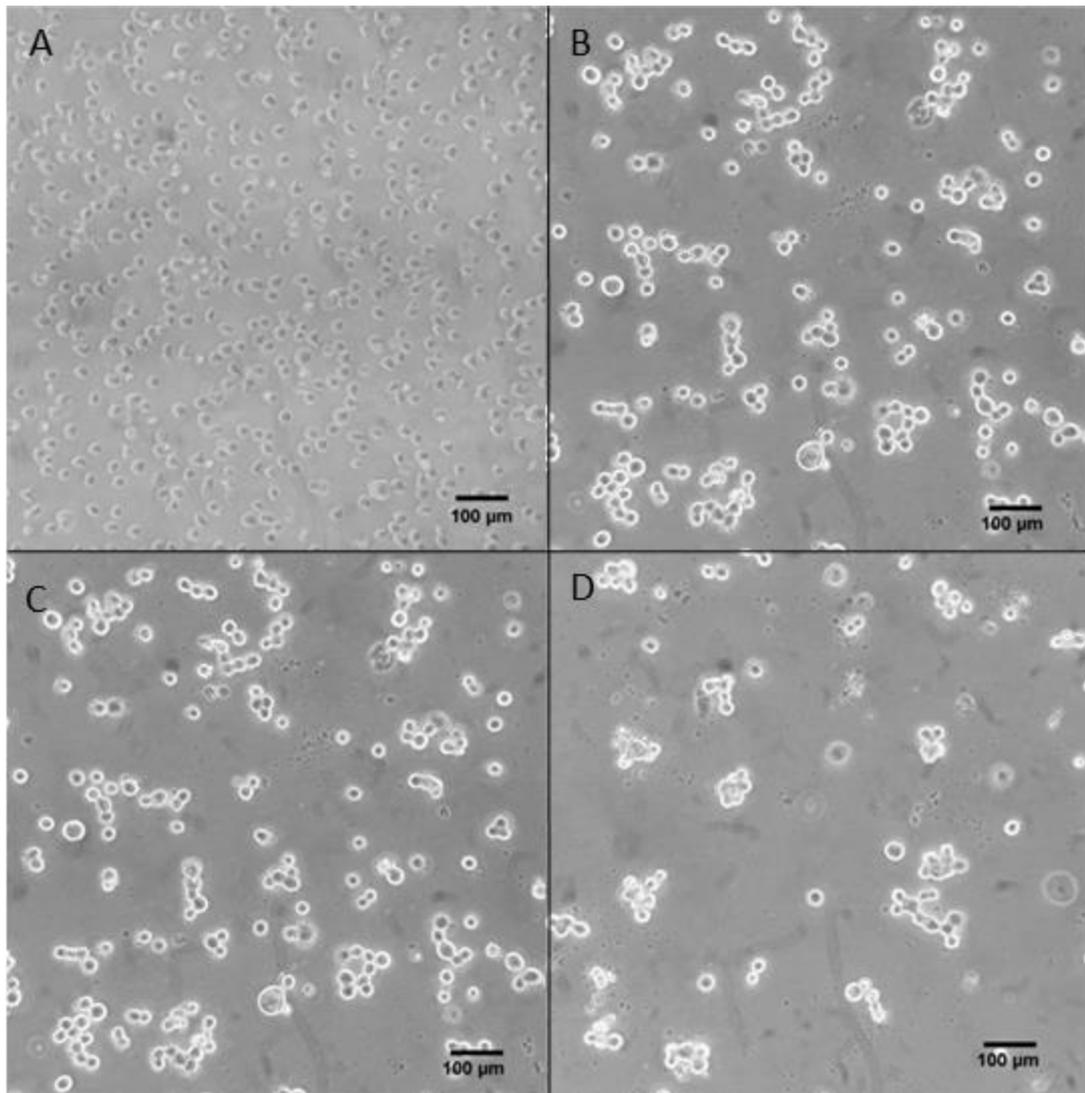
### 4.1.1. Differentiation Strategy #3

Differentiation combinations of three drugs (PMA, VD3, All-trans-retinoic acid (ATRA)) have been previously investigated and applied by others e.g. (Daigneault et al., 2010; Valdés López and Urcuqui-Inchima, 2018), so combinations of these were tested. As most of the changes in the single-drug experiment were observed with PMA treatment, further drug treatments focused on combinations involving PMA. In double-drug combination treatments, the highest concentrations of VD3 (0.8  $\mu$ M) and ATRA (1  $\mu$ M) from the previous experiment were each used in combination with PMA. Lower concentrations of all three drugs were combined as a drug cocktail (as well as the highest concentrations) to compare toxicity and desired effects on cells.

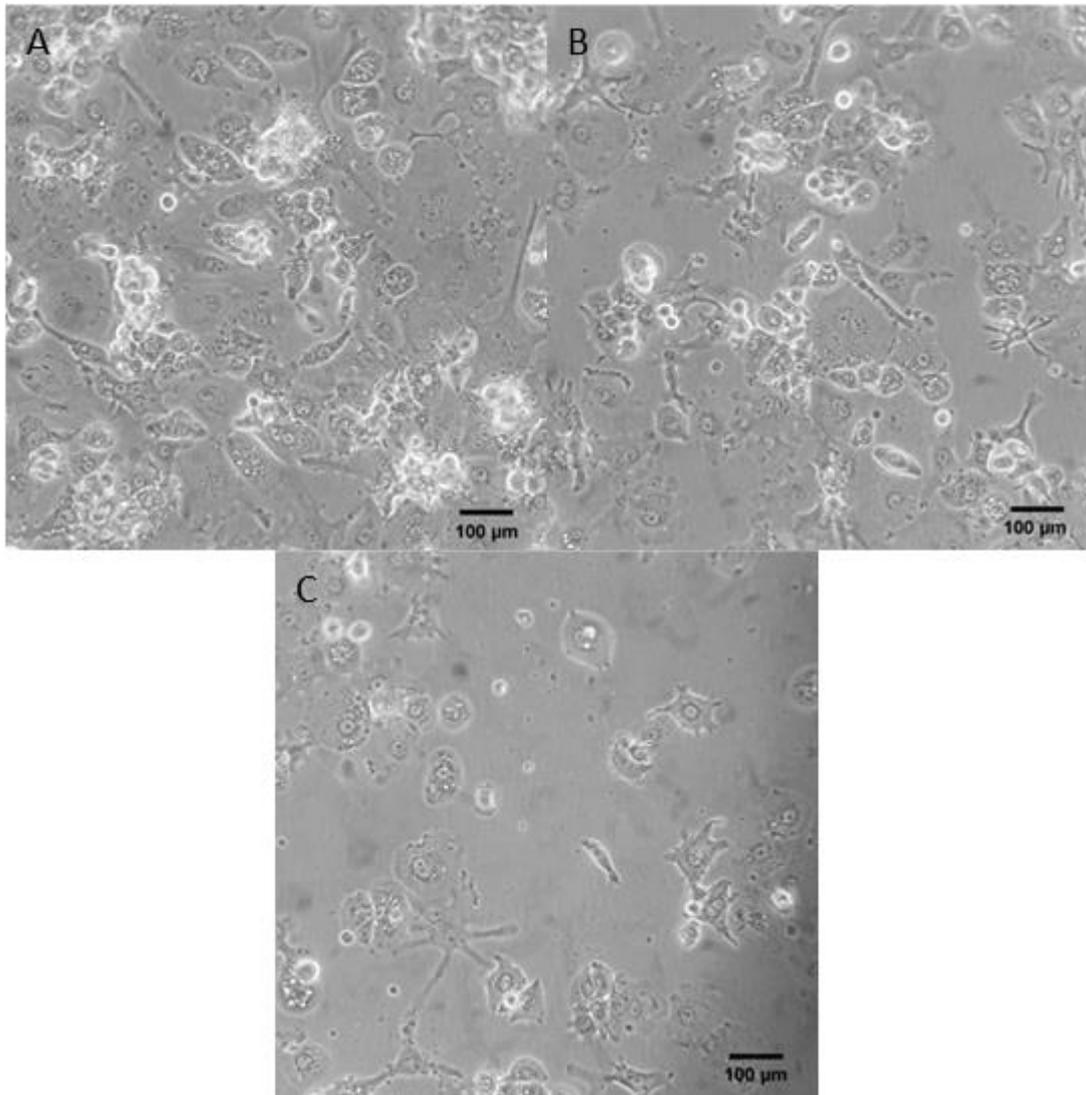
Cell population density was another factor that potentially hindered the acquisition of good micrographs during differentiation strategy #2 (

Figure A.4, Figure A.5). Monocyte seeding densities for macrophage differentiation vary substantially in the literature, but lower seeding densities are often used together with protocols designed to generate more homogeneous cell culture morphology with longer filopodia (Jin and Kruth, 2016) or bigger flattened 'pancake' shapes (Seo et al., 2015), perhaps due to the lack of contact inhibition. Sulahian et al., (2008) also note that higher cell densities reduce human monocyte derived macrophage (MDM) size. Therefore, in Strategy #3, cells were seeded at  $5 \times 10^4$ /well in 6 well plates to obtain both better micrographs and potentially larger cells. The cells were treated with the following drug combinations at both 3% and 10% FCS; 0.2  $\mu$ M PMA with 0.8  $\mu$ M VD3, 0.2  $\mu$ M PMA with 1  $\mu$ M ATRA, 0.1  $\mu$ M PMA with 0.4  $\mu$ M VD3 and 0.5  $\mu$ M ATRA and, finally, 0.2  $\mu$ M PMA with 0.8  $\mu$ M VD3 and 1  $\mu$ M ATRA for 3 days. By day 3, large flattened adherent cells were already present in 0.2  $\mu$ M PMA with 0.8  $\mu$ M VD3 treated wells (data not shown). Small round adherent cells were found in clusters in all wells of treatment combinations containing ATRA. Subsequently, the media was replaced with fresh media without drugs for the resting phase before imaging 2 days later (Figure 4.1 and Figure 4.2). All combinations containing PMA and ATRA produced a common cell phenotype; circular in shape, except for a tight frilly border that was highly light-scattering, frequent clustering, and adherence to the plastic. Upon adjusting the focal plane to focus on the cell edges, the highly light scattering border was found to result from membrane blebbing, characteristic of apoptosis - see D'Arcy (2019) for a review of cell death. Combined 0.2  $\mu$ M PMA and 0.8  $\mu$ M VD3 treatment at 10% FCS (Figure 4.2) produced morphologies (elongated or 'pancake' shaped cells, some with filopodia) similar to those produced with 0.2  $\mu$ M PMA plus resting period in previous differentiation strategies (Figure A.2 and Figure A.5), but cells had a larger main body, were flatter, and some had longer filopodia. Overall, these features appeared in much greater proportions than in previous differentiation strategies. A few highly light-scattering (likely apoptotic) cells like those produced by PMA and ATRA treatment were also found in PMA and VD3 treated wells. Differentiation treatments performed at 3% FCS (data not shown) produced very similar results to those at 10% FCS but resulted in fewer cells overall so, as before, further experiments were performed in 10% FCS.

The work described here allowed us to define the optimal differentiation conditions as a  $1.6 \times 10^4$  cells /cm<sup>2</sup> seeding density with 0.2  $\mu$ M PMA, 0.8  $\mu$ M VD3 treatment for 3 days (day 0-3) before resting in differentiation drug free media for a further 3 days (day 3-6) after which cells were ready for experimental use (timeline in Figure 2.5). This differentiation protocol (#3) is hereafter referred to as the 'differentiation drug-cocktail' protocol.



**Figure 4.1 Differentiation strategies including ATRA in combination with PMA with or without VD3 result in what appears to be apoptotic cell death.** *Differentiation strategy #3: U937 cells were seeded at  $5 \times 10^4$ /well in 6 well plates and treated with the following combinations at 10% FCS; A: untreated, B: 0.2 µM PMA with 1 µM ATRA, C: 0.2 µM PMA with 0.8 µM VD3 and 1 µM ATRA as well as D: 0.1 µM PMA with 0.4 µM VD3 and 0.5 µM ATRA for 3 days. Subsequently the media was replaced with fresh media without drugs before imaging by bright field microscopy 2 days later. Experiment performed once.*



**Figure 4.2 Combined PMA and VD3 treatment enhances differentiation as assessed by substantial morphological changes.** Differentiation strategy #3: U937 cells were seeded at  $5 \times 10^4$ /well in 6 well plates and treated with the following combinations at 10% FCS; 0.2  $\mu$ M PMA and 0.8  $\mu$ M VD3 for 3 days. **A**, **B** and **C** show different fields of view of the same conditions. Subsequently the media was replaced with fresh media without drugs before imaging by bright field microscopy 2 days later. This differentiation protocol has been repeated approximately 50 times with consistent results.

#### 4.1.2. Can macrophage-like cells be “retrodifferentiated”?

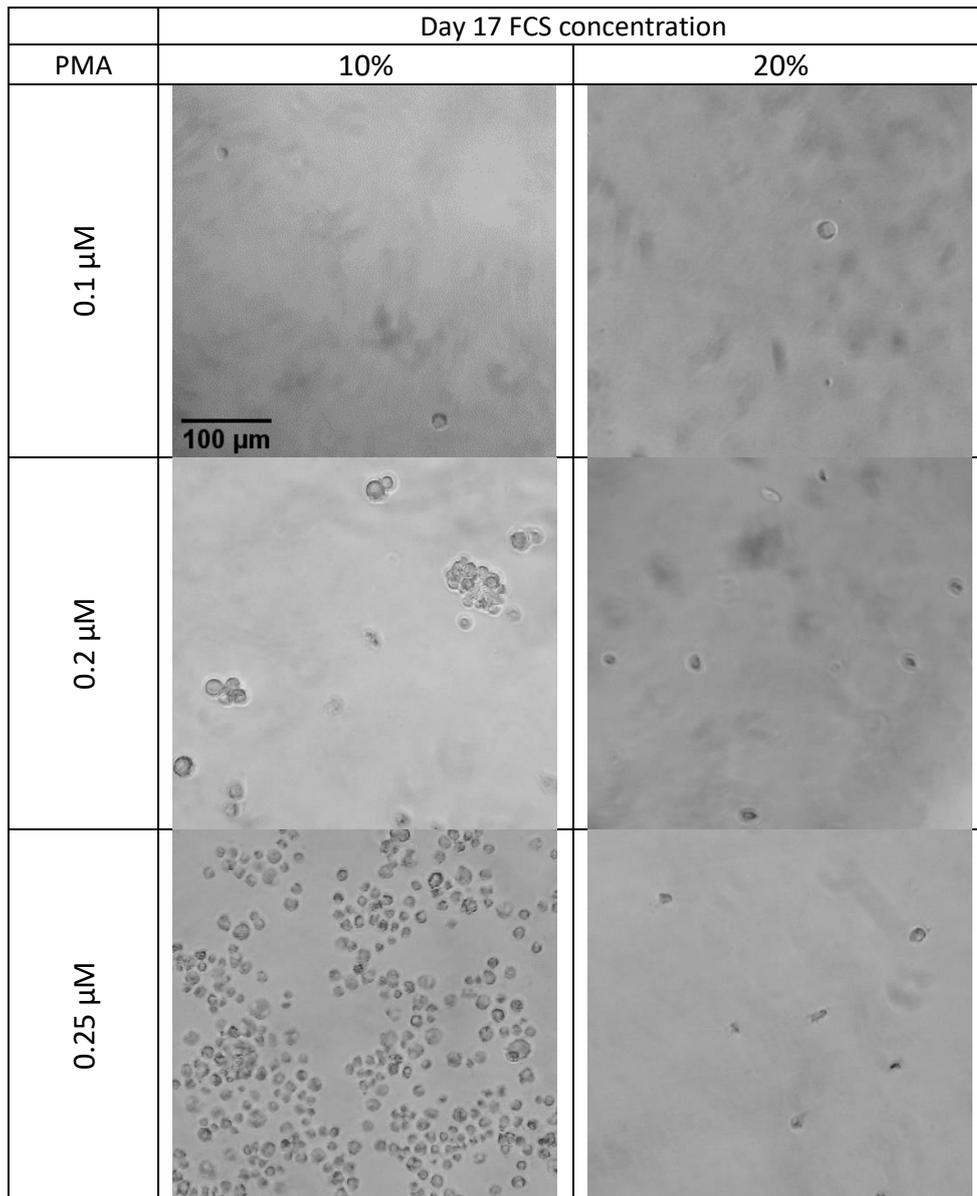
Retrodifferentiation, whereby previously differentiated cells are returned to monocyte-like suspension culture and regain proliferative activity, has previously been reported in the literature (Bertram et al., 2008; Hass et al., 1990; Meinhardt and Hass, 1995; Okada et al., 1995; Selle et al., 2007). Differentiated U937 cells do not proliferate, thus the ability to retrodifferentiate them would aid the discovery screen process. We hypothesised that differentiated cells surviving the infection process could be retrodifferentiated, amplified,

and archived for subsequent use. Iterative rounds of differentiation, selection by infection, and retrodifferentiation would therefore be achievable, allowing mutation effects to be clearly distinguished from stochastic 'noise'.

To assess the potential for retrodifferentiation, U937 cells were first treated with PMA following the differentiation strategy #2 timeline in Figure A.3A. Briefly, U937 cells were seeded at  $5 \times 10^5$ / well and treated with 0.1, 0.2, 0.25  $\mu\text{M}$  PMA for 7 days, before resting in drug free media for a further 7 days to complete the differentiation. Cells were then encouraged to retrodifferentiate, Figure A.3B, by washing and resting them in PMA-free culture media with 20% or 10% FCS on day 17. Higher FCS concentrations (e.g. 20%) often promote proliferation. By day 27, all the cells were circular, were more abundant, and were on the same focal plane as untreated cells in control wells (indicating detachment). On day 33 the wells were washed several times to retrieve any cells in suspension for counting as an indication of proliferative state – comparison to the number seeded (Table 4.1). Micrographs were taken (Figure 4.3) of cells remaining in the wells. Very few cells (approximately 3-5 cells/ well) remained in 0.1  $\mu\text{M}$  PMA wells after washing, regardless of FCS concentration. Many adherent cells remained in the 0.25  $\mu\text{M}$  PMA well with 10% FCS; demonstrating that at higher PMA concentrations, cells take longer to retrodifferentiate, or that fewer cells do so. Though similar numbers of adherent cells were seen in the wells at 0.2 and 0.25  $\mu\text{M}$  PMA at 20% FCS, the number of resuspended cells from 0.2  $\mu\text{M}$  PMA (Table 4.1) was roughly twice the number from 0.25  $\mu\text{M}$  PMA; indicating an increased ability to retrodifferentiate only at lower concentrations of PMA. Retrodifferentiation was also tested in full drug cocktail differentiated cells, but no cells reverted to suspension (data not shown); therefore, retrodifferentiation was found not to be a viable strategy in combination with the chosen differentiation protocol and a new screening strategy was required.

**Table 4.1 Attempts to retrodifferentiate macrophage-like cells after PMA treatment: differentiation drug and FCS concentrations influence the number of cells that return to suspension.** *The timeline for differentiation strategy #2 was followed (Figure A.3A and B) with the additional features specific to retrodifferentiation Figure A.3C). U937 cells were seeded at  $5 \times 10^5$ / well in a 24 well plate and treated with 0.1, 0.2, or 0.25  $\mu\text{M}$  PMA for 7 days before replacing with media without PMA. On day 17, the wells were thoroughly washed to resuspend any unattached cells. The media in one well for each concentration of PMA was replaced with 20% FCS and the remaining well per PMA concentration was maintained in 10% FCS. On day 33 the wells were washed several times to bring any detached cells back into the medium for counting. This experiment was performed once on PMA differentiated cells.*

Resuspended detached U937 cells			
Day 17 FCS concentration	0.1 $\mu\text{M}$ PMA	0.2 $\mu\text{M}$ PMA	0.25 $\mu\text{M}$ PMA
10%	$15.5 \times 10^5$	$2.58 \times 10^5$	$2.52 \times 10^5$
20%	$24.6 \times 10^5$	$14.3 \times 10^5$	$7.33 \times 10^5$



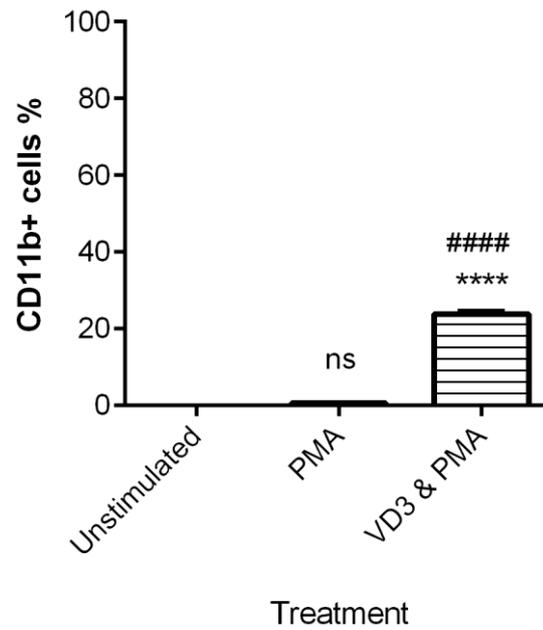
**Figure 4.3 FCS concentration increases U937 cell detachment after PMA treatment, though this was limited at desired PMA concentrations.** *Differentiation strategy #2: U937 cells were seeded at  $5 \times 10^5$ / well in a 24 well plate and treated with 0.1, 0.2, or 0.25  $\mu\text{M}$  PMA for 7 days before replacing the media without PMA. On day 17, the wells were thoroughly washed to resuspend any unattached cells. The media in one well for each concentration of PMA was replaced with 20% FCS and the remaining wells were maintained in 10% FCS. On day 33 the wells were washed several times to remove any detached cells and bright field micrographs were taken of what remained. This experiment was performed once on PMA differentiated cells.*

## 4.2. Characterisation of the differentiated U937 macrophage-like cell phenotype

### 4.2.1. Cell surface markers

To further characterise the nature of the differentiation of U937 cells using the 'differentiation drug-cocktail' and compare it to PMA treatment, an appropriate cell surface marker was required for use in flow cytometry. Multiple markers are used to distinguish between immune cell types *ex vivo*. The literature on U937 differentiation markers is conflicting, but CD11b was chosen because it appears to be the only marker showing increased cell surface expression with each of the drugs (Babina and Henz, 2003; Jensen Holly A. et al., 2015; Rahmani and Grant, 2002; Xiang et al., 2016; Yang et al., 2013; Zhang et al., 2008). Furthermore, CD11b is seen as a general differentiation marker, not specific to any immune cell type (Fang et al., 2017; Lu and Pitha, 2001; Xuening Wang et al., 2014).

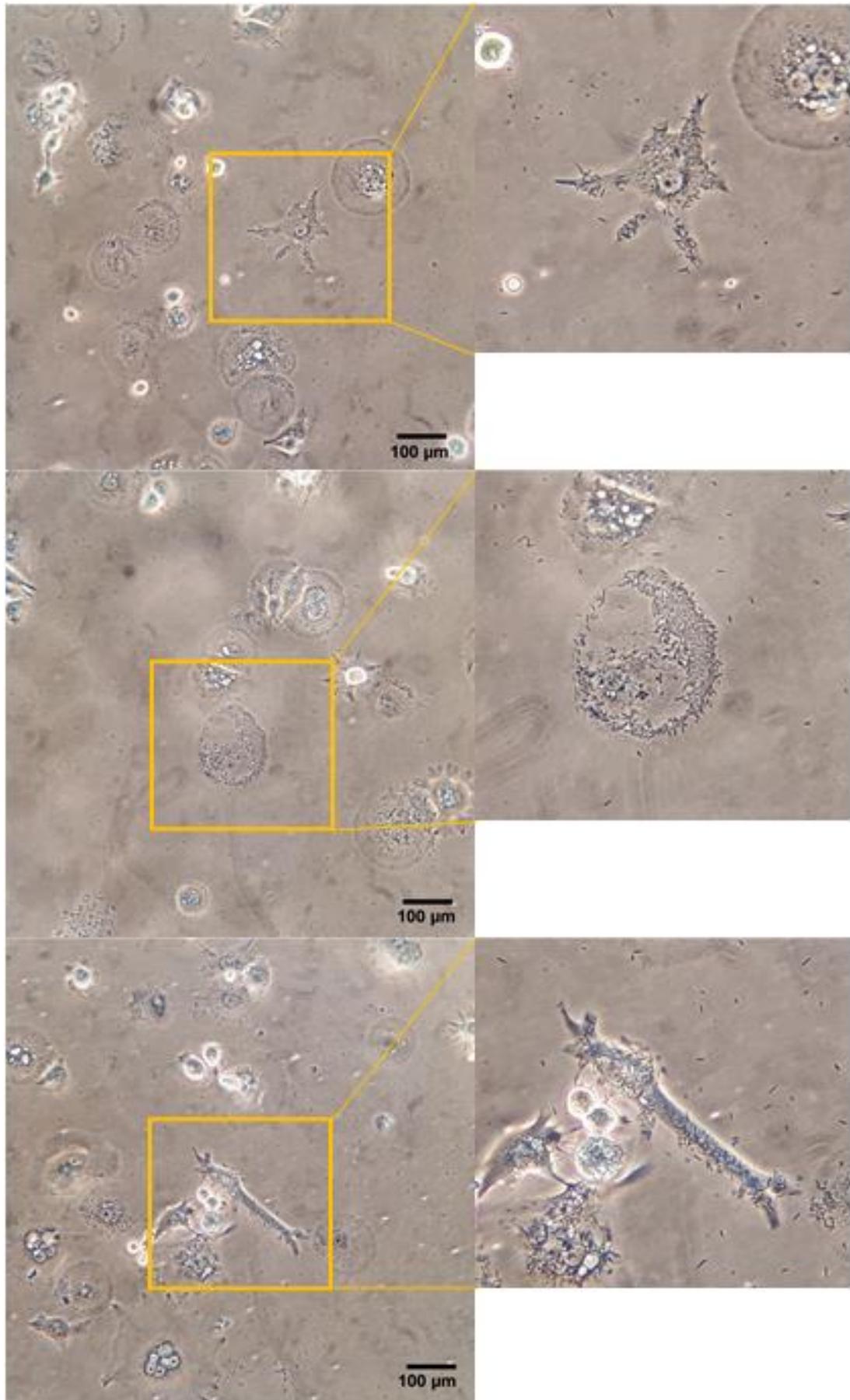
U937 cells were seeded at  $1.6 \times 10^4$  cells/cm<sup>2</sup> treated with 0.2 µM PMA or 0.2 µM PMA and 0.8 µM VitD3 (cocktail) or left untreated on day 0, before washing and feeding with differentiation-drug free media on day 3 and assessing by flow cytometry on day 6 (timeline in Figure 2.3). Cells were prepared for flow cytometry as described in Materials and Methods (Chapter 2), stained with CD11b, an isotype, or left unstained. The proportion of CD11b<sup>+</sup> cells as a percentage of live cells (gating strategy in Figure 2.4) is presented in Figure 4.4. PMA treatment alone failed to produce more CD11b<sup>+</sup> cells than the unstimulated population at this time point. The VD3 and PMA treated population had significantly more CD11b<sup>+</sup> cells than unstimulated cells or PMA treated cells, confirming its improved potential as a method for generating U937 macrophage-like cells.



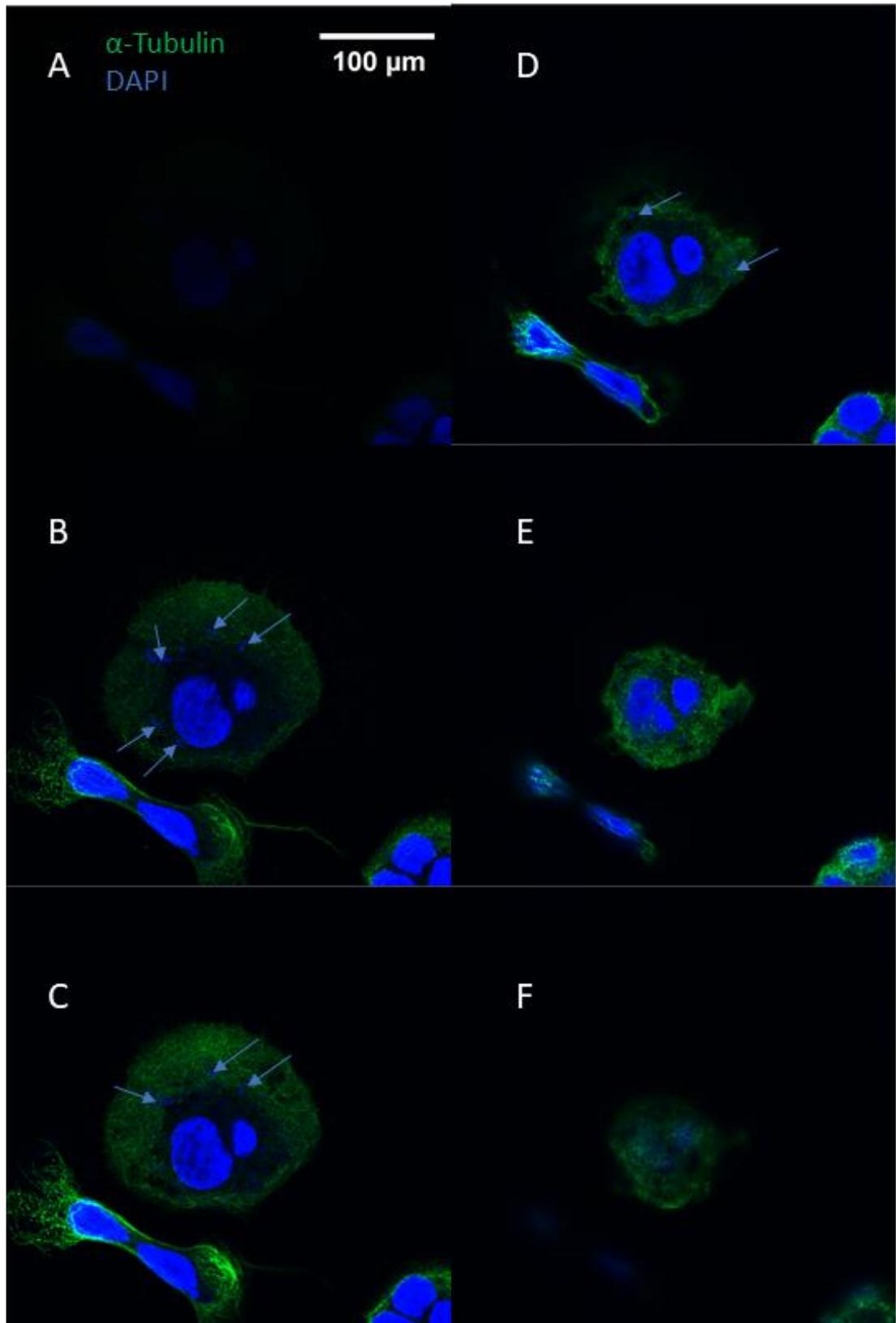
**Figure 4.4 The differentiation drug-cocktail treatment causes increased CD11b surface expression of U937 cells.** *CD11b<sup>+</sup> U937 cells as a percentage of live cells. U937 cells were treated with 0.2  $\mu$ M Phorbol 12-myristate-13-acetate (PMA), 0.8  $\mu$ M 1,25-dihydroxyvitamin D3 (VD3) in combination with 0.2  $\mu$ M PMA (VD3 & PMA), for 3 days before culturing for a further 3 days in drug-free media. On day 6, cells were tested by flow cytometry for CD11b surface expression. The height of the bars represents the mean of 3 replicate flasks and the error bar represents the standard deviation. The experiment was performed once. The data was analysed by one-way ANOVA and multiple comparisons were made with Tukey's multiple comparisons test between each treatment. ns stands for not significant, \*\*\*\* refers to  $p \leq 0.0001$  compared to unstimulated cells and ##### refers to  $p \leq 0.0001$  compared to PMA treated cells.*

#### 4.2.2. Bacterial phagocytosis

In addition to characterisation of cell surface marker expression, the ability of these cells to phagocytose bacteria was assessed. Bacterial phagocytosis is an important feature of bacterial-macrophage interaction *in vivo*, and the method of bacterial entry (e.g. through a T3SS or host initiated uptake) influences downstream intracellular interactions (Drecktrah et al., 2006; Lathrop et al., 2015). Therefore, a macrophage model that can phagocytose bacteria would be more physiologically relevant. The bacteria planned for use in the screening assay (*S. sonnei*, STM, and *F. tularensis* LVS) can all actively invade human cells. To distinguish host-mediated uptake from bacterial invasion, a non-invasive *Escherichia coli* strain (DH5 $\alpha$ ) was used. Drug-cocktail treated U937 cells were exposed to *E. coli* K12 strain DH5 $\alpha$  on day 6 post-drug treatment and imaged live by brightfield microscopy (Figure 4.5) or fixed with PFA for 15 minutes, washed with PBS and stained with mouse anti- $\alpha$ -tubulin, anti-mouse Alexa Fluor 488 and 4',6-diamidino-2-phenylindole (DAPI). The images in Figure 4.5 show a variety of macrophage-like U937 cell morphologies colocalised with *E. coli*, some of which have been digitally zoomed to highlight cells of interest. It is impossible in these images to determine whether the *E. coli* colocalised with the U937 cells are inside the cell or simply attached to the cell surface. Therefore, the fixed cells were imaged as a z stack on a confocal microscope, selected slices of which are presented (Figure 4.6). Based on these images, it is clear that the bacteria (visible as small DAPI stained rods in B, C, D and E) are within the cell (cell structure is stained green) and, thus, phagocytosis is occurring.



**Figure 4.5 *Escherichia coli* DH5 $\alpha$  bacteria colocalise with heterogeneous U937 differentiation drug-cocktail treated cells.** *Drug cocktail treated U937 (as described in differentiation strategy #3) exposed to E. coli K12 strain DH5 $\alpha$  on day 6 post-drug treatment. Micrographs taken by brightfield microscopy, example U937 cells displaying morphological changes and co-localising with bacteria (orange boxed area) have been digitally magnified (right). The experiment was performed once.*

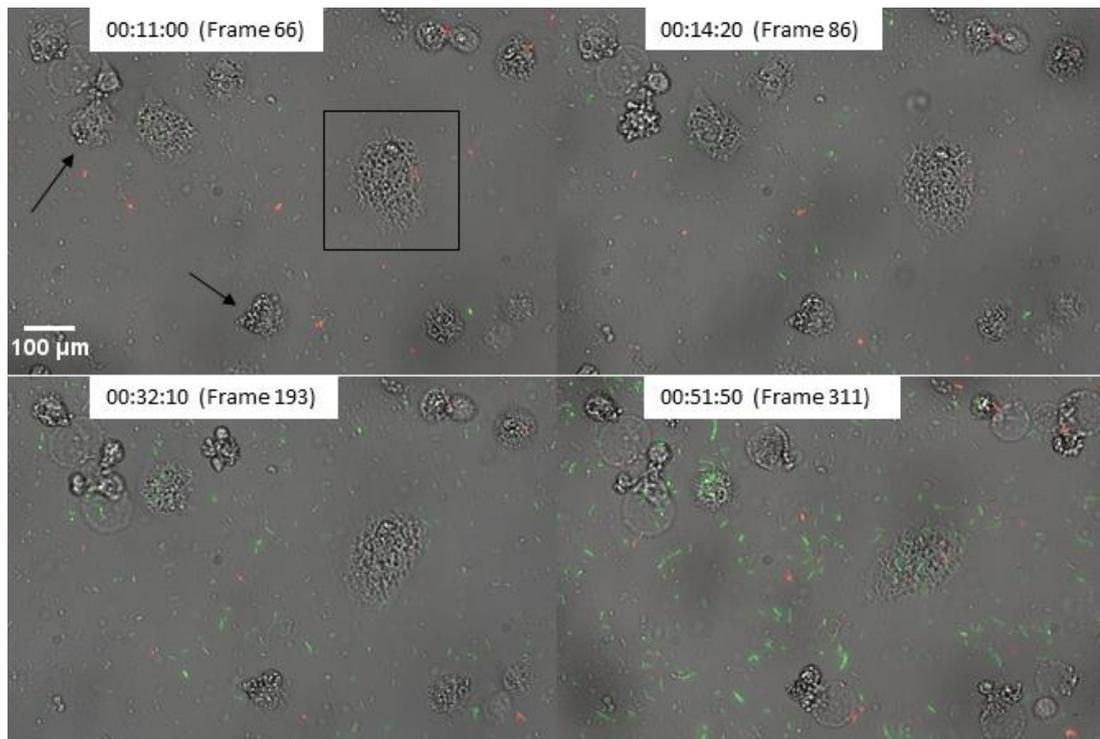


**Figure 4.6 Z-Stack: Non-invasive *E. coli* are actively internalised by drug-cocktail differentiated U937 cells.** Differentiation drug-cocktail treated U937 cells (as described in differentiation strategy #3) exposed to non-invasive *E. coli* K12 DH5 $\alpha$  6 days post-drug treatment. Cells were fixed with ice-cold PFA for 15 minutes, washed with PBS and stained with mouse anti- $\alpha$ -tubulin, anti-mouse Alexa Fluor 488 and 4',6-diamidino-2-phenylindole (DAPI). **A-F** Selected images from confocal z stack composed of 15 slices (slice 3, **A**; 5, **B**; 6, **C**; 9, **D**; 11, **E**; 14, **F**) of differentiated U937 cells colocalised with *E. coli* (DAPI-stained rods), 1575x. The experiment was performed once.

#### 4.2.2.1. Live imaging of differentiated U937 cells phagocytosing *Escherichia coli* DH5 $\alpha$

As an additional assessment of phagocytosis, Day 12 differentiation drug-cocktail treated U937 cells were inoculated with GFP expressing *E. coli* K12 strain DH5 $\alpha$  and mCherry expressing *E. coli* K12 strain DH5 $\alpha$ . A series of brightfield, green and red fluorescent images were taken at 200x magnification at 10 second intervals over a 90-minute time-period approximately 30 minutes post-inoculation. These images were then combined into a 38 second movie using ImageJ (Supplementary Figure 1). Selected images from the movie are presented in Figure 4.7 to highlight events of interest. The z value in the bar at the top of the movie gives the frame number, which is also listed in the images below. The macrophage-like cell indicated by the black box moves around the culture dish, which is likely chemotaxis (movement in response to signals) towards the bacteria; however, the MOI in this experiment is very high making it difficult to say with confidence whether the cell is migrating towards bacteria or simply moving at random. Black arrows indicate cells that appear to undergo cell death, these cells can be seen contracting and 'exploding' (the latter is also apparent in Figure 4.7). The host cell death could be due to the high MOI or a general susceptibility to cell death in response to the presence of bacteria.

Importantly, some bacteria do appear to become trapped upon interaction with the macrophages (follow the macrophage indicated by the black box in Figure 4.7 in the movie). Together Figure 4.5, Figure 4.6 and Figure 4.7 conclusively show that differentiation drug cocktail treated U937 cells are able to phagocytose bacteria; therefore, this protocol was taken forward as the 'macrophage model' of choice for screening purposes.



**Figure 4.7 Differentiated U937 cells move, engulf bacteria, and eventually die in response to bacterial inoculation.** Selected images taken from different time points of a time-lapse video at a 200X magnification. Day 12 differentiation drug-cocktail treated (strategy #3) U937 cells inoculated with mCherry and GFP expressing *E. coli* K12 strain DH5 $\alpha$  and imaged over a 90-minute time-period. Bacteria eventually overwhelmed culture due to lack of antibiotics, causing a high exposure to LPS and intracellular bacteria. Images were taken at 10 second intervals, thus frame 66 is 660 seconds (or 11 minutes) in. Cells indicated by black arrows are examples of those that appear to undergo cell death and the cell in the black box moves around. The experiment was performed once.

### 4.3. Generation of a U937 GT library

A randomly mutated cell line library suitable for use in a high throughput genetic screen must have high genome coverage to ensure that as many genes as possible are ‘queried’ in the screen. Not all genes can be screened, however, as successful insertions of the gene trap cassette require introns, which not all human genes contain. To generate a GT library with a large genome coverage, three independent GT sub-libraries were created in the U937 cell line using the pGTIV3 plasmid vector (plasmid details in Chapter 2). To further this effort, sub-libraries #1 and #3 were created with five separate nucleofections of the plasmid in 1,000,000 cells each, which were then pooled. Sub-library #2 was created with 4 separate nucleofections, but this generated an insubstantial number of mutations – no cells grew at the lowest serial dilution (data not shown). To estimate the mutational efficiency, an aliquot of the newly transfected cells was serially diluted until no resistant cells were observed after selection with antibiotic – the dilution factor allowed cell number in the neat library to be approximated. While not especially precise, this is best that can be achieved

with a suspension cell line that does not proliferate when differentiated (and adhered). Using this method, it was determined that the greatest coverage sub-libraries (#1 and #3) had an estimated 1,000-10,000 mutational events each. These two sub-libraries were deemed sufficiently dense in mutational events to create a pooled library, which would theoretically receive representatives of all of the mutated cells present in each. To compensate for the likely relative coverage (based on the method of generation), the sub-libraries were combined in a 1:2 ratio of #1 and #3, respectively.

#### 4.4. Determining suitable infection conditions for the differentiated U937 gene trap library as preparation for an infection-host cell survival screen

Due to the inability to use retrodifferentiation, a gene trap-based infection screen in differentiated U937 cells was required to be performed in one round only. To achieve an effective screen in this situation, three conditions have to be met. Firstly, a sufficient level of selection must be achieved: a host cell death rate of 80-95% should be targeted to reduce false positives/background noise. Secondly, screens should involve large numbers of differentiated cells. In this way, each individual mutation in the gene trap library will be represented in numerous cells in the screen, providing a large dynamic range that can be quantified after the screen. Thirdly, to adjust for a likely greater proportion of false positives and background noise, a comparatively large number of biological replicates should be prepared (e.g. six experiments). While the second and third conditions are simply issues of scale, the first of these issues, cytotoxicity, must be fully controlled to ensure a productive screen.

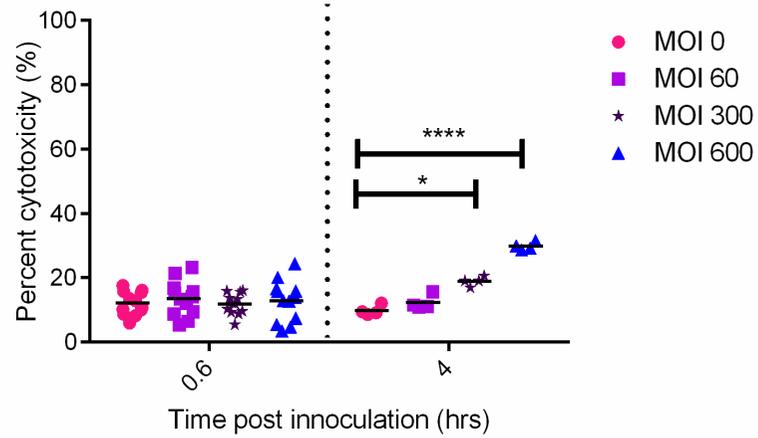
The LDH assay (used in Chapter 3) is one of multiple commonly used assays that measure cell death or cellular injury, including in infection assays (Shuai-Cheng et al., 2016; Volf et al., 2010; L. Wang et al., 2019; Wu et al., 2018). It is one of a few quantitative, directly correlated measures of inflammatory cell death and the only one that can be performed at scale on a budget. To fine tune the screening conditions, LDH activity (as a proxy for host cell death during STM infection) was measured over a number of time-points and MOIs in three experiments (Figure 4.8). The infections were performed according to the method outlined in Chapter 2 (Lactate dehydrogenase (LDH) activity assay to determine the extent of cell death after infection). At the time of the experiments, the *S. sonnei* OD<sub>600</sub> to CFU equation (Figure 2.5) was used to calculate the volume of bacterial culture required for the inoculum due to lack of a *Salmonella* specific equation. As the correlation between OD<sub>600</sub> and CFU for *S. sonnei* and STM are very different (Figure 4.3), this made it more difficult to prepare consistent MOI using subcultures with slightly different OD. As a result, different MOI (but of the same order of magnitude) were used in individual experiments. LDH max controls (positive cell death, lysis buffer treated) were inoculated with the same respective MOIs to control for increasing LDH expressing organisms (e.g. STM).

At 4 hours post inoculation, MOIs of 300 and 600 generated ~20% and ~30% percent cytotoxicity, respectively (Figure 4.8A). In an independent experiment an MOI of 300

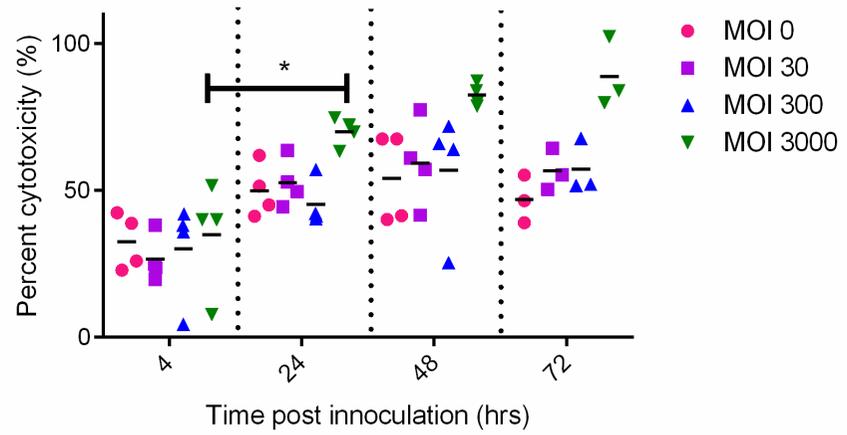
generated a mean cytotoxicity of 45% at 24 hours and 55% at 48 hours, a level that was retained 24 hours later (Figure 4.8B). In contrast an MOI of 3,000 generated much higher cytotoxicity, with means of 70%, 80% and 90% at 24 hours, 48 hours and 72 hours post inoculation, respectively (Figure 4.8B). A further experiment compared the cytotoxicity (presented as changes in OD due to lack of positive Max LDH controls) generated by MOIs of 50, 500 and 5,000 (Figure 4.8C) and demonstrated that the percent cytotoxicity peaks at 24 hours at an MOI of 5,000. The half-life of the LDH enzyme under the conditions of the assay is stated as 9 hours in the assay manual, indicating that the majority of remaining activity at 48 hours is likely due to enzyme accounted for at the 24-hour time point. Therefore, the vast majority of cell death at an MOI of 5,000 most likely occurred at 24 hours, making this MOI unsuitable for an infection screen, due to a potential lack of viable cells with good quality RNA. In comparison, an MOI of 3,000 provides a balance of a high level of cytotoxicity that does not fall as sharply with time, suggesting that some host cells are yet to die and leak LDH. Indeed, percent cytotoxicity remains high at 72 hours, suggesting that sufficient numbers of viable cells are likely to remain at 48 hours post inoculation – therefore, an MOI of 3,000 for 48 hours of infection was chosen for the screen.

Despite following the storage instructions for the cytotoxicity reagent, we observed that the reagent activity decreased over the course of use (Figure 4.9). The decrease in reagent activity decreased the OD range over which differences in LDH activity (and therefore percent cytotoxicity) could be seen. The decreased response range results in a flattening of the response – Figure 4.8B (in comparison to Figure 4.8A) is particularly affected by this.

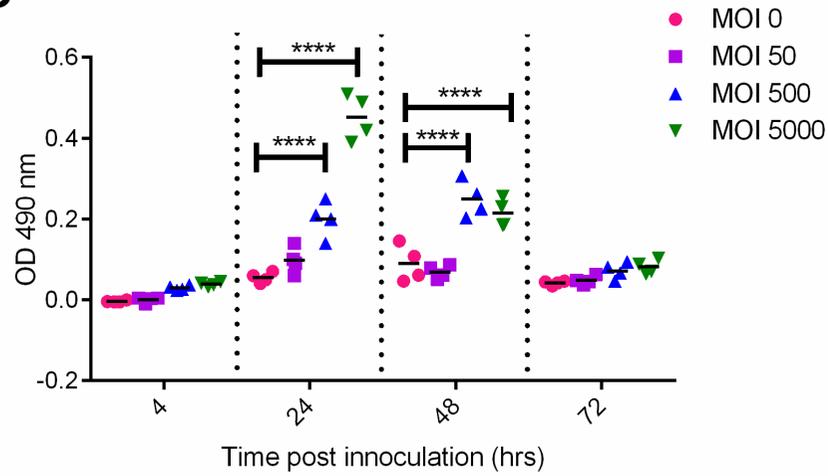
**A**



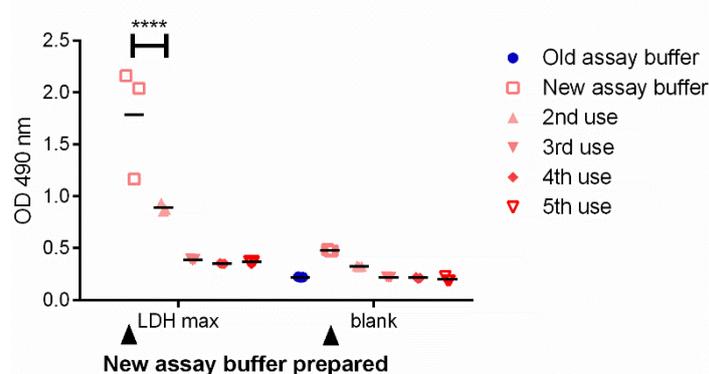
**B**



**C**



**Figure 4.8 STM infection cytotoxic effects on differentiated U937 cells assessed by LDH release: Cytotoxic effect increases with increasing MOI and time.** Differentiation drug cocktail-treated wild type U937 cells were prepared in 96 well plates (A), or 24 well plates (B, C). Cells were infected with STM on day 6 at the stated MOI, which was calculated using concentration at which the cells were seeded and the equation in Figure 2.8. The LDH assay was performed immediately after supernatant collection at each time point. Percent cytotoxicity was calculated with the equation given in the cytotox 96 kit instructions. Optical density (OD) values (minus the blank) are provided where no positive control was measured (C). Each graph shows data from one experiment (all of which were performed once), individual shapes represent technical replicates and the lines represent the mean. Individual experiments were statistically analysed with two-way ANOVA, after which post hoc comparisons were made between every condition using the Tukey test (using Graph Pad version 6.01). P values are only reported for comparisons between the same MOI across different time points and to MOI 0 at the same time point; \*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ , \*\*\*\*  $p \leq 0.0001$ .



**Figure 4.9 Cytotox96 reagent breaks down over multiple freeze-thaws, reducing the OD activity range.** The changes in LDH max (LDH activity from triplicate wells treated with kit lysis buffer) and blank (media without cells) are shown over the course of each freeze-thaw. Cytotox96 reagent was thawed from  $-20^{\circ}\text{C}$  and re-frozen before and after each use. The arrows below point to when fresh cytotox96 reagent was prepared from the substrate and assay buffer. The experiment was performed once.

#### 4.5. Use of the U937 gene trap library in a host cell survival screen during *S. enterica* ser. Typhimurium infection

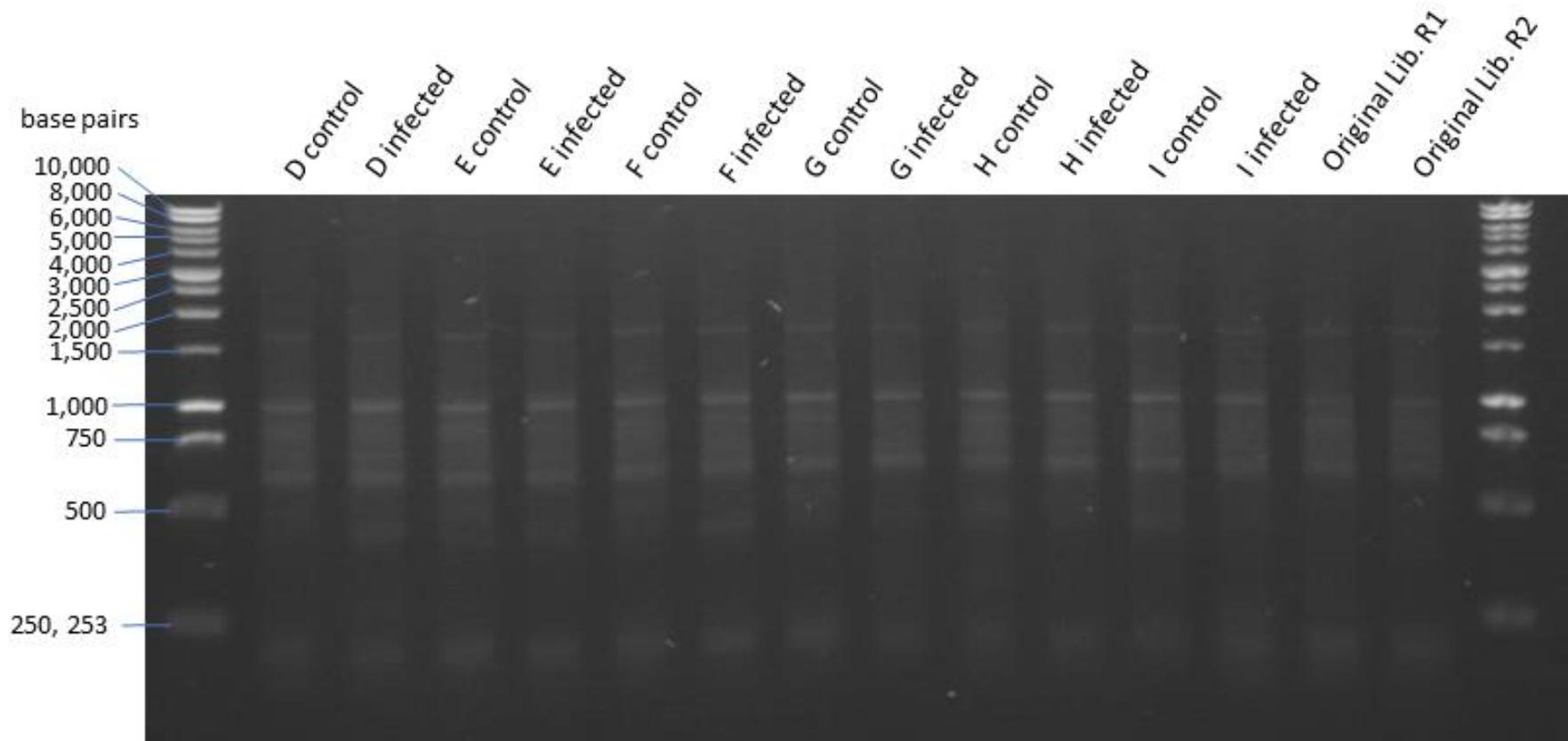
The combined U937 GT library (from sub-libraries #1 and #3) created in Section 4.3 was interrogated with STM infection as stated in Chapter 2, Materials and methods. To ensure a sufficient representation of the GT cell library, 10,000,000 U937 cells per sample were differentiated with the drug cocktail; which would mean an average of 1000 cells with the same mutation. Naturally, there would be some variation in the number of cells per mutant prior to selection. Aside from stochastic deviation, some mutant cells might possess a greater potential for proliferation, becoming more prevalent with time in culture. Therefore, the GT library was used for selection within a week of defrosting the combined library and a maximum of three passages were allowed. An  $\text{OD}_{600}$  reading of 0.6 from STM subcultures was sought as this optimises the SPI-1 TTSS protein expression for invasion (J. Wu et al., 2014). An MOI of 3,000 was intended, but due to the centrifugation methods available and the volume of STM subculture required, an approximate MOI of 2,400 (empirically

determined from bacterial suspensions prepared in the same way as the inoculum, n=3) was reached. The experiment was repeated six independent times over several weeks (biological replication), each with a duplicate pair of plated drug cocktail differentiated U937 GT library populations – one infected, the other uninfected. The uninfected population also controlled for replicative phenotypes observed in the absence of infection (Québatte and Dehio, 2017), as the cells are subjected to the same conditions aside from infection. After 2 days of infection, prior to RNA extraction, extensive cell death was seen in the infected GT population (data not shown); whereas the control population looked healthy. The concentration and quality of the extracted RNA was measured on a 2100 Bioanalyzer (see Appendix B). The RNA quality in three samples was poor, therefore the experiments for each of these were repeated to retain the number of biological replicates. The RNA quality from all remaining samples was sufficient for next generation sequencing (NGS) (see Appendix B for further detail).

To compare the representation of mutant cells in the originally synthesised library to the uninfected, differentiated cells, RNA was extracted from 1,000,000 undifferentiated U937 GT library cells (labelled Original Lib. R1 and R2). The RNA quality and concentration for these samples was also very good (see Appendix B).

The infected samples consistently had lower RNA concentrations than the control samples for each experiment (Appendix B); which was expected as the number of surviving cells would be considerably lower and the RNA from any dead cell would have likely degraded (Thomas et al., 2015). To have the same level of sequencing depth, the total RNA was normalised to the sample with the smallest concentration (I infected, Appendix B) in preparation for cDNA synthesis.

As the intention was to make a sequencing library only from the GT fusion transcripts, a specific primer matching mRNA poly-A tails and an additional tail sequence was used for cDNA synthesis. The trapped gene sequences were amplified in 2 consecutive rounds of nested PCR, as described in Chapter 2. Nested PCR was used to improve the amplification selectivity, as only one primer of the initial pair was specific to the GT vector (the other matched the tail of the cDNA primer). Figure 4.10 shows the purified results of the consecutive rounds of PCR which have been separated on an agarose gel by electrophoresis and the relationships between samples is described in Table 4.2. Each lane containing RACE PCR material had a smear between 500-3,000 bp as well as several distinct bands (as expected) and similar to others' GT library RACE PCR results (B Pickard, personal communication, 2019 20<sup>th</sup> May). The smear indicated the presence of many DNA products of varying sizes which is desired as it should indicate many sampled mutations. The distinct bands emerging from the background are likely overrepresentation of single PCR products.



**Figure 4.10 Final GT screen RACE PCR material prior to DNA library preparation and sequencing.** Purified RACE2 PCR material of STM U937 GT library screen samples. The letter at the beginning of each sample name refers to the experiment which it came from (experiments D – I). RACE PCR samples labelled ‘infected’ are from the STM-infected population of U937 GT library cells from the respective experiment; while samples labelled ‘control’ are derived from the mock-infected population. Experiments A and C resulted in low quality RNA, so were discarded. Experiment B was terminated before RNA extraction. Original Library (Original Lib.) biological replicates 1 (R1) and 2 (R2) are RACE PCR samples each prepared from  $1 \times 10^6$  U937 GT library cells without differentiation treatment or infection, defrosted and passaged from different frozen aliquots of the combined GT library. Total RNA was normalised to the I infected sample prior to first strand cDNA synthesis, RACE1 and RACE2 PCR. A total volume of  $180 \mu\text{L}$  of RACE2 PCR material was purified and eluted in  $28 \mu\text{L}$  of nuclease free water. Purified samples were loaded with water ( $0.5 \mu\text{L}$  and  $4.5 \mu\text{L}$ , respectively).

**Table 4.2 Description of RACE PCR amplicon samples from the STM U937 screen presented in Figure 4.10.** Each experiment from the infection screen functions as a replicate pair, the sample derived from the mock-infected differentiated U937 GT population (Control) serving as a control for the sample derived from the STM-infected differentiated U937 GT population (Infected). The control population is expected to have a similar representation of GT fusion transcripts for mutations that promote differential cell representation regardless of infection (such as differences in proliferation). Original Library (Lib.) R1 and R2 are derived from independent frozen aliquots of the GT library after two passages each. The Original Library samples were intended as representatives of the scope of the GT library without differentiation or infection.

Sample name	Experiment	Description of population from which sample was derived
D Control	Infection screen - Experiment D	Mock-infected
D Infected		STM-infected
E Control	Infection screen - Experiment E	Mock-infected
E Infected		STM-infected
F Control	Infection screen - Experiment F	Mock-infected
F Infected		STM-infected
G Control	Infection screen - Experiment G	Mock-infected
G Infected		STM-infected
H Control	Infection screen - Experiment H	Mock-infected
H Infected		STM-infected
I Control	Infection screen - Experiment I	Mock-infected
I Infected		STM-infected
Original Lib. R1	Untreated GT library – Experiment 1	Undifferentiated, uninfected
Original Lib. R2	Untreated GT library – Experiment 2	Undifferentiated, uninfected

#### 4.6. Use of the same U937 gene trap library in a host cell survival screen using *F. tularensis* subsp. *holarctica* Live Vaccine Strain infection

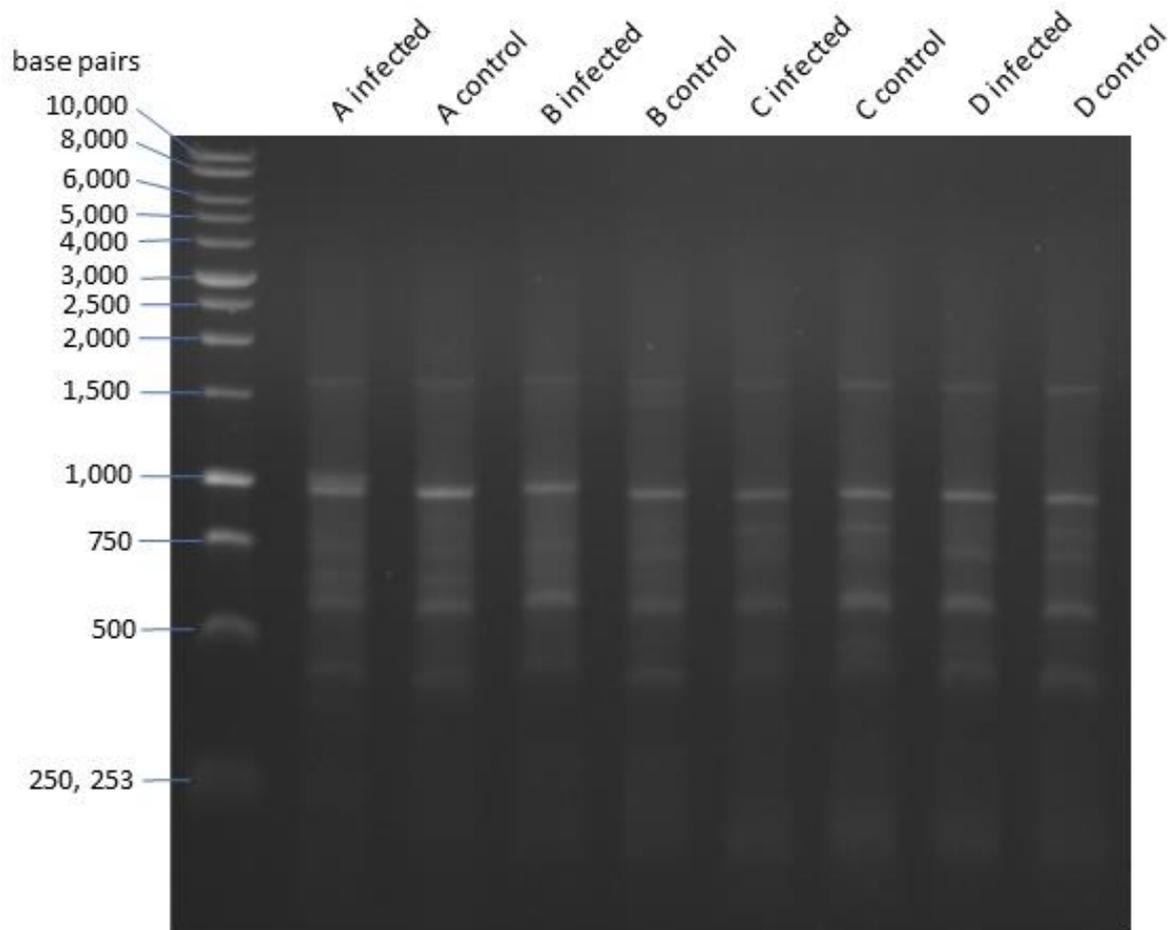
The combined U937 GT library was interrogated with *F. tularensis* LVS infection as stated in Chapter 2, Materials and methods in collaboration with Helen Flick-Smith at DSTL. A second screen using a different intracellular bacterium with different mechanisms of entry and different cellular niche was of interest to identify common genes or cellular processes involved in host survival or resistance to infection. *F. tularensis* LVS is a model pathogen for the *F. tularensis* subsp. *tularensis* Schu S4 strain which has been listed as a category A bioterror agent by the Centers for Disease Control (Centers for Disease Control and Prevention, 2018), thus requiring urgent biomedical research. The U937 GT library was prepared and differentiated at the same scale as in the STM U937 GT screen (Section 4.5). The inoculum from each experiment was serially diluted and plated on agar to generate CFUs to determine the MOI (Table 4.3).

**Table 4.3 Summary details of *F. tularensis* LVS U937 GT infection.** *U937 GT library cells defrosted from the initial cryotubes sent down to DSTL Porton Down were defined as passage 1. The concentration of the inoculum in CFU /mL and the MOI were calculated from CFU counts from the serially diluted inoculum.*

Experiment	Passage number	Inoculum CFU /mL	MOI
A	11 (passage control)	8450000	70
B	3	7300000	61
C	3	8800000	73
D	3	14000000	117

The experiment was repeated 3 times at passage 3 (Experiments B, C and D) with 1 technical replicate for infected and control (mock-infection). Experiment A was performed at passage 11 and served as a passage control to determine how much of an effect the passage number had on the mutation representation in the population. After 2 days of infection, prior to RNA extraction, a substantial amount of cell death was seen, though not as much as in the STM U937 GT screen. The quality of the total RNA for all samples as measured by a Tapestation was high (Appendix B).

Total RNA was checked for sterility (method in Chapter 2) before shipment to the University of Strathclyde where the RNA was processed and reverse transcribed into cDNA for RACE PCR (in the same manner as in Section 4.5). The purified results of consecutive nested rounds of RACE PCR were separated on a agarose gel by electrophoresis to provide evidence of their nature and quality (Figure 4.11), which was comparable to the RACE PCR products generated from the STM screen (Figure 4.10). The source and relationship of each of RACE PCR sample is described in Table 4.4. Sequencing library creation and sequencing (by Glasgow Polyomics) was performed in the same manner as the previous screen.



**Figure 4.11 Final GT screen RACE PCR material prior to DNA library preparation and sequencing.** Purified RACE2 PCR material of the *F. tularensis* U937 GT library screen samples. The letter at the beginning of each sample name refers to the experiment which it came from (experiments A – D). RACE PCR samples labelled ‘infected’ are from the *F. tularensis* LVS infected population of U937 GT library cells from the respective experiment; while samples labelled ‘control’ are derived from the mock-infected population. Total RNA normalised to the experiment C infected sample prior to first strand cDNA synthesis, RACE1 and RACE2 PCR. A total volume of 140  $\mu\text{L}$  of RACE2 PCR material (half GC enhanced) was purified and eluted in 30  $\mu\text{L}$  of nuclease free water. Purified samples (1  $\mu\text{L}$ ) were loaded with 7.3  $\mu\text{L}$  of water onto the gel.

**Table 4.4 Description of RACE PCR amplicon samples from the *F. tularensis* U937 screen presented in Figure 4.11.** Each experiment from the infection screen functions as a replicate pair, the sample derived from the mock-infected differentiated U937 GT population (Control) serving as a control for the sample derived from the *F. tularensis* infected differentiated U937 GT population (Infected). The control population is expected to have a similar representation of GT fusion transcripts for mutations that promote differential cell representation regardless of infection (such as differences in proliferation). Experiment A (see Table 4.3) serves as a passage control to determine how strongly passaging the GT library prior to differentiation affects changes in GT fusion transcript representation.

Sample name	Experiment	Description of population from which sample was derived
A Control	Infection screen - Experiment A, passage controls	Mock-infected
A Infected		<i>F. tularensis</i> -infected
B Control	Infection screen - Experiment B	Mock-infected
B Infected		<i>F. tularensis</i> -infected
C Control	Infection screen - Experiment C	Mock-infected
C Infected		<i>F. tularensis</i> -infected
D Control	Infection screen - Experiment D	Mock-infected
D Infected		<i>F. tularensis</i> -infected

## 4.7. Discussion

A U937 differentiation protocol suitable for use in macrophage infection screening assays was developed through three rounds of optimisation. The final protocol generates differentiated U937 cells that are adherent, morphologically different to monocytes, and have reduced or terminated proliferation as well as the ability to phagocytose bacteria. Indeed, the drug cocktail differentiated U937 cells described here have a similar appearance to PMA treated and rested THP-1 cells prepared by Daigneault et al., (2010), who describe these cells as predominantly classically ‘M1’ activated but also sharing some alternative ‘M2’ activation features. Additionally, a greater proportion of the drug cocktail differentiated U937 cells express the differentiation marker CD11b than PMA treatment alone.

A U937 GT library suitable for phenotype discovery screening was generated from a combined pool of multiple independent rounds of random GT mutagenesis. The conditions for a large scale STM infection screen were optimised before use in six replicate experiments (each with an infected and control pool of cells). A second infection screen using *F. tularensis* LVS as the infectious agent was also performed. Each of these sets of experiments generated RNA of suitable quality for downstream analysis.

### 4.7.1. An optimal infection MOI and time frame for stringent selection

In general, STM infection of drug cocktail differentiated U937 cells did not produce significant cytotoxicity below MOIs in the hundreds and so MOIs in the thousands were required to produce levels of cytotoxicity required for efficient screen selection (summarised in Table 4.5). The majority of STM-generated cell death described here occurs between 4 and 24 hours at high MOIs. By contrast, the situation *ex vivo* / *in vitro* when using primary macrophage cultures is very different. STM causes SPI-I T3SS mediated inflammasome activation and cell death in human LPS-primed

monocyte-derived macrophages (MDMs) at early time points at comparatively low MOIs (e.g. 50% cytotoxicity at an MOI of 20 over 4 hours) (Casson et al., 2015). Low MOIs also cause significant cell death in mouse macrophages; 85% cytotoxicity during 8 hours in *ex vivo* infection (Hersh et al., 1999). In experimental comparisons of primary and cell line macrophages the former often have stronger and more complete transcriptional and proteomic responses (Andreu et al., 2017; Levenson et al., 2018; Tedesco et al., 2018); therefore primary macrophages may also have a stronger cell death response.

4.7.1.1. The selection procedure was not too stringent - the total RNA was still of good quality

An STM MOI of 2,400 (technical detail in Section 4.5) was used for U937 GT screen and extensive cell death was seen in the infected GT population. Overall, good quality RNA was obtained from infected GT populations, indicating that surviving cells of sufficient health were available to extract RNA from; as RNA in cells undergoing apoptosis becomes degraded (Del Prete et al., 2002; Thomas et al., 2015). Using the same argument, the contribution of RNA from potential remaining dead 'susceptible' host cells to the final sequenced material is likely to be low.

**Table 4.5 Summary of STM infection of cocktail differentiated wild type U937 cells to find optimal infection conditions.** LDH activity as a proxy for host cell death increased ( $\uparrow$ ), decreased ( $\downarrow$ ) or was the same (=) as without bacteria; where not statistically significant but still mentioned, the indicated direction has a \*. All experiments listed were performed once, except STM infection at MOI 300 for the 4-hour time point. In this case, the pattern was different between two experiments, the result for each is listed with a (1) adjacent to it. Time points or assays not performed are listed as ND.

MOI	Time post inoculation (hours)				
	0.6	4	24	48	72
30	ND	=	=	=	=
50	ND	=	=	=	=
60	=	=	ND	ND	ND
300	=	$\uparrow$ (1), = (1)	=	=	=
500	ND	=	$\uparrow$	$\uparrow$	=
600	=	$\uparrow$	ND	ND	ND
3000	ND	=	$\uparrow$ *	$\uparrow$ *	$\uparrow$ *
5000	ND	=	$\uparrow$	$\uparrow$	=

#### 4.7.2. Limitations of U937 differentiation characterisation

Many authors have presented cell surface or transcript expression of CD11b as a means of confirming PMA-induced U937 differentiation (Babina and Henz, 2003; Basoni et al., 2005; Boukes and van de Venter, 2012; Deszo et al., 2004; Guest et al., 2008; Kang et al., 2004; Paulsen et al., 2011). Boukes and van de Venter et al., (2012) observe that PMA treatment encourages more cells to express CD11b (~70%) at 48hr compared to 24hr, but also that higher concentrations of PMA produce lower CD11b surface expression. Paulsen et al., (2011) demonstrate that longer periods of treatment with PMA (72hr>48hr>24hr) at 25 nM increases the mean fluorescence intensity of CD11b. Both use higher seeding densities than have been used here ( $1.6 \times 10^4$  cells/ cm<sup>2</sup> vs  $1 \times 10^5$  cells/ cm<sup>2</sup> (Boukes and van de Venter, 2012),  $5\text{-}2.5 \times 10^5$  cells/ cm<sup>2</sup> (Paulsen et al., 2011)). Thus, if lower concentrations of PMA were used and cell surface expression levels of CD11b was measured on day 3 of PMA treatment, expression levels would likely have been much higher.

Cell surface markers have a limited ability to characterise subsets of mononuclear phagocytes derived *ex vivo* as this changes based on tissue and inflammatory state (Guilliams and van de Laar, 2015). Interestingly, CD11b cell surface expression actually decreases in human monocytes differentiated with M-CSF into macrophages (Da Costa et al., 2018) and CD11b has also been used as a M2 “alternatively activated” macrophage marker (Pagie et al., 2018). Daigneault et al., (2010) report an alternative method of characterising monocyte cell line differentiation; primarily by comparing differentiated cell line responses to (human) monocyte-derived macrophage (a common, physiologically closer macrophage model) responses in a series of physiological tests, as well as comparisons of cell surface marker expression. Interestingly, the majority of cells in CD14<sup>+</sup> CD16<sup>-</sup> and CD14<sup>+</sup> CD16<sup>+</sup> human blood monocyte subsets express CD11b (Cros et al., 2010). Perhaps U937 cells immediately after PMA treatment are more representative of these subsets than tissue macrophages, though comparisons based on the expression of one cell surface marker are superficial at best. The characterisation of the chosen differentiation protocol was limited due to time constraints, but further characterisation would provide more information about the context in which the screens were performed. Features such as expression of (further) cell surface macrophage

markers, production of ROS and RNS as well as phagocytosis of STM would help comparisons with human MDMs, *in vivo* phenotypes and other cell line infection screens.

#### 4.7.2.1. Exposure of cocktail differentiated U937 cells to bacteria

Upon imaging of *E. coli* K12 strain DH5 $\alpha$ -inoculated differentiated U937 cells, it was a surprise to find that many of the mammalian cells died over the course of the 90-minute imaging period. One would think that it would be evolutionarily advantageous to retain phagocytic cells such as macrophages during an active infection if they were able to destroy the invading species without resorting to programmed cell death (PCD). Indeed, Yang et al., (2014) report that mouse peritoneal macrophages are able to modulate their PCD response to cytosolic flagellin according to the source (*E. coli* K12 strain MG1655 or *S. enterica* ser. Typhi). Nevertheless, some macrophages succumbed to PCD in response to cytosolic *E. coli* flagellin, indicating that PCD is a physiological response to non-pathogenic bacteria. Additionally, the MOI used in the time lapse experiment is clearly very high and the experiment was performed in near atmospheric CO<sub>2</sub> (thus reducing the buffering capacity of the media). A combination of these factors was probably what caused the high frequency of PCD.

#### 4.7.3. Retrodifferentiation is not possible in drug cocktail differentiated U937 cells

U937 cells differentiated with low concentrations of PMA (in differentiation strategy #2) appeared able to return to suspension and begin proliferating as previously described (Bertram et al., 2008; Hass et al., 1990; Meinhardt and Hass, 1995; Okada et al., 1995; Selle et al., 2007). Unfortunately, this process occurred to lesser extent at the required PMA concentration (0.2  $\mu$ M) and failed to occur at all subsequent to drug cocktail differentiation (strategy #3). As the PMA concentrations tested produce a heterogeneous morphology, it is unclear if the different phenotypes (such as PMA- vs drug-cocktail generated) result in different retrodifferentiation ability. Selle et al., (2007) found that U937 cells treated with 5 nM of PMA for up to 72 hours generated a morphologically heterogeneous population much like that presented in differentiation strategy #2 with circular and elongated shapes. In that publication, the majority of the cells 17 days post PMA treatment initiation were visibly in suspension. Therefore, there may be no difference in the ability to retrodifferentiate between PMA-treated cells demonstrating different observable morphology. However, different differentiation regimes are likely to produce more profound differences in phenotype than within a single differentiation treatment. The major difference in differentiation treatment between strategy #2 and #3 was the addition of VD3; indicating that this may have prevented retrodifferentiation.

#### 4.7.4. Future work

A common theme to these infection screens is the idea of identifying a universal novel host gene or process associated with permissive/ successful bacterial infection that can then be pharmacologically inhibited. The U937 differentiated GT screen has thus far already been performed for two pathogens, namely STM and *F. tularensis* LVS. Using the same GT library in screens of different pathogens could prove useful for identifying potential genes common to all, though given the potential for higher false positives and false negatives in high throughput experiments, any common differentially represented trapped genes should be confirmed independently.

#### 4.7.4.1. *S. sonnei* strain 86 as candidate pathogen for a future infection screen

As a globally important intracellular pathogen with potential as a bioterror agent, *S. sonnei* would be a good candidate for use in a potential third infection survival screen in differentiated U937 GT library cells. *In vivo*, *S. sonnei* mainly infects epithelial cells, microfold cells, dendritic cells and macrophages; the last of these is an intracellular niche *S. sonnei* uses to avoid neutrophils.

Therefore, macrophage-like cells are a particularly relevant *in vitro* model to study host factors important in *S. sonnei* infection. To initiate an investigation of suitable infection conditions for selection, a preliminary experiment was performed using *S. sonnei* strain 86 at an MOI of 20 in wild type U937 cells. At 24 hours post inoculation almost all host cells were still alive (as observed by microscopy), but 29 hours later approximately 95-99% were not viable (data not shown).

The capacity of *S. sonnei* to cause a near complete level of cell death would be a powerful selection pressure for a mutant U937 library; but the near totality of death at this time point is also cause for concern if used in a discovery screen. Enough living U937 cells must survive to analyse subsequently (e.g. by RNA sequencing (RNA-Seq)) and the size of the surviving U937 population observed (green cells) is small enough to perish in replicate experiments due to experimental variation of infection conditions. Nevertheless, this is an important step towards identifying suitable *S. sonnei* infection conditions for a future screen. On the other hand, maintaining the virulence of *S. sonnei* in large experiments has proven challenging in my hands (data not shown); as *S. sonnei* is well known to stop expressing the virulence proteins that (in addition other activities) increase the rate of macrophage cell death, if the culture temperature drops below 37°C for a sufficient length of time (McVicker and Tang, 2017; Sandstrom et al., 2019; Suzuki et al., 2014). This challenge would need to be overcome as a factor to enable use of the pathogen in an infection survival screen.

#### 4.7.4.2. Iterative rounds of selection as an additional approach in a monocyte-like model

Another approach to finding resistance/susceptibility genes would be to screen the U937 GT library as an undifferentiated suspension population on monocyte-like cells in iterative infections/recoveries. This would have the advantage of identifying greater numbers of differentially represented genes, as the repeated selection would have a greater success of removing genes identified by chance. However, the immature state of the cells may mean that mature macrophage host defence processes are not functioning and are therefore 'invisible' to this functional screening approach.

# Chapter 5

## Bioinformatics analysis and phenotypic validation of U937 GT library infection screens

## 5. BIOINFORMATICS ANALYSIS AND PHENOTYPIC VALIDATION OF U937 GT LIBRARY INFECTION SCREENS

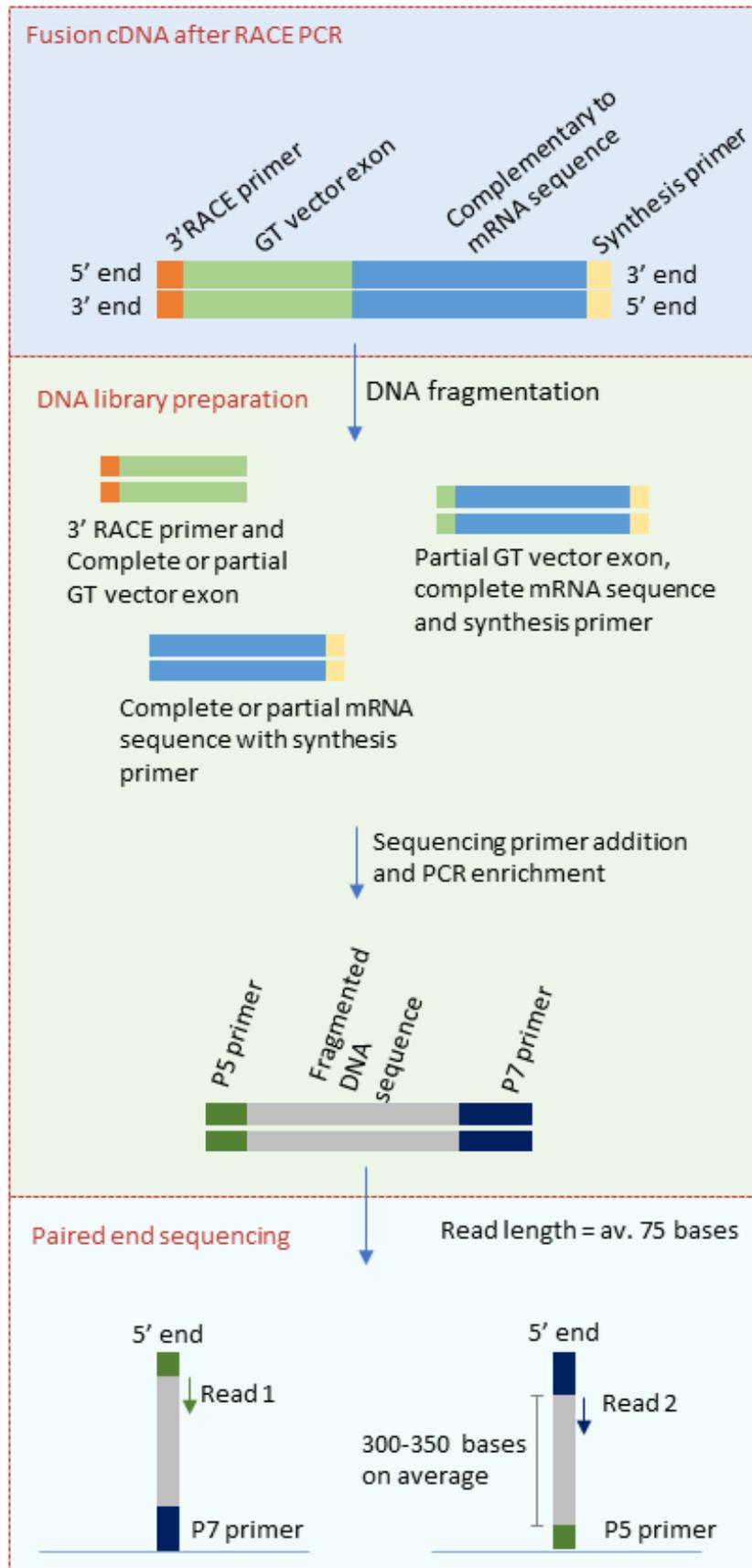
Chapter 4 identified a suitable timeframe and MOI for *Salmonella enterica* serotype Typhimurium SL1344 (STM) infection of differentiated U937 macrophage-like cells. These conditions generated a sufficient level of cell death (selection pressure) while still providing sufficient quality of RNA for analysis. This timeframe and MOI were therefore used to select for differentiated U937 gene trap mutants with greater or lesser capacity for survival than mock-infected U937 GT library cells. A similar set of experiments was performed with *Francisella tularensis* Live Vaccine Strain (LVS) as the infectious agent. Fusion cDNA amplicons (made up of both gene trap and trapped endogenous gene sequences) from these screened populations of cells are each present in a quantity proportional to the number of cells that carried them. Therefore, NGS of these amplicons permits the accurate quantification of allele frequency changes (reflecting benefit or harm to cell survival) for every mutation in the population after selection.

In this chapter, the process of identifying over- and under-represented trapped gene sequences using an online bioinformatics tool kit is described. Each step in the process has been examined, and optimisations made and discussed. Gene ontology enrichment analysis was performed for each screen to identify the processes, functions and components that may be particularly important to STM infection and, or, the host response. Protein interaction network analyses of the combined list of differentially represented genes is undertaken to infer potentially important proteins missing from the screens, as well as potential pathway connections between the screens. A single over-represented trapped gene was taken forward for validation through independent genetic mutation and the results of an initial infection experiment are presented. Finally, the preliminary results of a pharmacological intervention based on this new bioinformatics data are presented.

### 5.1. Summary of pre-sequencing and sequencing processes

As described in Chapter 4, the STM U937 GT screen was composed of six replicates of experiments (D, E, F, G, H, I) each with a pair of infected and uninfected U937 cell populations, with an additional two samples taken as representative of the original library. The *F. tularensis* LVS U937 GT screen was composed of four independent replicates (A, B, C, D) each with a similar infected and uninfected U937 cell population pair. RNA was extracted from each of these populations. Once the RNA had been converted to cDNA, it was amplified via RACE-PCR and all amplicons sent to Glasgow Polyomics where they were enzymatically fragmented with an NEB FS Ultra II DNA kit to create shorter, more uniformly distributed, sequences (Figure 5.1). Sequencing primers were also added to each end before sequencing individual strands from the 5' end to generate paired-end 75 base reads. Most reads would likely have terminated before the middle of the DNA fragment, given that the fragments were an average of 300-350 bp in length (depending on the

experimental sample). This would leave a gap in the centre of the fragment that was left unsequenced. Despite this, paired-end sequencing is particularly useful for this application, as it ensures the trapped gene is sequenced, even if one read of the pair is composed solely of GT vector. Paired reads also allow more accurate identification of genetically similar but functionally different genes (e.g., gene vs. pseudogene, paralogues); provided both reads in the pair contain DNA from the endogenous gene sequence, as well as being of sufficient length and quality. A minority of DNA fragments may have been shorter than 75 bases, allowing a portion of the 3' primer to be sequenced as well. The presence of sequencing primers was evaluated during quality assessment (described in Section 5.2.1.4).



**Figure 5.1 Summary of RACE PCR amplicon composition and technical processes required for Next Generation Sequencing.** *The DNA output from RACE PCR is a fusion of the GT vector exon (containing the 3'RACE primer sequence (Table 2.4)), cDNA complementary to a partial endogenous 'trapped' mRNA transcript and a synthesis primer (Table 2.4) sequence. To begin DNA library preparation, the DNA was enzymatically fragmented, splitting the PCR products into shorter sequences (potential sequence composition has been represented above). (Sequencing) adaptors are ligated to each of these fragmented sequences on both ends and the library is enriched by PCR, (during which additional nucleotides are added to comprise sequencing primers with the adaptors (labelled P5 and P7). To sequence the fragmented DNA, the PCR products are denatured and the 3' ends of each strand is affixed to a chip before sequencing from the 5' end. Read 1 begins immediately after the 3' end of the P5 primer and read 2 begins immediately after the 3' end of the P7 primer. The sequencer is set up to generate an average read length of 75 bases, but the DNA fragments have an average size of 300-350 bases.*

## 5.2. Bioinformatics analysis of next generation sequencing (NGS) results from the *S. enterica* ser. Typhimurium infection-host cell survival screen with the U937 gene trap library

The aim of the bioinformatics analysis of the NGS data from the STM infection host cell survival screen was to identify a list of trapped gene sequences that were present in significantly greater or lesser numbers in infected GT samples compared to uninfected GT controls. The tools used to arrive at this goal were required in a format that did not require extensive programming experience or use of a command line. Galaxy is such a toolkit; a web-based platform, which allows bioinformatics analysis of NGS data in an user-friendly manner (Afgan et al., 2018). The chief analyses would be to identify over- and under-represented mutated genes after the infection screens, to identify overlaps between the two screens, as well as any overlaps with the original screen of the HEK-293 GT library. These comparisons would identify potential 'pathogen-specific' and 'universal' host factors involved in response to infection processes. Furthermore, the gene lists would be used in gene ontology (GO) analysis to identify enriched GO terms relating to process or function that might shed light on the key biological mechanisms in the host response. Similarly, protein interaction analysis would further clarify important functional pathways in the host response.

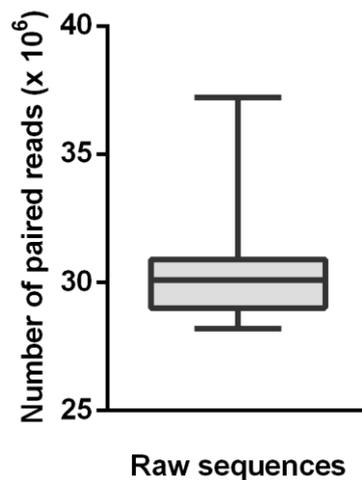
In our particular NGS application, we sought to identify significantly over- or under-represented mutated genes in one experimental group compared to another. This required several steps including matching reads to a reference (known as alignment or mapping), transcript assembly, counting the number of transcripts per gene per sample and statistically analysing the gene counts across groups. The most comparable form of NGS application to this is differential gene expression, as it also detects significant differences in representation between experimental samples to reach a similar end goal. The major difference between the two is the output: differentially expressed genes compared to

differentially represented trapped genes (our screen). However, the similarity allowed Galaxy tools normally applied to differential gene expression to be applied here.

### 5.2.1. Raw DNA sequence quality assessment

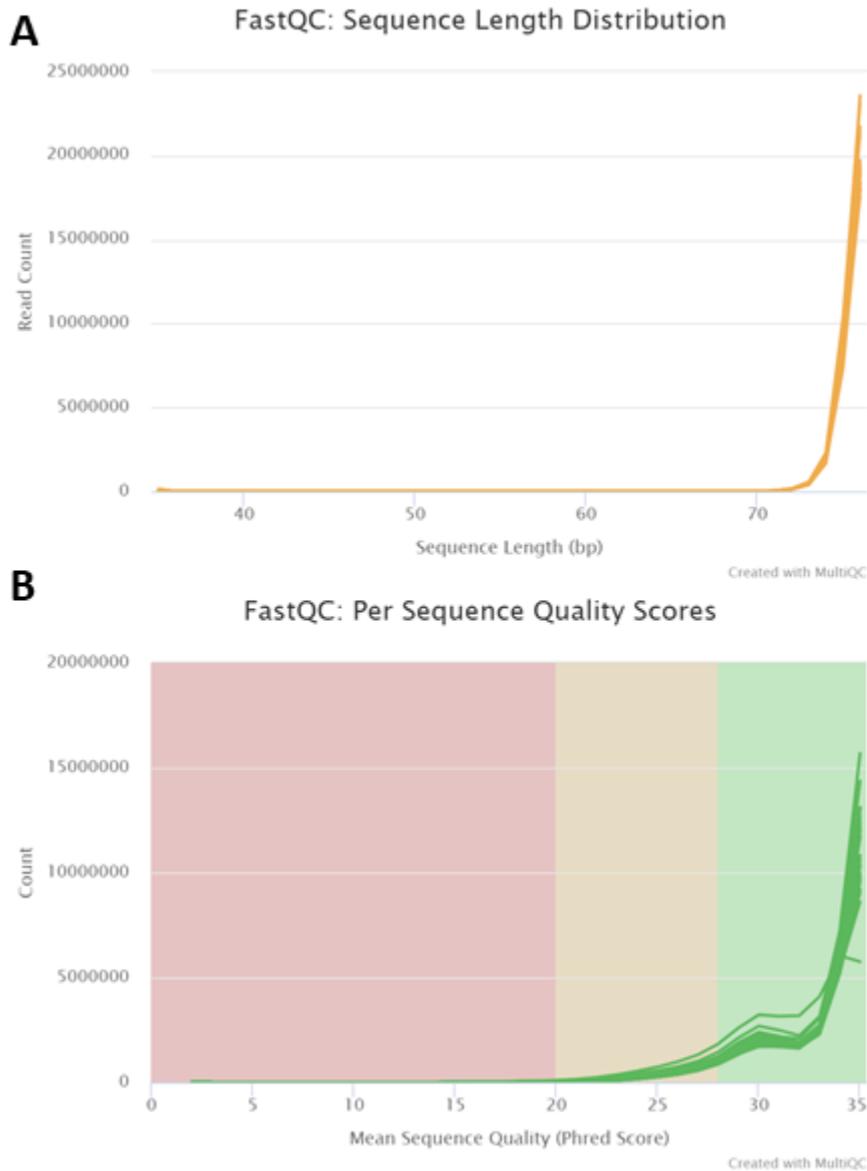
#### 5.2.1.1. Base call quality and read length

The raw NGS data files returned from Glasgow Polyomics contained an average of 30,000,000 paired reads (Figure 5.2) and approximately 2,250,000,000 bases per file.



**Figure 5.2 STM U937 screen raw NGS files have an average of 30 million paired reads, as requested during procurement.** Statistics taken from MultiQC file of raw NGS data provided by Glasgow Polyomics and plotted using GraphPad. Data comes from 8 experiments (A-D, OR1 and OR2). The whiskers extend from the lowest to highest values.

Post-sequencing quality assessment of raw sequence reads is an important step prior to read alignment so any technical problems can be identified and adjusted for (if possible) before alignment. The results of the quality checks used on the sequencing data were compared where appropriate to RNA-Seq data ideals. The length of the reads and accuracy of the bases produced by the sequencer (represented as a 'Phred score') were checked as these factors could affect sequence alignment. A Phred score is given to each base depending on how clear the signal was for that base in the sequencer. The vast majority of read lengths are more than 75 bp (Figure 5.3A), as requested at the time of the NGS service order. This is beneficial as longer reads allow more accurate alignment to the reference genome. Figure 5.3B shows that the majority of reads from the STM screen had an average Phred score of 30 or more (across the read) and an even greater proportion had a Phred score of 20 or more. A Phred score of 30 indicates the base has a 99.9% chance of being correct and a Phred score of 20 means the probability of the base call being correct is 99%. Thus, the majority of our sequencing was of good quality and could be used for downstream analysis. A small minority of sequences were of lower quality; later sections in the chapter detail trimming criteria which include low Phred scores.

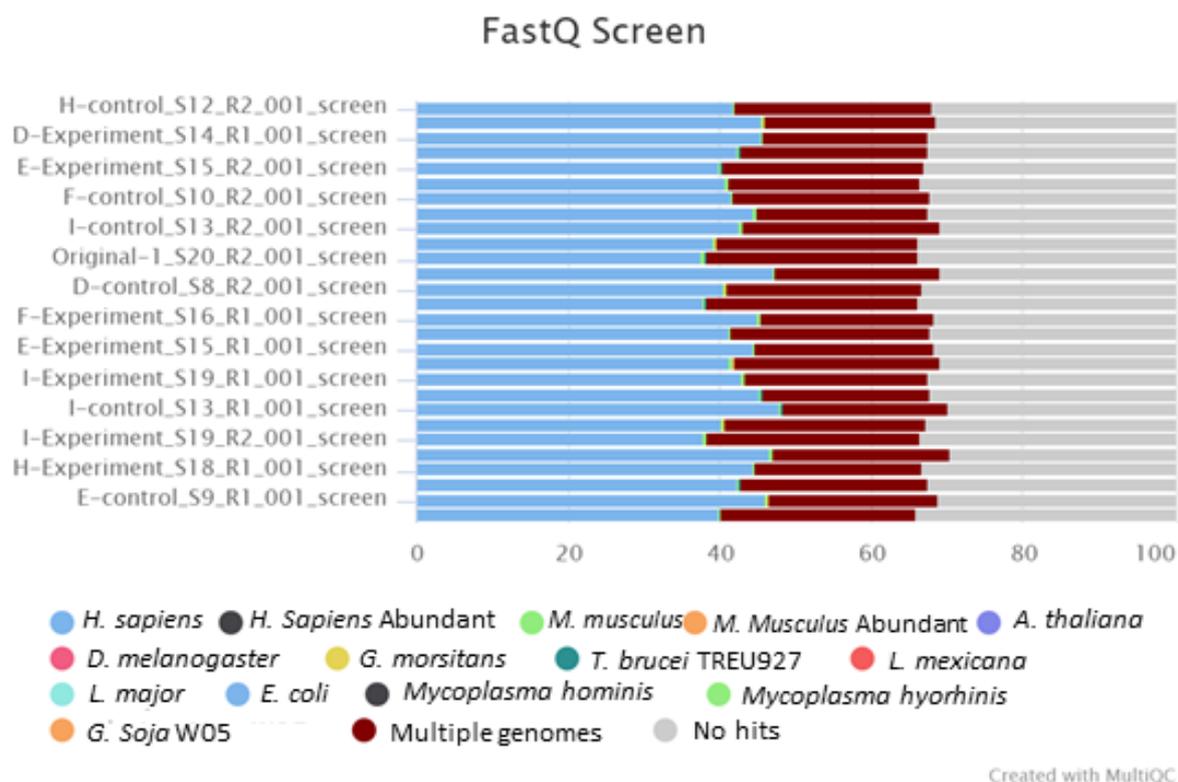


**Figure 5.3** The read length is as requested and the majority of reads in all samples have good quality (Phred) scores averaged across the read. Sequence length distribution, **A** and Per sequence quality scores of sequence reads, **B**. In **B**, each green line corresponds to an individual experimental sample, converted into a DNA library, and then taken through high-throughput sequencing. The plot provides an overview of sequence quality within and between samples. A Phred score is given to each base depending on how clear the signal was for that base in the sequencer. The red region of the graph indicates poorer quality scores, the orange region indicates good quality scores (a Phred score of 20 indicates the probability of a base call being correct is 99%) and the green region indicates high quality scores. The number of reads in a sample (y axis) is plotted against the average (mean) sequence quality represented as the mean Phred score (x axis). Each line is representative of a sample (Infected, Control or Original Lib., data comes from 8 experiments). The Phred scores were collated by FastQC (Andrews, 2019) and presented in graph form with MultiQC (Ewels et al., 2016) by Glasgow Polyomics.

#### 5.2.1.2. Quality check for species and vector contamination

The sequences derived from all experimental samples were checked for potential contamination by other species (e.g. from other samples at the sequencing facility) using a programme called FastQ screen (Figure 5.4). Sometimes contaminating sequences from other species can misalign to sequences in the human genome. Usually this is the consequence of sequence contamination with species of similar homology, but extremely short sequences require less species homology for a mismatch. The screen was initiated with an STM MOI of 2,400 and the bacteria had a chance to grow in nutrient-rich media for 2 days; therefore, bacterial genomic contamination after host cell lysis was likely. The RNA extraction included a DNase digestion step, which would have eliminated all, or at least the vast majority of any DNA present. The RACE PCR protocol included primers with specificity for messenger RNA (mRNA) poly-A tails and the GT vector exon, as well as an intermediate 500-fold sample dilution step. Thus, it is highly unlikely for such sources to be present in the final sequencing samples. Nevertheless, it is prudent to check the quality of the sequence data for signs of remaining sequence contamination, from bacteria, ribosomal RNA, mitochondrial RNA and sequencing primers. The potential for these sources of contamination will be additionally discussed and analysed in later sections of the chapter (5.2.1.3, 5.2.1.4 and in Appendix C).

Besides contamination detection, FastQ Screen can provide a rough estimate of the proportion of reads that can be aligned to the human genome. At later stages in the analysis, a more nuanced alignment programme is used as part of gene identification; thus, the proportion of mappable reads may be higher than indicated here. The experimental samples consistently contained approximately 40% of sequences that uniquely matched the human genome, as well as another 25-30% of sequences that matched multiple genomes of those tested (which likely includes human). The final 30-35% of sequences do not match any genome tested (no hits). As the *Salmonella* genome was not queried, the sequences without matches may result from STM contamination; however, evidence in Sections 5.2.1.3 and in Appendix C strongly suggests this is not the case. The majority, or perhaps all, of the sequences with no matches likely originate from the GT vector exon or GT vector exon-gene spanning sequences (as predicted in Figure 5.1).

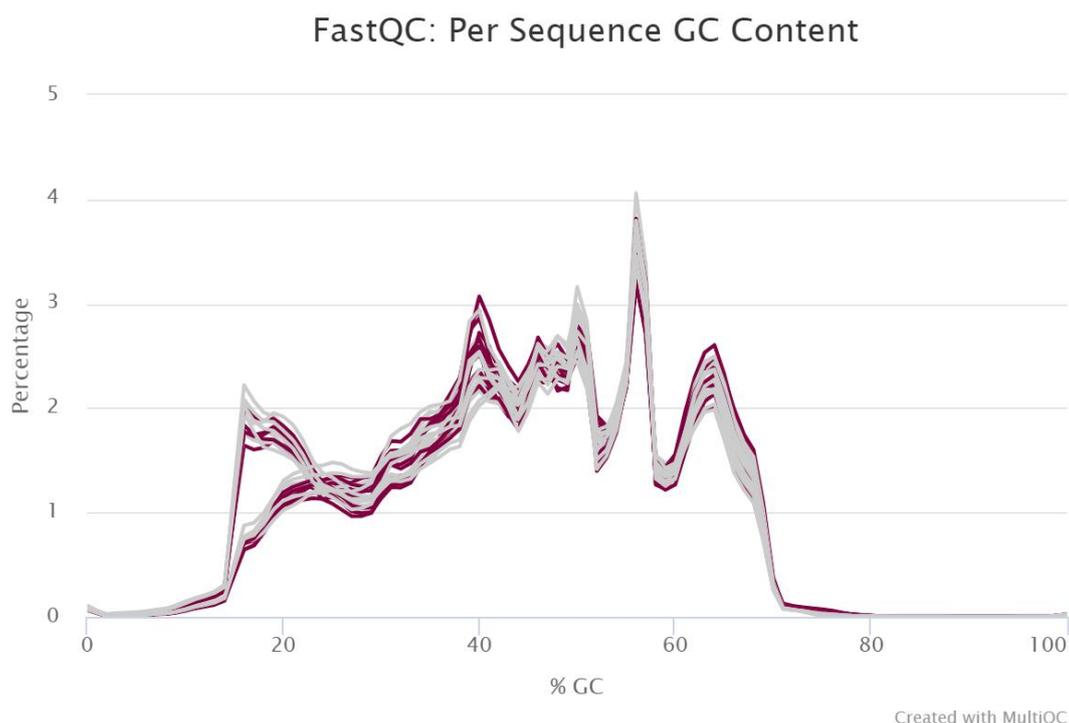


**Figure 5.4** At least 40% of sequence reads across all infected (experimental), uninfected (control), and original U937 GT library samples match the human genome. The light blue regions on the far left of the graph indicate reads matching the human genome, the red regions in the middle match multiple genomes (including human) and the grey regions on the right do not match any genome listed (and most likely reflect sequencing of the gene trap vector). The sample labels for each bar are on the left in grey, with each letter (D-I) at the start of the label corresponding to a specific experiment (8 experiments including Original-1 and 2). The percentages were calculated by the programme FastQ screen (Wingett and Andrews, 2018), presented in graph form with MultiQC by Glasgow Polyomics courtesy of Rachael Munro.

### 5.2.1.3. GC content quality check

Testing the GC content distribution of sequences in all experimental samples is another useful way of checking for potential contamination, as purely human mRNA sequences would have a normal distribution centred around 40.9% GC content (Piovesan et al., 2019). Figure 5.5 shows a very broad hump between 20-70% GC content with several sharp peaks at approximately 20, 40, 50, 55 and 70% GC content. The cDNA synthesis primer (**Table 2.4**) used to reverse transcribe transcripts containing poly-A tails from the extracted total RNA has a high AT content, so this is likely the cause for the peak at 20% GC content. The GT vector exon preceding trapped gene sequences prior to DNA fragmentation (Figure 5.1) has an average GC content of 56% (calculated with the GC content tool of Biologics International Corp, <https://www.biologicscorp.com/tools/GCContent/>). However, it is necessary to look more widely for potential sources of this GC content profile. For example, ribosomal, mitochondrial RNA and STM DNA could be potential sources of contamination, as described in Section 5.2.1.2. Ribosomal and mitochondrial RNA are GC rich, so could generate a peak at 70% GC content; however, as neither contain poly-A tails or GT vector exon sequences, it is highly unlikely either would be represented enough to affect the shape of the graph.

The STM genome has an average GC content of 52.3% (Fookes et al., 2011), so could produce a peak at roughly 50%. However, no difference in GC content distribution was seen between infected and uninfected (control and original) samples (Figure 5.5), strongly suggesting that STM sequences were not a source of contamination. An alternative reason for the sharp peaks in GC content may be low complexity of the DNA library. This is realistic as both of the GT libraries (1 and 3) had an estimated 1,000-10,000 mutational events prior to mixing (Section 4.3) and the sequencing was comparatively high depth at an average of 30,000,000 reads per sample. Further analysis of the source of these quality issues was performed during optimisation of sample quality processing described in Appendix C.

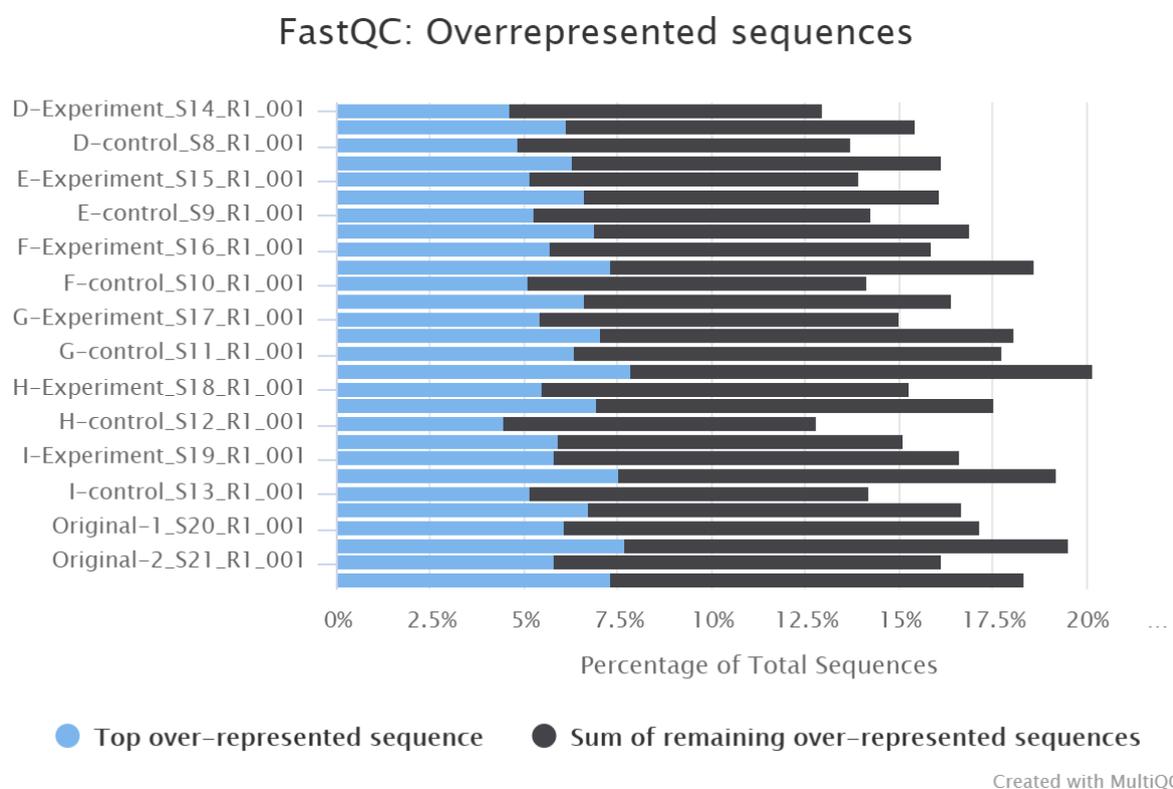


**Figure 5.5** The average per sequence GC content for the raw STM screen sequence samples is around 40% as expected, but the bell curve is interrupted by numerous sharp peaks. A small proportion of total reads per sample was tested for its GC content (%GC, x axis) and the percentage of reads per sample for a given GC content is plotted against the y axis. Each line represents a single sample (from a total of 8 experiments) each containing around 30 million reads. Infected samples are drawn in maroon, control and original library samples are grey. The sharp peaks are likely due to a combination of factors, for example the relatively high content of GT vector sequences.

#### 5.2.1.4. Per sample sequence overrepresentation

Testing for overrepresented sequences as a percentage of the total reads in a sample is another way to test quality, as it can indicate specific sources of contamination (e.g. with sequencing primers (Figure 5.1), bacterial, viral or fungal DNA, or vector sequences). An overrepresented sequence can comprise all or part of a read and is defined as a constituent of more than 0.1% of total reads per sample. Overrepresented sequences can also indicate high biological significance after selection, or lower than expected diversity in the library. Figure 5.6 shows that 12.5-20% of reads per raw experimental sample contain overrepresented sequences. The greatest overrepresented sequence for each sample is contained in 5-7.5% of total reads alone. An individual FastQC report for the Raw D-Control reverse file (as a representative) found that the measured overrepresented sequences all

contained portions of the cDNA synthesis primer (Table 2.2) or GT vector. Therefore, these sequences are likely to be highly overrepresented in the other raw sequence files as well.



**Figure 5.6 At least 12.5% of reads per raw sample contain overrepresented sequences.** An Overrepresented sequence is defined as a constituent of more than 0.1% of the total reads per sample. The blue bars represent the single most common overrepresented sequence per sample, which most likely reflect the cDNA synthesis primer, or GT vector sequences. Comparing all reads with each other takes significant computing power, so the first 100,000 are taken as representatives and compared with the remaining reads. A potential matching sequence must be over 20 bp to be considered and may have one mismatch or less. The sample labels for each bar are on the left in grey, with each letter (D-I) at the start of the label corresponding to a specific experiment (8 experiments including Original-1 and 2). The percentages of overrepresented sequences from raw sequencing files were calculated by FastQC and presented in graph form with MultiQC by Glasgow Polyomics.

## 5.2.2. Empirical comparison of quality-based processing techniques for use prior to analysis

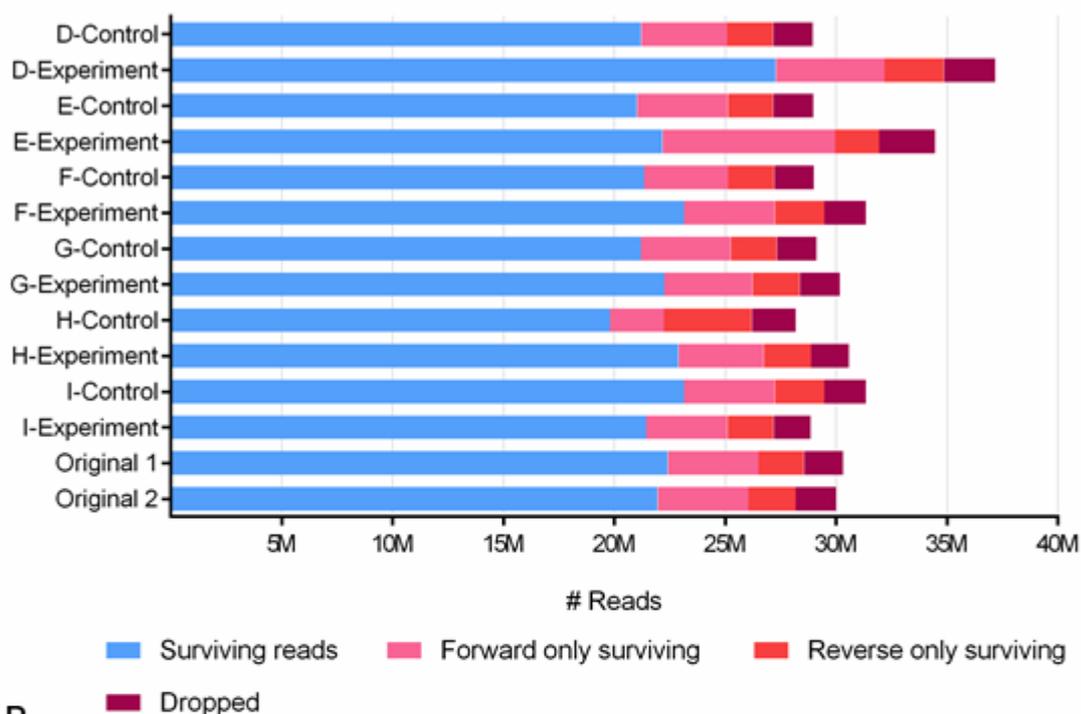
### 5.2.2.1. Trimming to an average Phred of 20 is appropriate

All bioinformatics analysis from hereon is performed on usegalaxy.org unless stated otherwise.

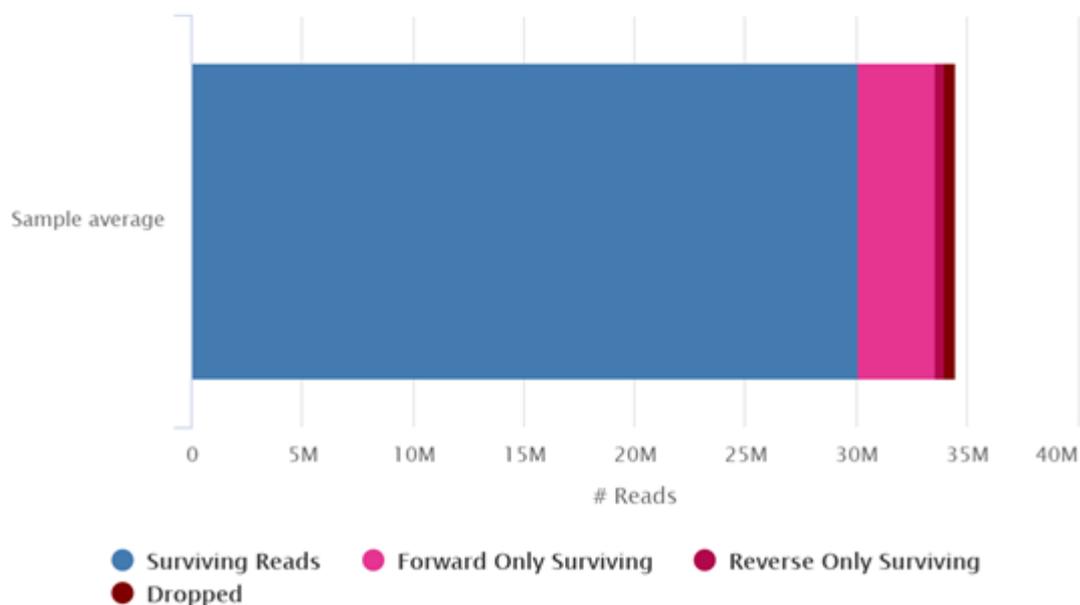
The following section describes the key considerations and processing steps used in the analysis. Trimming based on quality is a commonly used pre-processing adjustment to individual sequence reads - indeed some researchers suggest that some trimming is necessary for all RNA-Seq data to get reliably mapped reads (Fabbro et al., 2013). Percentage mappability (the percentage of reads in a

sample able to map to a reference genome) is often used as a reliability indicator. Mappability goes up with increased quality trimming, but the number of reads lost also increases; resulting in a need to find the best trade-off. Reads with low quality bases throughout may map inaccurately without trimming as well as taking extra time and computer memory. Reads with some low-quality regions may become mappable with careful trimming, provided the read is long enough afterwards. Aggressive levels of trimming (particularly to a minimum average Phred of 30 - 40) promote a bias in the number of reads aligning to a given transcript isoform or gene, where a family of genes is similar in sequence (Williams et al., 2016); because splice junctions are selected against. Splice junction spanning reads (or read pairs) help distinguish exon-sharing transcripts by specifying exon usage. The increased number of short sequences found in trimmed datasets also contributes to gene mapping bias, but this can be alleviated with minimum length filtering with the best results seen at 20 bases or more. On the other hand, some authors suggest that relatively high-quality sequencing data sets should not be trimmed, and that the soft-clipping feature of the read aligner should instead be solely relied upon (Liao and Shi, 2019). Ideally, the settings for trimming should be optimised for each dataset, however the time and computer space required to determine these is lengthy. Due to time and storage space constraints, a brief comparison of trimming at two different quality levels was performed, to identify a suitable threshold for the current dataset. The programme 'Trimmomatic' was chosen for trimming as it was a popular pre-processing tool for short, paired-end RNA-Seq reads (Bolger et al., 2014), was user friendly, and was available on [usegalaxy.org](http://usegalaxy.org). Sliding window, a trimming method used by Trimmomatic, was used to generate trimmed reads with a minimum average Phred of 30 (Figure 5.7A) or 20 (Figure 5.7B). Surviving reads with a length below 20 bases were excluded from the output. The minimum length filter was implemented based on the optimisations described by Williams et al., (2016). Trimmomatic also separates output reads into three categories: (paired) surviving reads, forward only, and reverse only surviving reads. Trimming to a minimum average Phred of 30 reduced surviving read pairs to just over 20 million in most samples, with one sample retaining 27 million read pairs. In contrast, trimming to a minimum average Phred of 20 left an average of 30 million surviving read pairs. Applications such as *de novo* genome assembly require high quality base calls such as a Phred score of 30 (where probability of the base being accurate is 99.9%); to produce a genome with high accuracy. For applications such as reference-based differential gene expression, where the outcome is a list of gene counts, highly accurate base calls are not essential, as long as the sequences can be aligned to correctly. Therefore, trimming to Q20 is often performed instead. Differential GT representation (the application to be implemented here) has a similar output and requirements to reference-based differential gene expression, thus, a minimum average Phred score of 20 is likely to be appropriate. Considering the lower Phred score assumed to be required for the application and the higher number of read pairs remaining after processing, trimming to a minimum average Phred of 20 was chosen as the threshold for this step. Simply trimming and length filtering will be known as pre-mapping procedure one.

A



B



Created with MultiQC

**Figure 5.7** Q20 trimmed and size-filtered sequences from STM U937 screen samples retain a greater majority (30 million surviving read pairs) of the input sequences than Q30 for genome alignment. Raw sequencing files trimmed to an average Phred of 30 (Q30) (A) or 20 (Q20) (B) per 4 bases using the sliding window operation before filtering for reads longer than 20 bases in Trimmomatic. Trimmomatic separates surviving reads into those in a complete pair (surviving reads), forward only reads and reverse only reads. The sample labels for each bar are on the left, with each letter (D-I) at the start of the label corresponding to a specific experiment (8 experiments including Original-1 and 2). (B) The bar shows the average number for each output group from all 14 STM U937 GT sequencing samples and was generated by the programme MultiQC.

### 5.2.2.2. GT vector removal from sequence reads leads to loss of mappable sequences

In theory removing (“cutting”) GT vector sequences from the reads would make some read pairs more mappable, therefore, this was assessed empirically (see Appendix C) through a combination of low quality base trimming and cutting off exogenous GT vector sequences. In practice this led to the loss of a substantial number of read pairs that could be partially (one read per pair) and uniquely mapped to the human genome (referred to as single-mapped). Extensive read loss can lead to changes in differential gene representation if either the loss affects samples differently, or some genes are lost completely from representation. As substantially more reads could be uniquely mapped (including single-mapped reads) by simply trimming – Section 5.2.2.1 – in contrast to cutting and trimming, this was decided on as the best approach.

### 5.2.3. Statistics of differentially represented human genes during STM infection and identification of the most significantly differentially represented genes

NGS reads from STM U937 screen samples were trimmed to an average Phred of 20 (see Section 5.2.2.1), mapped to the human genome (Appendix C, Section C.2) and bundled into transcripts before the number of transcripts per gene were counted (performed by the programme StringTie – see Appendix C, Section C.3). The gene counts were then in an appropriate format to determine differential GT fusion gene representation.

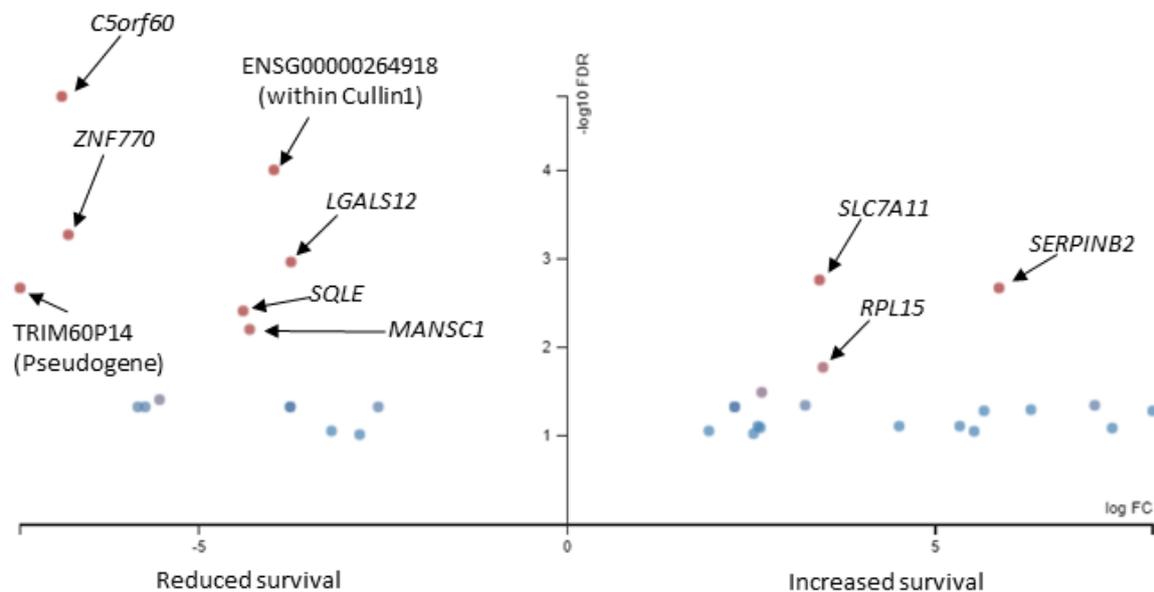
To identify trapped genes that are differentially represented in the infected U937 GT library population compared to uninfected controls; the gene counts output from StringTie must be statistically analysed. Many tools are available for the statistical analysis of differential gene expression datasets, of which limma (Ritchie et al., 2015) was the one selected for this study. limma was originally made for microarray data but has been adjusted to accommodate RNA-Seq data with the voom function. limma has also been suggested for use with low sample numbers (Pertea et al., 2016) and was built with that purpose in mind (Ritchie et al., 2015). The purpose-built online platform Degust (David Powell, 2015) was used to analyse the gene count data with limma as it is more accessible than Galaxy.

When multiple comparisons are made from data in one experiment, adjustments to the statistical analysis must be made to account for this. A popular method controls the family-wise error rate, or the probability of finding false positives based on the number of comparisons made. Controlling the family-wise error rate becomes too conservative when applied to large numbers of comparisons such as Omics data, resulting in the loss of many positive hits (false negatives). An alternative method was proposed by Benjamini and Hochberg (1995) that adjusted the false discovery rate (FDR); which takes into account only those comparisons initially deemed significant (discoveries) rather than all the comparisons made. Two levels of selection using the FDR threshold were desired for different purposes, a stricter level (threshold of 0.1), for identification of genes to be taken through independent genetic and pharmacological validation; and a second more liberal level (threshold of 0.3) for a broader analysis of potentially important processes in host infection response or survival.

#### 5.2.3.1. Selection stringency #1: FDR threshold of 0.1

A FDR threshold of 0.1 means that 10% of discoveries will be false positives (not true hits); such a threshold is useful for discovery screen analysis where the intention is to confirm any hits independently. This initial selection threshold indicated that 34 gene sequences were differentially

represented in infected vs control samples (Figure 5.8). Of these, trapped versions of *C5orf60*, *TRIM60P14*, *MANSC1*, *SQL*, *ENSG00000264918*, *ZNF770*, *LGALS12*, and 8 others were significantly less represented on average in the infected samples compared to the controls. *ENSG00000264918* is an Ensembl gene ID for a predicted microRNA transcript within an intron inside *Cullin1*. *SERPINB2*, *SLC7A11*, *RPL15* and 16 others were significantly over-represented in the infected samples compared to the controls. The molecular functions of these genes are briefly described in Table 5.1.



**Figure 5.8 Successful identification of trapped genes that show a strong effect on U937 cell survival or death (FDR threshold of 0.1) as measured by read depth differences between Infected and Control samples across the six paired replicates.** Volcano plot of significantly differentially represented genes identified from the STM U937 GT screen under the first level of selection. Mutant genes (dots) promoting survival in the infected population have positive values of log fold change (FC) plotted on the x axis and mutant genes reducing survival in the infected population have negative values of log FC.  $-\log_{10}$  FDR plotted on the y axis indicates how significant the fold change (Infected,  $n=6$  vs Control,  $n=6$ ) is for a given mutant gene (experiment performed 6 times). Less significant mutant genes are in blue and the more significant ones are in red. The comparison analysis was achieved using limma: voom in Degust with a FDR threshold (Benjamini Hochberg correction) of 0.1. The top 10 trapped genes (in red) are labelled.

**Table 5.1 Molecular function of differentially represented genes at a FDR threshold of 0.1.** *The Ensembl IDs are from the reference annotation provided to StringTie. The IDs are used to search for the gene sequence information on the Ensembl database <http://www.ensembl.org/index.html> (Cunningham et al., 2019), using Ensembl release 99 (January 2020). The molecular function information is obtained from the Ensembl and Uniprot databases <https://www.uniprot.org/> (The Uniprot Consortium, 2019). The log<sub>2</sub> fold change (FC) reports the average direction and magnitude of representation for infected samples (n=6) compared to the average of control samples (n=6). The data is from 6 experiments. The false discovery rate (FDR) column reports the p value adjusted for multiple testing.*

Ensembl gene ID	Symbol	Molecular function	Log <sub>2</sub> FC	FDR
ENSG00000204661	C5orf60	Uncharacterised protein	-6.86	1.47E-05
ENSG00000264918	miRNA (within CUL1)	microRNA, Cullin1 (CUL1) is part of an E3 ubiquitin-protein ligase complex	-3.98	9.93E-05
ENSG00000198146	ZNF770	May be involved in transcriptional regulation	-6.77	5.35E-04
ENSG00000133317	LGALS12	Recognises glycan groups/PAMPs	-3.75	1.08E-03
ENSG00000151012	SLC7A11	Cystine/glutamate antiporter, a.k.a. xCT	3.42	1.73E-3
ENSG00000197632	SERPINB2	a.k.a Plasminogen activator inhibitor-2 in extracellular form.	5.86	2.13E-3
ENSG00000250227	TRIM60P14	TRIM60 Pseudogene	-7.43	2.13E-3
ENSG00000104549	SQLE	Catalyses the stereospecific oxidation of squalene to (S)-2,3-epoxysqualene in cholesterol synthesis pathway	-4.40	3.87E-3
ENSG00000111261	MANSC1	Contains a MANSC domain	-4.31	6.25E-3
ENSG00000174748	RPL15	Ribosomal Protein L15	3.47	1.67E-2
ENSG00000103502	CDIPT	CDP-Diacylglycerol--Inositol 3-Phosphatidyltransferase	2.64	3.21E-2
ENSG00000137819	PAQR5	Progesterin and AdipoQ Receptor Family Member 5. Steroid binding	-5.53	3.88E-2
ENSG00000116459	ATP5F1	Mitochondrial ATP synthase subunit	3.23	4.50E-2
ENSG00000185774	KCNIP4	Potassium Voltage-Gated Channel Interacting Protein 4	7.16	4.50E-2
ENSG00000084710	EFR3B	EFR3 Homolog B. PI4K localisation to membrane	-2.56	4.69E-2
ENSG00000090487	SPG21	CD4-binding activity	-3.76	4.69E-2
ENSG00000091157	WDR7	WD repeat protein family. Regulator of Rab3 small G proteins	-5.73	4.69E-2
ENSG00000134243	SORT1	Sortilin 1. Trafficking of vesicles	-3.76	4.69E-2
ENSG00000259488	lncRNA, antisense to DUT	lncRNA, antisense to DUT	-5.83	4.69E-2
ENSG00000278259	MYO19	Myosin XIX. Actin-binding, mitochondrial localising	2.27	4.69E-2
ENSG00000278372	MYO19	Myosin XIX. Actin-binding, mitochondrial localising	2.27	4.69E-2
ENSG00000244247	LINC01995	lncRNA	6.29	5.04E-2

Ensembl gene ID	Symbol	Molecular function	Log2 FC	FDR
ENSG00000136379	ABHD17C	De-palmitoylating enzyme	5.66	5.19E-2
ENSG00000280515	SALRNA2	Senescence Associated Long Non-Coding RNA 2	7.95	5.19E-2
ENSG00000054803	CBLN4	Cerebellin 4 Precursor. Neurexin signalling pathway	5.32	7.73E-2
ENSG00000198804	MT-CO1	Mitochondrially Encoded Cytochrome C Oxidase I	4.50	7.73E-2
ENSG00000224778	CENPIP1	Unprocessed pseudogene (centromere protein I pseudogene 1)	2.58	7.73E-2
ENSG00000224228	lncRNA overlaps TNFSF18	lncRNA, TNFSF18 is a cytokine, it may control monocyte migration to sites of inflammation <sup>1</sup>	2.62	8.06E-2
ENSG00000173085	COQ2	Coenzyme Q2, Polyprenyltransferase	7.40	8.13E-2
ENSG00000235770	LINC00607	lncRNA	-3.20	8.75E-2
ENSG00000270276	HIST2H4B	Histone	1.92	8.75E-2
ENSG00000121964	GTDC1	Glycosyltransferase Like Domain Containing 1	5.52	8.83E-2
ENSG00000226791	LINC02611	long intergenic non-protein coding RNA 2611	2.53	9.39E-2
ENSG00000134028	ADAMDEC1	Disintegrin metalloproteinase family secreted from dendritic cells	-2.81	9.63E-2

In support of a successful outcome of this screen, a number of the genes listed in Table 5.1 have published evidence in support of a role in macrophage-pathogen interaction. The expression of *LGALS12*, a cell-surface galectin, which recognises carbohydrate species such as lipopolysaccharide (LPS), has been shown to change in the presence of PAMPs (Asiamah et al., 2019), and its deletion has been shown to promote macrophage differentiation in the M2 direction (Wan et al., 2016).

The serine protease serpin family B member 2 (*SERPINB2*) has established intracellular roles in senescence (Sossey-Alaoui et al., 2019; X. M. Zhang et al., 2019) and inhibition of cell movement (Schroder et al., 2019a; X. M. Zhang et al., 2019), in addition to its extracellular role in the clotting pathway (Schroder et al., 2019b). *SERPINB2* is upregulated in macrophages in response to LPS stimulation (Udofa et al., 2013) and it appears to promote macrophage polarisation towards an M2-like phenotype (Schroder et al., 2019a; Zhao et al., 2013). Furthermore, its knockout increases inflammatory gene expression, including *Nos2*, *Tnf* and subsequent Tnf secretion in macrophages (Schroder et al., 2010). This suggests that *SERPINB2* may function as a regulatory phenotype switch. *NOS2* (which produces RNS) and *TNF(α)* are important in controlling STM infection (Eriksson et al., 2003; Lahiri et al., 2008; Mastroeni et al., 2000; Pham et al., 2020; Stevanin et al., 2002); though perhaps the switch from an M2- to M1-macrophage phenotype and thus the loss of a replicative niche is most important (Saliba et al., 2016; Stapels et al., 2018).

The cholesterol synthesis pathway enzyme squalene epoxidase, encoded by the *SQLE* gene, has been linked with resistance to *Shigella* and *Salmonella* infection when deleted in embryonic stem cells (Yu et al., 2009).

The product of Cullin1 (*CUL1*) functions as part of a SKP1-CUL1-F-box protein (SCF) E3 ligase complex that ubiquitinates proteins, leading to their degradation (for review see Zhou et al., (2013)). The SCF

E3 ligase complex is an important regulator of cell cycle proteins (Burrows et al., 2012; Choudhury et al., 2017; Rechem et al., 2011; Rossi et al., 2013). SCF activity is prevented by bacterial effector Cycle inhibiting factor (Cif) (during enterohemorrhagic *E. coli* infection) through stabilisation of post-translational protein modification of CUL1 (Morikawa et al., 2010). This halts cell cycle progression at G1 phase. Cif-mediated inhibition of CUL1 activity suggests that the activity is deleterious to the bacteria in some way. Indeed, NLRP3 inflammasome activation is inhibited by CUL1 independent of SKP1 (Wan et al., 2019). In macrophages NLRP3 activates the inflammasome in response to signs of cellular stress or injury, resulting in IL-1 $\beta$  and IL-18 secretion and pyroptosis (Compan et al., 2012; Franchi et al., 2007; Groß et al., 2016; Hornung et al., 2008; Muñoz-Planillo et al., 2013; Perregaux and Gabel, 1994; Pétrilli et al., 2007; Shimada et al., 2012; Zhou et al., 2011).

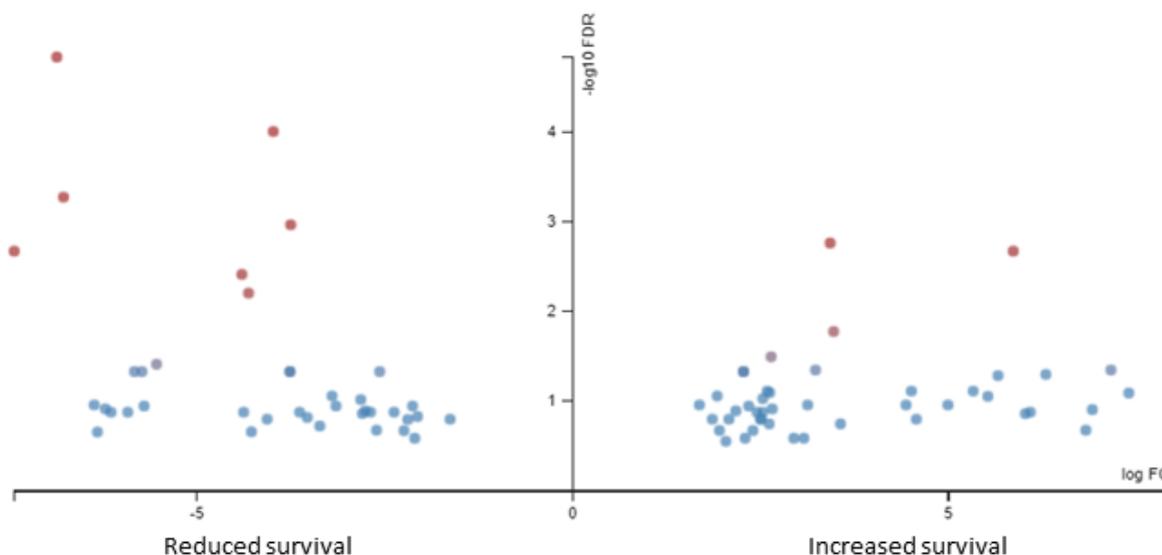
Solute carrier family 7 member 11 (*SLC7A11*) encodes half of a high affinity ionic amino acid channel, xCT, specific for cystine and glutamate. Cysteine transport (in this case in an oxidised form - cystine) into the cell is a rate limiting factor in glutathione synthesis (a potent reducing agent) see Lewerenz et al., (2013) for a review. *SLC7A11* was found upregulated in *Chlamydia trachomatis* infection (Dlugosz et al., 2014) and in *Mycobacterium tuberculosis* infected macrophages (Cai et al., 2016) to regulate the intracellular redox state. Moreover, *Slc7a11*<sup>-/-</sup> mice were more resistant to *M. tuberculosis* due to reduced glutathione levels attenuating intracellular ROS activity (Cai et al., 2016).

*Sort1*-KO mice have higher *Mycobacterium tuberculosis* burdens (Vázquez et al., 2016). Additionally, the human orthologue is upregulated in sepsis patients and methicillin-resistant *Staphylococcus aureus* infected children (Gaviria-Agudelo et al., 2014; Ma et al., 2015).

The product of WD repeat domain 7 (*WDR7*) regulates the formation of V-ATPase (Li et al., 2020), the protein complex that acidifies pathogen containing vacuoles (an important feature of host defence) (Kissing et al., 2018).

#### 5.2.3.2. Selection stringency #2: A reduced FDR threshold of 0.3

A FDR threshold of 0.3 means that 30% of discoveries will be false positives (not true hits); this looser statistical selection reports more truly differentially represented mutated genes as well as false positives. Indeed, a similar infection discovery screen found that a FDR threshold of 0.2 identified further hits that were confirmed in addition to those found at a threshold of 0.1 (Jeng et al., 2019); indicating that higher thresholds can identify additional 'true' discoveries. The mutated genes identified at this threshold will be put through further procedures such as gene ontology (GO) enrichment and induced network analysis, which look for patterns in the data. This pattern analysis may ameliorate the effect of false positives within the looser selection level by directing the focus of discovery onto genes fitting the broader pattern within the data. This selection level will additionally be used for comparison with a further screen in the same background (U937 cell line) but with a different pathogen as a selection agent (Section 5.3.4, 5.3.6) and with the previous GT screen in a HEK-293 background (Section 1.10). A further 50 genes were found differentially represented at this selection level in addition to genes identified at the stricter FDR threshold, all of which are displayed in the volcano plot in Figure 5.9. The molecular functions for the additional differentially represented genes are briefly described in Table 5.2.



**Figure 5.9** Sequence analysis resolved 84 mutant genes differentially represented in the STM infected U937 GT population compared to the control at a more liberal FDR threshold (0.3).

Mutant versions of genes sequences (dots) promoting survival in the STM infected population have positive values of log fold change (FC) plotted on the x axis and mutant genes reducing survival in the infected population have negative values of log FC.  $-\log_{10}$  FDR plotted on the y axis indicates how significant the fold change (Infected,  $n=6$  vs Control,  $n=6$ ) is for a given mutant gene. Less significant mutant genes are in blue and the more significant ones are in red. The comparison analysis was done with Limma: voom in Degust with a FDR threshold (Benjamini Hochberg correction) of 0.3.

**Table 5.2** Molecular function of additional differentially represented genes at a FDR threshold of 0.3 in STM infected U937 GT cells. At a FDR threshold of 0.3 these genes in addition to those in Table 5.1 were identified as significantly differentially represented in the STM infected U937 GT population. The Ensembl IDs are from the reference annotation provided to StringTie. The IDs are used to search for the gene sequence information on the Ensembl database <http://www.ensembl.org/index.html> (Cunningham et al., 2019), using Ensembl release 99 (January 2020). The molecular function information is obtained from the Ensembl and Uniprot databases <https://www.uniprot.org/> (The Uniprot Consortium, 2019). The  $\log_2$  fold change (FC) reports the average direction and magnitude of representation for infected samples ( $n=6$ ) compared to the average of control samples ( $n=6$ ). The false discovery rate (FDR) column reports the  $p$  value adjusted for multiple testing. <sup>1</sup> TNFSF18 function predicted due to similarity to mouse *tnfsf18* (Liao et al., 2014).

Ensembl gene ID	Symbol	Molecular function	Log <sub>2</sub> FC	FDR
ENSG00000125538	IL1B	Interleukin 1B	4.44	1.10E-1
ENSG00000147872	PLIN2	Perilipin 2. Coats lipid droplets in cells	4.99	1.10E-1
ENSG00000198695	MT-ND6	Mitochondrially Encoded NADH: Ubiquinone Oxidoreductase Core Subunit 6	3.13	1.10E-1
ENSG00000223561	LOC646588	Uncharacterised Protein-coding gene	1.69	1.10E-1
ENSG00000229557	LINC00379	long intergenic non-protein coding RNA 379	-6.37	1.10E-1

Ensembl gene ID	Symbol	Molecular function	Log <sub>2</sub> FC	FDR
ENSG00000077514	POLD3	DNA Polymerase Delta 3, Accessory Subunit	-2.13	1.11E-1
ENSG00000140853	NLRC5	NLR Family CARD Domain Containing 5. Inflammasome pathway	-3.15	1.13E-1
ENSG00000163637	PRICKLE2	Prickle Planar Cell Polarity Protein 2	2.34	1.13E-1
ENSG00000166862	CACNG2	Calcium channel.	-5.70	1.13E-1
ENSG00000006625	GGCT	Gamma-Glutamylcyclotransferase. Glutathione metabolism	2.66	1.22E-1
ENSG00000162782	TDRD5	Tudor Domain Containing 5	-6.22	1.22E-1
ENSG00000155761	SPAG17	Sperm Associated Antigen 17. Cilia function	6.92	1.24E-1
ENSG00000119917	IFIT3	Interferon Induced Protein With Tetratricopeptide Repeats 3	-2.76	1.28E-1
ENSG00000272807	lncRNA, antisense to KANSL1L	KANSL1L is a subunit of histone acetyltransferase	2.17	1.28E-1
ENSG00000004468	CD38	Transmembrane glycoprotein that synthesizes and hydrolyses cyclic adenosine 5'-diphosphate-ribose	2.53	1.33E-1
ENSG00000176853	FAM91A1	Family With Sequence Similarity 91 Member A1	-2.69	1.33E-1
ENSG00000187688	TRPV2	Transient Receptor Potential Cation Channel Subfamily V Member 2. High temperature-detecting channel	-2.38	1.33E-1
ENSG00000223587	LINC01986	long intergenic non-protein coding RNA 1986	6.09	1.33E-1
ENSG00000270948	MTDHP1	metadherin pseudogene 1	-5.92	1.33E-1
ENSG00000277443	MARCKS	Myristoylated Alanine Rich Protein Kinase C Substrate. Cell motility and phagocytosis	2.45	1.33E-1
ENSG00000186417	GLDN	Gliomedin. Nodes of Ranvier formation	-3.63	1.33E-1
ENSG00000238057	ZEB2-AS1	Antisense to Zinc Finger E-Box Binding Homeobox 2	-6.14	1.33E-1
ENSG00000260586	lncRNA, intronic to THSD4	THSD4 promotes Fibrillin-1 extracellular matrix assembly, reduces TGF- $\beta$ signalling	-4.38	1.33E-1
ENSG00000274461	lncRNA	lncRNA	-2.80	1.38E-1
ENSG00000254275	LINC00824	long intergenic non-protein coding RNA 824	6.02	1.38E-1
ENSG00000117682	DHDDS	Dehydrololichyl Diphosphate Synthase Subunit. Dolichol phosphate synthesis which allows glycosylation of proteins	-2.06	1.48E-1
ENSG00000253439	CDC42P5	cell division cycle 42 pseudogene 5	-3.53	1.52E-1
ENSG00000102760	RGCC	Regulator Of Cell Cycle	4.57	1.59E-1
ENSG00000135241	PNPLA8	Patatin Like Phospholipase Domain Containing 8	2.08	1.59E-1

Ensembl gene ID	Symbol	Molecular function	Log <sub>2</sub> FC	FDR
ENSG00000149781	FERMT3	Fermitin Family Member 3. Cell adhesion	2.50	1.59E-1
ENSG00000175147	TMEM51-AS1	Antisense transcript of TMEM51 (Transmembrane Protein 51)	-1.63	1.59E-1
ENSG00000198786	MT-ND5	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 5	2.51	1.59E-1
ENSG00000226012	lncRNA (within KCNJ15)	KCNJ15 is an ATP-sensitive inward rectifier potassium channel	-4.07	1.59E-1
ENSG00000231028	LINC00271, overlapping AHI1	lncRNA, AHI1 is involved in vesicle trafficking and genesis of cillum	1.86	1.59E-1
ENSG00000129515	SNX6	Sorting Nexin 6. Intracellular trafficking	-2.19	1.60E-1
ENSG00000163704	PRRT3	Proline Rich Transmembrane Protein 3	2.62	1.79E-1
ENSG00000173068	BNC2	Basonuclin 2	3.56	1.79E-1
ENSG00000250999	NA	lncRNA	-3.37	1.89E-1
ENSG00000132669	RIN2	Ras And Rab Interactor 2. Endocytosis	-2.61	2.11E-1
ENSG00000236819	LINC01563	lncRNA	6.83	2.11E-1
ENSG00000185634	SHC4	SHC Adaptor Protein 4	1.95	2.12E-1
ENSG00000228058	LINC01736	lncRNA	2.40	2.13E-1
ENSG00000076043	REXO2	RNA Exonuclease 2. DNA repair pathway	-2.24497	2.13E-1
ENSG00000231740	NA	lncRNA	-4.28	2.19E-1
ENSG00000133808	MICALCL	MICAL C-Terminal Like. MAPK pathway, vesicle trafficking	-6.33	2.21E-1
ENSG00000100767	PAPLN	Papilin, Proteoglycan Like Sulfated Glycoprotein. Peptidase activity	-2.10	2.59E-1
ENSG00000117242	PINK1-antisense	PINK1 is a Serine/threonine-protein kinase, it shields against mitochondrial dysfunction throughout cellular stress	2.94	2.59E-1
ENSG00000131446	MGAT1	Begins the formation of complex N-linked carbohydrates	2.29	2.59E-1
ENSG00000198431	TXNRD1	thioredoxin reductase activity, effect is isoform dependent, e.g. mediates cell death, promotes actin and tubulin polymerization	3.08	2.59E-1
ENSG00000100292	HMOX1	cleaves the heme ring at the alpha methene bridge to form biliverdin, cytoprotective in excess free heme conditions	2.04	2.80E-1

Within these differentially represented gene sequences are a few with previously described involvement in infection or immune responses, indicating potential genuine hits outwith the initial FDR threshold. For example, the cytokine IL-1 $\beta$  (IL1B) is a well-known product of macrophage inflammasome activation in response to *Salmonella* infection (see Bierschenk et al., (2017) for review) and the trapped version of the *IL1B* gene was identified as significantly overrepresented at the looser FDR threshold. In addition to secreting IL-1 $\beta$ , macrophages respond to exogenously applied IL-1 $\beta$  with Interferon regulatory factor 3 (IRF3) activation, production of IRF3 response genes and multiple positive regulators of programmed cell death (Aarreberg et al., 2019). The action of IL-1 $\beta$  in priming the host cell for programmed cell death fits with an improved survival advantage for cells with reduced functional *IL1B* expression in an environment stimulating IL-1 $\beta$  production.

Trapped Interferon-induced protein with tetratricopeptide repeats (IFIT) 3 (*IFIT3*) was identified as significantly underrepresented within the looser FDR threshold. *IFIT3* is expressed in response to LPS stimulation (Billing et al., 2011; Øvstebø et al., 2008) or IRF3 activation (Ashley et al., 2019; Grandvaux et al., 2002) indicating an indirect connection to IL-1 $\beta$  activity. *IFIT3* inhibits apoptosis promoting IFIT 2 (*IFIT2*) activity – another protein expressed in response to IRF3 activation (Guinn and Petro, 2019; Nakaya et al., 2001). The anti-apoptotic activity of *IFIT3* fits with the data presented here suggesting a reduced survival advantage with reduced functional *IFIT3* expression. Additionally, *ifit3* expression is directly associated with *Borrelia burgdorferi* and *Helicobacter pylori* infection in mouse and rat, respectively (Petzke et al., 2016; Yang et al., 2020).

Further differentially represented trapped genes identified within the looser FDR threshold for which functional versions have associations with infection response are sorting nexin-6 (*SNX6*) and Myristoylated alanine-rich C-kinase substrate (*MARCKS*). The *SNX6* protein sorts membrane proteins from early endosomes to the trans Golgi network – see Bonifacino and Hurley, (2008) for review. It is thought to be an important part of the delivery of acid hydrolases to the lysosomes, allowing for effective degradation of the contents (Seaman, 2018; Simonetti et al., 2019; Yong et al., 2020). Indeed, *snx6* knockdown in *Drosophila melanogaster* cells prevents effective autophagy (Maruzs et al., 2015). This theory fits with the underrepresentation of the trapped gene (which suggests that having one mutated copy reduces survival during STM infection). Besides this, the *SNX6* protein is hijacked by pathogen proteins during infection with *Listeria monocytogenes* (David et al., 2018) or *Chlamydia trachomatis* (Paul et al., 2017; Sun et al., 2017) and its abundance is reduced during STM infection (Shi et al., 2009); suggesting that it may have an important function in the immune response.

*MARCKS* regulates inflammatory cytokines in response to *Streptococcus* infection (Q. Zhang et al., 2019) or LPS stimulation (Lee et al., 2015) and its phosphorylation is associated with *B. thailandensis* survival within host cells (Micheva-Viteva et al., 2017); though its role in STM infection is unknown.

A total of 18 predicted long noncoding RNAs (lncRNAs) and 3 antisense transcripts have been reported as differentially represented in the STM U937 screen (Table 5.1 and Table 5.2). Many lncRNAs and antisense transcripts have been characterised with roles in gene expression regulation, see Kopp and Mendell, (2018) for review. However, characterising the function of predicted lncRNAs and antisense transcripts identified here is out the intended scope of this work, as they are difficult to query and cannot be pharmacologically targeted through traditional methods. The functional and biological relevance of many predicted lncRNAs is ambiguous, therefore, where reported uncharacterised lncRNAs and antisense transcripts are present within the proximity of a protein coding gene, it is the latter that is interrogated further (as in Section 5.2.4).

#### 5.2.4. Gene ontology analysis of STM screen-derived genes

A gene set enrichment analysis was performed to aid identification of biological processes that are important in the infection process. Gene set enrichment analysis (a type of gene ontology, GO) determines if associated gene ontology process or component terms are statistically enriched in the sample dataset compared to a reference dataset. GO process terms are labels for biological outcomes, which are as specific as “positive regulation of intrinsic apoptotic signalling pathway in response to hydrogen peroxide” or as broad as “response to stimulus” (Thomas, 2017). GO component terms refer to a cellular location in relation to cellular or macromolecular structures association with a given gene product.

Characterised genes have associated gene ontology terms, which are held in a central gene ontology database <http://geneontology.org/> (The Gene Ontology Consortium, 2019; The Gene Ontology Consortium et al., 2000). The associations are updated frequently within the database in response to new scientific literature. Many tools offering gene set enrichment analysis are available, but the methods, accessibility, output, and tool-to-database update regularity vary. GOrilla <http://cbl-gorilla.cs.technion.ac.il/> (Eden et al., 2009, 2007) was chosen for use, allowed a gene list to be compared to a provided reference list, was available online, provided the output in an informative manner and was updated monthly.

The protein coding genes within whole genome can typically be used as a reference list. However, the total number of open reading frames (including non-coding RNAs and transcript readthroughs) represented in the gene trap library was 15,449 (calculated as all Ensembl ID entries detected with a read count of one or more), approximately two-thirds of the human genome. A comparison of gene hits to the list of these represented open reading frames (rather than to the whole genome) makes more scientific sense, as well as reducing unnecessary multiple testing, and was applied in the analysis.

Input lists of over- and underrepresented trapped genes were curated consisting of distinct genes from the looser threshold (Table 5.1 and Table 5.2) without lncRNAs, pseudogenes or entries which lacked an official gene symbol, as none of these are recognised by GOrilla. Additionally, antisense transcripts for three genes were differentially represented at the more liberal threshold (Table 5.2), and these were replaced with official gene symbols for the complementary coding genes. Similarly, some lncRNAs present in the differentially represented list were located within or overlapped protein coding genes (Table 5.2), in such cases the former was replaced with the latter. Where the affected lncRNA overlaps a protein coding gene, the latter may also have been affected (which is of more interest than the lncRNA itself).

The GO processes ‘response to oxidative stress’, ‘smooth muscle adaptation’, ‘electron transport chain’ as well as the GO components ‘organelle inner membrane’ and ‘astrocyte projection’ were enriched at an FDR threshold of 0.1 in the overrepresented trapped gene list (**Table 5.3**). However, no GO processes or components were enriched at a threshold of 0.1 FDR for the underrepresented trapped gene list, even though the list was of a comparable size to the overrepresented one. The low number of genes associated with, and the inappropriate biological context of, the GO processes ‘smooth muscle adaptation’ and ‘astrocyte projection’ indicate that they are unlikely to be biologically relevant. Where the majority of associated input genes are shared between identified GO terms (as is the case for ‘response to oxidative stress’, ‘electron transport chain’ and ‘organelle inner membrane’), inferences about the specific processes should be taken with care, as one or the other may not be relevant. In this case, the reduced availability of proteins of input genes highlighted below may be disrupting the electron transport chain as well as reducing the oxidative

stress response (see analysis below for further detail). The number of input trapped genes associated with 'response to oxidative stress' and 'electron transport chain' respectively, with low FDR q values (a multiple testing adjusted p-value) are good indications that one or more of these processes is genuinely important in STM-infected U937 cells.

**Table 5.3 Gene Ontology enrichment analysis of a lower stringency selection of 35 overrepresented mutant genes identified enriched GO terms including the GO process ‘response to oxidative stress’.** *GO process and component terms enriched in the queried list of 35 protein coding genes compared to the GT library background (15,466 distinct open reading frames including those of uncertain function with a gene count  $\geq 1$  in at least one sample). Of 9019 recognised genes (queried and background open reading frames), 8,143 were associated with a GO term. The q-value is the Benjamini and Hochberg (1995) corrected version of the p-value after testing for 12974 GO terms. Trapped protein coding genes overrepresented (with a FDR less than 0.3) in STM-infected samples were entered into the GOrilla (Gene Ontology enRiChment anaLysis and visualizAtion tool) website <http://cbl-gorilla.cs.technion.ac.il/> with GT library background open reading frames as Ensembl IDs to ensure GOrilla recognised a greater proportion of the protein coding genes. Ensembl IDs for overrepresented non-coding RNAs within or overlapping protein coding genes were replaced with the associated protein coding gene Ensembl ID before the GOrilla search. Process, function and component ontologies were searched using default settings, though no function ontologies were found enriched. Only GO terms which had an FDR q-value of 0.1 or less were retained. This is derived from 6 experiments.*

GO Term	Description	FDR q-value	Enrichment	Genes
GO:0006979 (Process)	response to oxidative stress	4.00E-3	11.4	<i>PINK1, CD38, SLC7A11, MT-ND5, HMOX1, MT-CO1, TXNRD1, MT-ND6</i>
GO:0014805 (Process)	smooth muscle adaptation	9.10E-2	263	<i>HMOX1, IL1B</i>
GO:0022900 (Process)	electron transport chain	8.20E-2	14.8	<i>PINK1, MT-ND5, MT-CO1, TXNRD1, MT-ND6</i>
GO:0019866 (Component)	organelle inner membrane	7.70E-2	7.33	<i>PINK1, ATP5F1, MT-ND5, COQ2, MT-CO1, MT-ND6</i>
GO:0097449 (Component)	astrocyte projection	6.00E-2	105	<i>PINK1, SLC7A11</i>

#### 5.2.4.1. GO Process - Electron transport chain

The electron transport chain (ETC) couples the TCA cycle with oxidative phosphorylation, allowing the efficient transfer of energy from fuel sources such as glucose and fatty acids to ATP. In resting macrophages Nicotinamide adenine dinucleotide + hydrogen (NADH)-ubiquinone oxidoreductase (complex I) transfers electrons from NADH to ubiquinone and succinate dehydrogenase (complex-II) transfers electrons from reduced flavin adenine dinucleotide (FADH<sub>2</sub>). Inoculation of BMDMs with live *E. coli* or STM or M1 polarisation of macrophages dysregulates the ETC (Garaude et al., 2016; Jha et al., 2015), causing excess production of mROS (Chouchani et al., 2014; Guarás et al., 2016; Jin et al., 2014; Lapuente-Brun et al., 2013; Lopez-Fabuel et al., 2016). Reducing the number of active complex I also increases mROS (Arena et al., 2018; Jin et al., 2014; Lopez-Fabuel et al., 2016; Pignataro et al., 2017) and reducing the expression of Cytochrome c Oxidase (complex IV) would likely have the same effect (Chouchani et al., 2014; Guarás et al., 2016; Lapuente-Brun et al., 2013).

Excess mROS reduces intracellular bacterial burden (Roca and Ramakrishnan, 2013; West et al., 2011), but the mechanisms for this remain enigmatic. mROS may be directly bactericidal (Heijden et al., 2015; West et al., 2011), induce proinflammatory gene expression (Bulua et al., 2011; Jin et al., 2014), or promote inflammasome-assisted host cell death (Groß et al., 2016; Shimada et al., 2012);

Zhou et al., 2011). The latter mechanism of mROS-mediated reduction of bacterial burden does not fit with the results presented here, so will be discounted in this case.

Trapped versions of genes encoding subunits of complex I – Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Core Subunit 5 (*MT-ND5*), Subunit 6 (*MT-ND6*) – as well as a gene encoding a core subunit of complex IV – Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1*) were overrepresented in the STM U937 screen. The trapped genes Phosphatase And Tensin Homolog (PTEN) Induced Kinase 1 (*PINK1*), and *TXNRD1* were also found overrepresented in the U937 GT population after STM infection. The gene trap cassette inserts at sufficiently low frequency that it can be assumed that only one mutation will occur in a given mutated cell. If the mutated gene is present on an autosome (Panning, 2008), mutating one copy of two will likely result in a reduction of gene and protein expression rather than a complete cessation of function. It is possible that expression from the remaining functional copy is upregulated to compensate, but this is unlikely to produce any signal in a gene trap screen (Zhipeng Ma et al., 2019).

Reduced protein expression from any of *MT-ND5*, *MT-ND6*, *MT-CO1*, *PINK1* and *TXNRD1* likely either leads to mROS production (Morais et al., 2014; Narendra et al., 2010) or reduces the redox buffer (thus increasing general cellular oxidation)(Nalvarte et al., 2004). *MT-ND5*, *MT-ND6*, *MT-CO1*, *PINK1* and *TXNRD1* all are associated with the GO process ‘electron transport chain’ (Table 5.3).

#### 5.2.4.2. GO Process - Response to oxidative stress

The gene ontology enrichment analysis in Table 5.3 indicates that the host oxidative stress response is a focus of mutations that promote host cell survival during STM infection. In response to oxidative stress, human cells activate a series of stress responses to return the intracellular environment back to a homeostatic balance. *MT-ND6*, *MT-CO1* and CD38 molecule (*CD38*, aka adenosine diphosphate (ADP)-ribosyl cyclase/cyclic ADP-ribose hydrolase 1) transcription are repressed (Arena et al., 2018; Crawford et al., 1997; Ozawa et al., 2007; Yu et al., 2020), Cystine transporter (e.g. xCT, of which SLC7A11 is a subunit) activity is enhanced (Bannai et al., 1989; Dun et al., 2006; Li et al., 1999) and *HMOX1* gene expression is upregulated (Alam et al., 1999, 1994; Choi and Alam, 1996). Furthermore, *PINK1* protein levels are stabilised in response to pharmacological inhibition of ETC complex I (Xuejing Wang et al., 2014). Under conditions of intracellular oxidative stress, *TXNRD1* reduces oxidised forms of ubiquinone (Xia et al., 2003), thioredoxin as well as other antioxidants, thus replenishing these important redox buffers, see Lu and Holmgren, (2014) for review.

The same genes (*MT-ND5*, *MT-ND6*, *MT-CO1*, *PINK1*, *TXNRD1*) as those associated with the GO process ‘electron transport chain’ are also associated with the GO process ‘response to oxidative stress’, as well as three others *CD38*, *SLC7A11* and *HMOX1* (Table 5.3). The effect of reducing the number of ETC complex I (*MT-ND5*, *MT-ND6*) and complex IV subunits (*MT-CO1*) on mROS, and the effect of increased mROS on bacterial burden, are both described above in Section 5.2.4.1.

Heme, the substrate of *HMOX1*, can be a source of oxidative stress, such that excessive exposure of heme to *HMOX1* deficient cells causes cell death (Yachie et al., 1999). Nevertheless, *Hmx* knockdown in Raw cell line mouse macrophages have lower STM burden, potentially due to the resulting increased oxidative stress (Mitterstiller et al., 2016; Silva-Gomes et al., 2013). *CD38* controls intracellular calcium signalling through cleavage of NAD<sup>+</sup> to produce the signalling molecule cyclic ADP ribose, see Lee (2006) for review. *CD38*<sup>-/-</sup> promotes resistance to oxidative stress (e.g. H<sub>2</sub>O<sub>2</sub> injury), potentially through the increase of the redox buffer, NAD<sup>+</sup> (Ge et al., 2010; Xiao et al., 2018). As SLC7A11 is part of a cystine-glutamate transporter, it regulates the production of glutathione, a cellular antioxidant. Glutathione is used as a further redox buffering system to control ROS mediated damage.

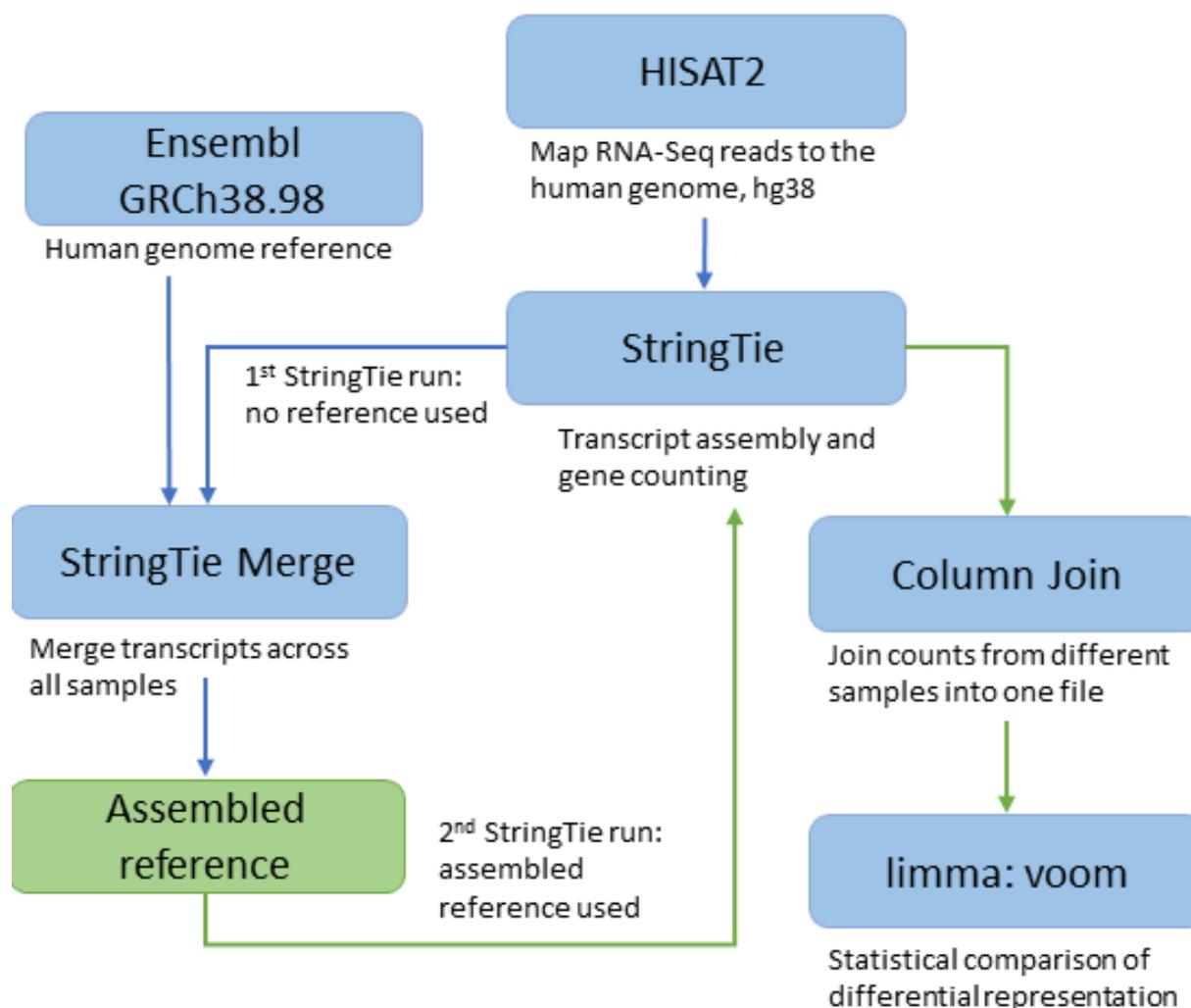
The reduced gene expression (and activity) of these enzymes and antiporter as a result of trapping a single allele of each is an increased oxidising intracellular environment, which will reduce bacterial burden (Roca and Ramakrishnan, 2013; West et al., 2011).

#### 5.2.5. Intergenic trapped sequences with altered representation after infection were identified using *de novo* transcript assembly

Within the reference-based transcript assembly used in the original analysis, the definition of the extent of a gene is strictly tied to the official reference transcript information provided; therefore, incomplete transcripts are discarded as aberrant. As trapped gene transcripts are inherently 'incomplete', this transcript assembly method requires known transcripts to be assembled from reads generated from one or, more likely, multiple 'incomplete' real transcripts. *De novo* transcriptome assembly can partially avoid this pitfall by instead attempting to identify all potential transcripts found within the sequencing files. A flow chart of the processes required for this analysis is presented in Figure 5.10.

*De novo* transcriptome assembly has the potential to incorporate sequences generated from non-canonical GT cassette insertion that would not normally be counted. Random GT mutation usually results in GT cassette insertion into an intron of a gene and a fusion transcript is created starting with the GT cassette spliced to the following exon. Sometimes the GT cassette is inserted into an intergenic sequence, but the cell still survives antibiotic selection indicating that a successful splice event must have occurred. Such transcripts are mechanistically harder to explain but may still act in a mutagenic fashion. In some cases, the GT cassette may aberrantly splice to a sequence similar to an exon - previously described in (Gow et al., 2013) - but in so doing, disrupts the function of a gene's regulatory sequence such as an enhancer. An assumption was made that an intergenic sequence present at statistically greater or lower frequency in infected samples compared to control samples would have functional relevance. *De novo* transcript assembly is documented for use in building previously unidentified alternative transcript isoforms in less studied contexts in model organisms as well as species without reference genomes (Hölzer and Marz, 2019; Pertea et al., 2016, 2015); therefore, some heuristic optimisation was required to fit it to our purposes.

*De novo* transcriptome assembly requires a reference to be built using a distinct list of all transcripts across all samples under study. Adding an official reference genome during transcriptome assembly (performed by StringTie merge) allows the subsequent StringTie runs to assemble known transcripts more accurately than it could otherwise. The *de novo*-built reference (rather than an official transcriptome reference) is used to guide the assembly of 'transcripts' in a second round of transcript assembly and gene sequence counting (in StringTie). Each distinct transcript is given a unique "MSTRG" identifier, allowing comparison between control and infected samples. Based on these hypotheses, *de novo* transcript assembly would be predicted to identify a mixture of intergenic sequences and alternative transcript isoforms of previously identified genes. The programmes used, their functions and the references implemented are presented in Figure 5.10.



**Figure 5.10** Flow chart of processes used to identify over- or underrepresented trapped intergenic sequences and additional overrepresented genes in control vs infected U937 GT cells through *de novo* transcript assembly. Flow chart of processes undertaken to achieve a list of over- or under-represented sequences in control vs infected samples. Blue arrows indicate the processes performed to generate a *de novo* reference file from distinct transcripts in all samples. Green arrows indicate processes performed after this reference was incorporated into the analysis. (Adapted from Yiwen Ding project dissertation, 2019).

Within the dataset, the defined genomic regions displaying a bias in representation often contained a gene which may be genuinely disrupted (as described in Table 5.4) or functionally perturbed by local gene trap vector insertion. To assess the potential for these latter cases, up- and downstream genes as well as the relative locations to the biased genomic region are additionally reported (Table 5.4). There are a number of disrupted gene overlaps with the prior conventional analysis (e.g., *SQLE* and *C5ORF60*) because conventional GT transcripts were not actively removed. Other disrupted genes identified with this method may have been missed by the previous reference-based analysis (see Appendix C, Section C.3) due to the strict use of the open reading information provided by the reference library (by StringTie) to define a gene's extent. By the nature of a complete GT transcript, it is likely to start part through a gene.

**Table 5.4 Intergenic mutant transcripts with altered representation in STM-infected samples.** A *de novo* method of transcriptome assembly to identify additional trapped sequences with altered representation. These might be formed from GT cassette insertion into intergenic regions spliced with open reading frame-like sequences; or aberrantly spliced transcripts produced by GT cassette insertion. The MSTRG identifier refers to a 'gene' locus identified as significantly over or underrepresented in control samples compared to infected. These are defined by gene coordinates in the centre of the table. Directly disrupted genes are listed to the left. Known genes lying up or down-stream of the MSTRG location are listed, together with their distances from the MSTRG. It is proposed that these proximal genes could be indirectly perturbed in expression by the gene trap insertion. Note that the fold change presented here is in the inverse direction (control, n=6 vs infected, n=6) to that presented in Section 5.2.3. <sup>1</sup>Log FC of averaged Control samples. Data from 6 experiments. (Adapted from Yiwen Ding's project dissertation).

MSTRG	Disrupted gene	Upstream gene distance	Upstream gene	Coordinates	Downstream gene	Downstream gene distance	Log FC <sup>1</sup>	Adj. p-value
11443	<i>SQLE</i>			chr8:125,016,629-125,024,243			4.51	7.10E-3
10644	<i>CTTNBP2</i>			chr7:117,713,682-118,232,581	<i>LSM8</i>	110700	8.90	7.10E-3
5218		40000	<i>KCNF1</i>	chr2:10,787,757-11,215,256	<i>FLJ33534</i>	48000	4.58	7.10E-3
9190	<i>C5orf60</i>			chr5:179,639,543-179,660,461	<i>AC136604.3</i>	2774	7.82	7.10E-3
11316	<i>VPS13B</i>			chr8:99,511,093-99,516,238			7.54	7.10E-3
8108	<i>AL136537</i> <i>NWD2</i>			chr4:36,876,010-37,515,509			-7.31	7.10E-3
9674	<i>SIM1</i>			chr6:100,351,929-100,446,802			3.14	1.46E-2
1865	<i>C11orf49</i>			chr11:46,922,094-47,082,011			1.90	1.66E-2
7615	<i>GSK3B</i>			chr3:119,960,133-120,061,238			-4.14	1.82E-2
2787		88415	<i>DIAPH3</i>	chr13:60,028,959-60,675,558	<i>TDRD3</i>	-55390	7.75	2.32E-2
9877	<i>SAMD5</i>			chr6:147,549,928-147,599,107			-7.58	2.45E-2
243	<i>KIF2C</i>			chr1:44,751,991-44,756,828			4.15	2.83E-2
4149	<i>HSBP1, MLYCD,</i> <i>CDH13, OSGIN1</i>			chr16:83,543,584-84,165,683			-7.46	2.83E-2
4015	<i>TENT4B</i>	16270	<i>HEATR3</i>	chr16:50,074,980-50,222,388			1.97	2.83E-2

<b>MSTRG</b>	<b>Disrupted gene</b>	<b>Upstream gene distance</b>	<b>Upstream gene</b>	<b>Coordinates</b>	<b>Downstream gene</b>	<b>Downstream gene distance</b>	<b>Log FC<sup>1</sup></b>	<b>Adj. p-value</b>
3895	<i>CLEC19A</i>			chr16:19,309,185-19,315,184			2.82	2.83E-2
6809	<i>SLC6A6P</i>			chr21:19,220,149-19,297,246			-9.09	3.30E-2
4622	<i>FHOD3</i>			chr18:36,338,580-36,501,382			1.09	3.60E-2
4392	<i>TEX2</i>			chr17:64,195,027-64,204,160			4.85	4.46E-2
11923		160125	<i>SMC2</i>	chr9:104,109,451-104,693,208	<i>OR13F1</i>	3846	7.69	4.46E-2

Of the newly identified, putatively disrupted, genes Vacuolar Protein Sorting 13 Homolog B (*VPS13B*) is associated with Cohen Syndrome, which features persistent minor infections among its numerous pathologies (Duplomb et al., 2019). *GSK3B* encodes a kinase that has recently been described as a key host infection factor for STM, as it promotes M2 polarisation through STAT3 activation (Gibbs et al., 2020; Panagi et al., 2020). Of its more well-known targets is beta catenin, as well as numerous others. Beta catenin, in turn, is known to regulate the distribution of cell adhesion molecules of the cadherin type, such as cadherin 13 (*CDH13*) also discovered here. *CDH13* has previously been identified as a host resistance factor against the malaria parasite, as has delta catenin (Mackinnon et al., 2016). Furthermore, avian *Cdh13* has an allele that offers protection against *Campylobacter jejuni* infection (Connell et al., 2013) and a *CDH13* variant in the Korean population protects against tuberculosis infection (Hong et al., 2017). C-Type Lectin Domain Containing 19A (*CLEC19a*) has been genetically linked to chronic periodontitis (Offenbacher et al., 2016).

This *de novo* transcriptome-based analysis additionally identified cortactin binding protein 2 (*CTTNBP2*) as a disrupted gene. *CTTNBP2* causes microtubule bundling in the dendrites of neurons for the generation of dendrite projections (Chen et al., 2012; Shih et al., 2014). The overrepresented StringTie bundled GT transcripts mapped to a nearly 600,000 bp region encompassing the majority of this gene, the subsequent downstream gene *LSM8*, as well as the intervening intergenic space. The expression of the *CFTR* gene (part of a postulated host response network based on literature analysis, Chapter 1) is dependent on multiple intronic and intergenic regulatory regions, one of which appears to be located in an intron of *CTTNBP2* (Ott et al., 2009; Zink et al., 2004). Altered regulation of *CFTR* gene expression stimulating an increased host cell survival advantage during STM infection as a result of the disruption of an intronic regulatory element seems plausible.

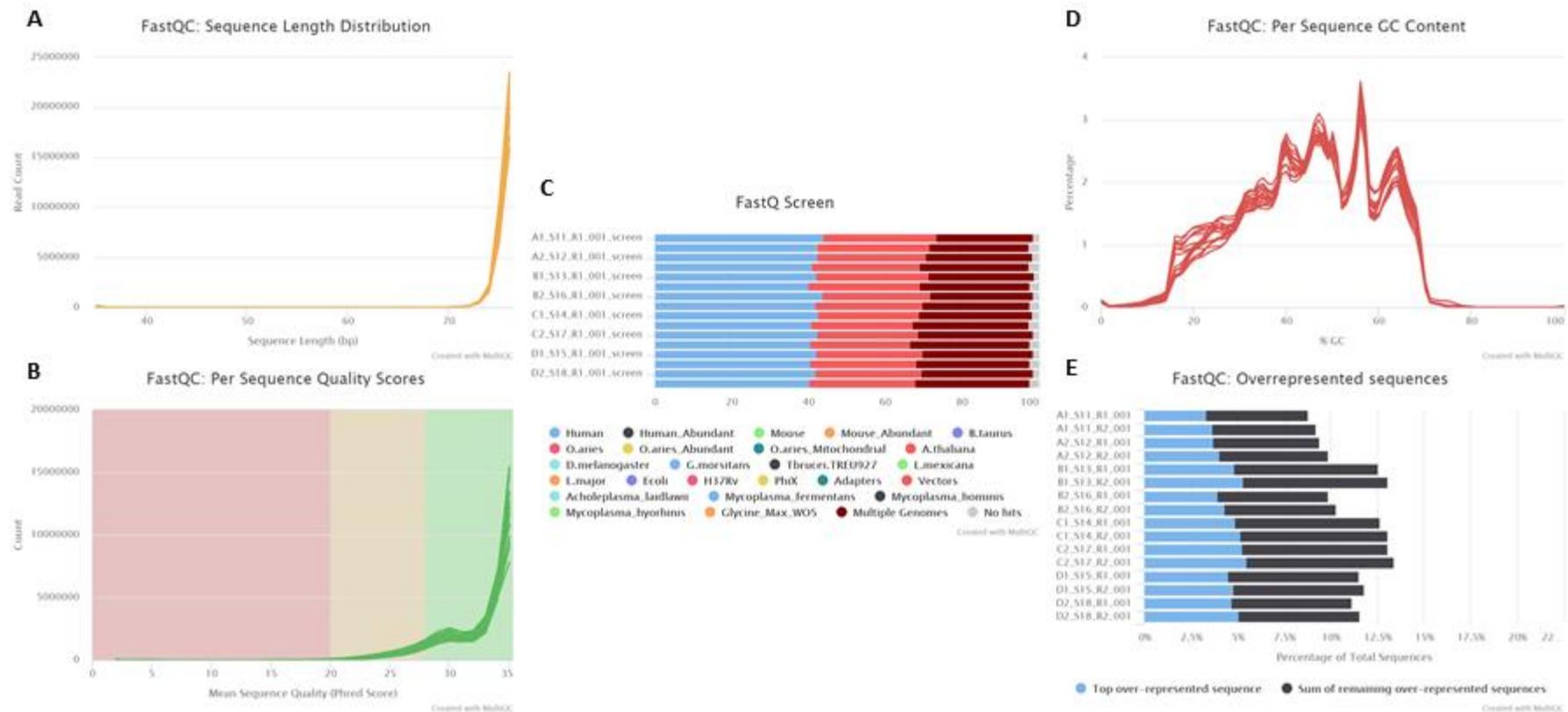
### 5.3. Bioinformatics analysis of the *F. tularensis* subsp. *holarctica* live vaccine strain (LVS) infection-host cell survival screen using the U937 gene trap library

The NGS files for the second screen using *F. tularensis* LVS as the infectious agent were received with the quality assessment data from Glasgow Polyomics. The infection, RNA extraction and molecular processing prior to sequencing to generate this data is described in Chapter 4.

#### 5.3.1. Raw DNA sequence quality assessment

Raw next generation sequence data from the LVS screen was returned with an initial quality assurance analysis using FastQ Screen and FastQC before collapsing the quality results generated into graph form with MultiQC. The equivalent graphs presented for the STM screen are presented here (Figure 5.11). The vast majority of reads are more than 70 bp long (Figure 5.11A) as expected. Most of the reads have a Phred score of  $\geq 30$  (Figure 5.11B), with a similar quality distribution to that found in the STM screen. Approximately 40% of reads are uniquely human (Figure 5.11C), about 25% are vector sequences, about 30% match multiple of the genomes tested (including human) and a small remaining proportion do not match any genome. This is comparable to the STM screen (Figure 5.4), and is acceptable. The range (about 15-70%) and overall shape of the per sequence GC content (Figure 5.11D) is similar to that generated by the STM screen. GT vector and primer sequences were found to cause the abnormally broad distribution as well as some of the sharp peaks in the GC content of sequences in the STM screen (Appendix C); thus this is likely the case for the LVS screen

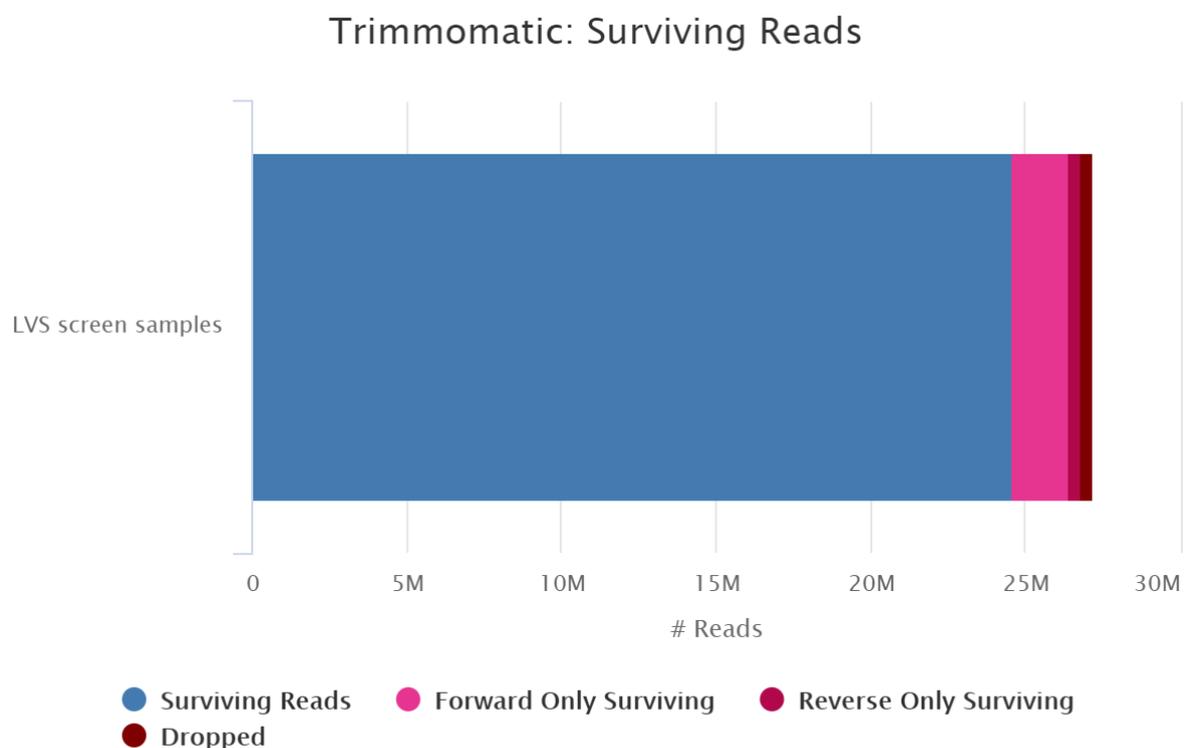
also. The percentage of overrepresented sequences as a proportion of the total number of reads per sample ranges from less than 9% (A- passage control) to over 13% (C-Infected) (Figure 5.11E). The topmost overrepresented sequence ranges from 3-6% across the samples. The percentage of total overrepresented sequences found in the raw LVS screen sequence files is less than that found in the raw STM screen sequence files (Figure 5.6). To get an impression of the identity of these overrepresented sequences, the FastQC data from a single sample file (B-Control forward) was compared to likely contaminant sequences. All reported overrepresented sequences for the raw sequence file were cDNA synthesis primer and GT cassette sequences (data not shown). On the basis of these graphs, the quality of the raw sequence files is comparable to the raw sequence files from the STM screen and are of sufficient quality to take forward.



**Figure 5.11 Raw NGS read quality for U937 GT library *F. tularensis* LVS screen is comparable to NGS read quality from U937 GT library STM screen.** Quality checks on raw NGS reads from U937 GT library *F. tularensis* LVS screen performed by Glasgow Polyomics and collated by MultiQC. Sequence length distribution, **A**, Per sequence quality (Phred) scores, **B**, FastQ Screen, **C**, Per sequence GC content, **D** and Overrepresented sequences, **E**, provided by FastQC (unless specified otherwise). Additional reference libraries were provided by FastQ screen to match to NGS reads, the vector reference library is in pink, otherwise the colours match those in the legend of Figure 5.4. The settings and colour coding for A, B, D and E are the same as those used in Figure 5.3A, B, Figure 5.5, Figure 5.6. Samples ( $n=8$  for all) are represented by single lines, **A**, **B**, **D**, or horizontal bars, **C** and **E**.

### 5.3.2. Quality processing

Quality trimming (to Q20) and minimum length filtering ( $\geq 20$  bp) was performed with Trimmomatic as described in Section 5.2.2.1 for the STM screen sequences. Surviving read pairs averaged at just under 25 million (Figure 5.12), which is not as many as those surviving for the STM screen (Figure 5.7B). The Trimmomatic paired output was used for the subsequent analysis.



Created with MultiQC

**Figure 5.12 Q20 trimmed and filtered LVS U937 screen sequencing samples retain more than 20 million surviving read pairs for genome alignment.** Raw sequencing files trimmed to an average Phred of 20 (Q20) per 4 bases using the sliding window operation before filtering for reads longer than 20 bases in Trimmomatic. Trimmomatic separates surviving reads into those in a complete pair, forward only reads and reverse only reads. The bar shows the average number for each output group from all infected and control LVS U937 GT sequencing samples (total  $n=8$ ) and was generated by the programme MultiQC.

### 5.3.3. Comparison of GT library representation between different passages of the GT library – greater passage number expected to affect representation

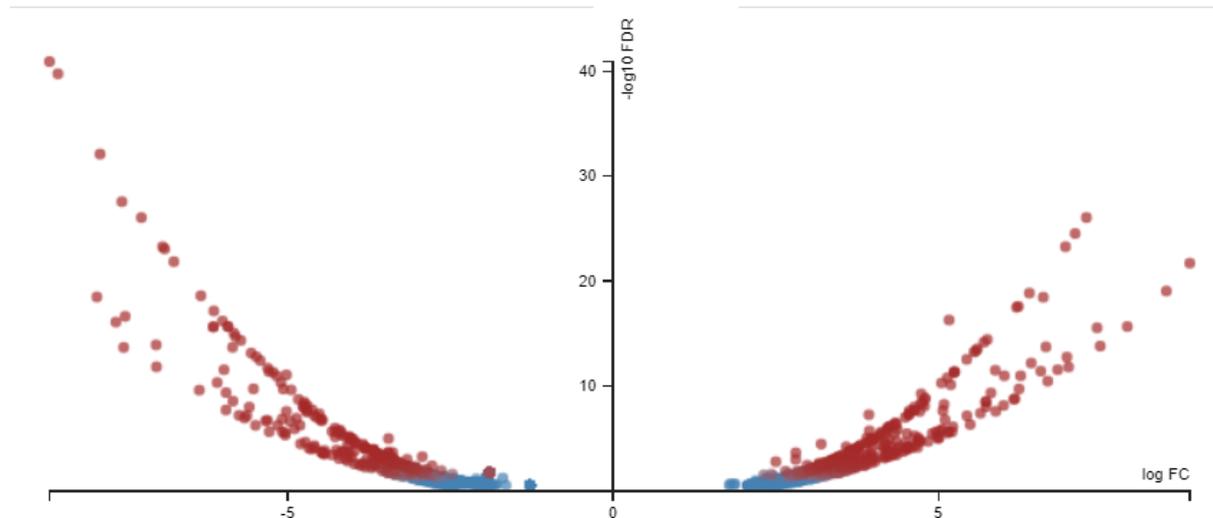
Alignment to the genome was performed with HISAT2, transcripts were assembled and counted with StringTie as performed for the STM screen analysis (software preferences can be found in Chapter 2). Experiment A of the *F. tularensis* U937 GT screen was performed with GT library cells at passage 11 compared to Experiments B-D which were performed with passage 3 library cells. The genetic representation of the library is expected to diverge and reduce with increasing passage number due to phenotypic differences in replication speed and fragility/ changes in vitality deriving from differing genotype. If the gene representation between uninfected and infected GT populations were

statistically analysed inclusive of a divergent GT library population (potentially Experiment A) infection relevant genes might be obscured. However, this must be balanced against the contribution in statistical power that a fourth biological replicate would provide (if Experiment A were included).

The specific effect of 8 additional passages on the library representation was unknown, therefore, a comparison between the Experiment A gene counts and the gene counts from Experiments B-D was required. The gene count data output from StringTie for Experiment A (passage 11) control was compared to B, C and D controls to identify how different Experiment A control data was from the other control samples. To do this the gene counts per gene were averaged with or without Experiment A for the control group before subtracting from the other. Upon examination of the data there was a visible difference between the gene count averages with and without Experiment A (data not shown). The potential reason for this is that mutations reducing proliferation or survival during normal monocyte culture might become underrepresented and *vice versa*. Additionally, stochastic variation in U937 replication and cell death could account for comparative under- and overrepresentation in the GT library prior to infection, and a greater passage number allows greater time for these effects to occur. Therefore, Experiment A data was excluded from the subsequent differential comparison.

#### 5.3.4. Identified Genes of Interest

For small sample sizes, estimating the variation per transcript is difficult and various tools to statistically analyse count data have been designed to account for this (Robles et al., 2012; Seyednasrollah et al., 2015; Sonesson and Delorenzi, 2013); including limma voom (Law et al., 2014). limma: voom performs well with 3 replicate samples per group and produces a lower rate of false positives when the FDR threshold is increased compared to other tools (Seyednasrollah et al., 2015; Sonesson and Delorenzi, 2013). Based on this, the tool was retained as the method of choice for statistical analysis of gene counts for the *F. tularensis* U937 GT screen dataset. The mutated gene sequences identified as differentially represented in this screen at an FDR threshold of 0.3 are presented in Figure 5.13.



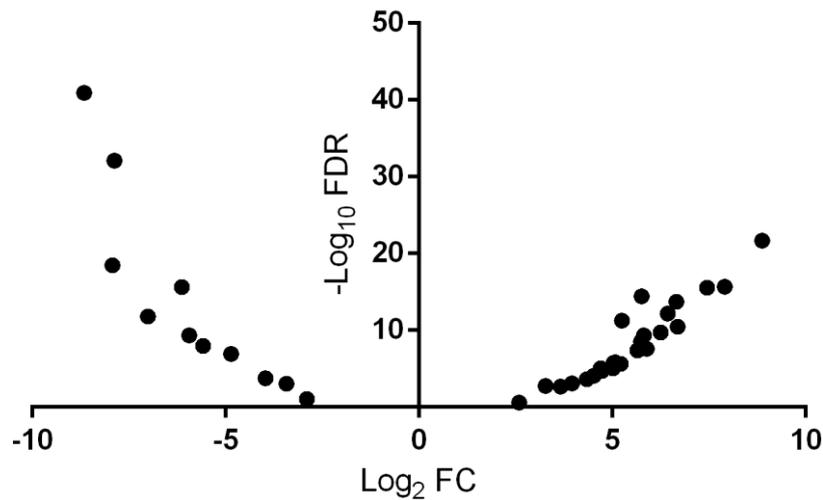
**Figure 5.13 *F. tularensis* differentiated U937 GT library screen resolves 1,867 mutant gene sequences differentially present in infected U937 population.** Mutant versions of genes (dots) promoting survival in the infected population have positive values of log fold change (FC) plotted on the x axis and mutant genes reducing survival in the infected population have negative values of log FC.  $-\log_{10}$  FDR plotted on the y axis indicates how significant the fold change (Infected vs Control) is for a given mutant gene. Less significant mutant genes are in blue and the more significant ones are in red. The comparison analysis was achieved using *limma: voom* in *Degust* with a false discovery rate threshold (Benjamini Hochberg correction) of 0.3.  $n = 3$  per group. Volcano plots like this one, Figure 5.8 and Figure 5.9 tend towards a parabola – this one is more clearly parabolic, likely due to the greater number (>20 x) of genes identified as significant by *limma*.

*limma: voom* compares average log fold change (FC) values per treatment group (mock-infected, infected), which may obscure large differences between experiments. An examination of log FC values between infected and mock-infected samples from each experiment for genes presented in Figure 5.13 showed a large amount of variation between experiments (in several cases only 2 of 3 experiments agreed in the direction of the fold change). In other cases, the log FC was due to very small changes in gene count at low numbers. The high count dispersion in the upper group per gene (e.g. ENSG00000225744) is concerning, given the small number of replicates; especially as *limma* struggles with comparisons between groups with differing amounts of dispersion (Soneson and Delorenzi, 2013). A similar amount of variation for some genes was seen in the STM screen (data not shown) though the higher replicate number provided a clear indication of what the inherent biological dispersion was. With lower numbers of biological replicates, what is biological variation or experimental noise is less clear.

To identify those genes which had more consistent differences in mutant gene representation, discovered genes were manually filtered to keep only those where all experiment log FC had the same direction. To focus on those genes in which the difference in representation were greatest, only genes with a log FC  $\geq 2$  or  $\leq -2$  for all experiments were kept.

In conventional differential gene expression analysis, low gene counts are the result of low transcript numbers. In contrast, low gene counts in gene trap screens are the result of low representation of the mutant transcript and perhaps a low population of that mutated cell prior to selection. Changes in the representation of these mutant transcripts with selection could still indicate a protective and/or deleterious effect; but experimental noise would impact the log FC to a greater degree, causing potential false positives. Therefore, a decision was made to select for genes, which had gene

counts higher than 10 in all the samples in the higher treatment group. The combination of these criteria reduced the number of differentially represented genes to 39 (Figure 5.14).



**Figure 5.14 Differentially represented gene sequences in *F. tularensis* LVS infected U937 samples reduced to 39 when further constrained.** Differentially represented genes output from *limma: voom* at a FDR threshold of 0.3 (Figure 5.13) were manually filtered to keep only gene sequences with fragment counts higher than 10 in all the samples in the higher treatment group. Only those gene sequences where all experiment log FC had the same direction were kept. Additionally, only genes sequences with a log FC  $\geq 2$  or  $\leq -2$  for all three experiments were kept. These criteria reduced the number of genes to 39.

The constrained list of genes (Table 5.5) includes one uncharacterised protein coding gene (*LOC730100*), 4 pseudogenes, 2 antisense versions of genes, 1 duplicate (*MARF1*) with different Ensembl gene IDs and 4 predicted long non-coding RNAs, 3 of which overlapped a protein coding gene.

**Table 5.5 Molecular function of differentially represented gene sequences in *F. tularensis* LVS infected U937 samples constrained by post-limma filters. Genes represented in Figure 5.14.**

*Ensembl gene IDs for differentially represented sequences at a FDR threshold of 0.3. The comparison direction is Infected vs Control (n=3 each). The Ensembl IDs are from reference annotation provided to StringTie. The IDs are used to search for the gene sequence information on the Ensembl database <http://www.ensembl.org/index.html> (Cunningham et al., 2019), using Ensembl release 99 (January 2020). The molecular function information is obtained from the Ensembl, Uniprot databases <https://www.uniprot.org/> (The Uniprot Consortium, 2019) and individual papers (Peng et al., 2019).*

Ensembl gene ID	SYMBOL	Log2 FC	FDR	Molecular Function
ENSG00000031691	CENPQ	-8.66	1.21E-41	Subunit of complex, involved in assembly of kinetochore proteins, mitotic progression and chromosome segregation
ENSG00000159459	UBR1	-7.88	7.93E-33	Ubiquitinates destabilised proteins to signal their degradation
ENSG00000231918	LOC730100	8.87	1.97E-22	lncRNA
ENSG00000189129	PLAC9	-7.93	3.25E-19	Inhibits cellular proliferation, involved in embryonic development
ENSG00000050327	ARHGEF5	7.91	2.15E-16	activates Rho GTPases, specifically RHOA, RHOB, weakly activates RHOC and RHOG
ENSG00000166783	MARF1	-6.14	2.31E-16	represses transposable elements during oogenesis, may protect from DNA double stranded breaks
ENSG00000277140	MARF1	-6.14	2.31E-16	See above.
ENSG00000048471	SNX29	7.44	2.84E-16	Produces circRNA to act as molecular sponge for miR-744, resulting in derepression of Wnt5a and CaMKIIδ
ENSG00000050730	TNIP3	5.75	3.69E-15	Binds to zinc finger protein TNFAIP3 and inhibits NF-kappa-B activation
ENSG00000198420	TCAF1	6.66	1.87E-14	Positively regulates the plasma membrane cation channel TRPM8 activity, involved in TRPM8 plasma membrane recruitment
ENSG00000103811	CTSH	6.43	6.53E-13	Involved in lysosome mediated protein degradation
ENSG00000274225	AP001065.3 (lncRNA)	-7.02	1.48E-12	lncRNA
ENSG00000232832	LMLN antisense transcript	5.24	5.15E-12	LMLN is a metalloprotease, moves from internal structures to leading edge of migrating cells
ENSG00000277406	SEC22B4P	6.68	3.43E-11	SEC22 homolog B4, pseudogene
ENSG00000132639	SNAP25	6.24	1.91E-10	diverse vesicle trafficking and membrane fusion processes
ENSG00000235137	HSP90AB6P	-5.95	4.29E-10	heat shock protein 90 alpha family class B member 6, pseudogene
ENSG00000213639	PPP1CB	5.81	4.63E-10	Functions as part of a holoenzyme, dephosphorylates vast numbers of proteins

Ensembl gene ID	SYMBOL	Log2 FC	FDR	Molecular Function
ENSG00000174473	GALNTL6	5.74	3.13E-09	Catalyses first reaction in O-linked oligosaccharide biosynthesis
ENSG00000187240	DYNC2H1	5.74	9.82E-09	Involved in generation of cilia
ENSG00000128918	ALDH1A2	-5.59	1.02E-08	Converts retinaldehyde to retinoic acid
ENSG00000231298	MANCR	5.89	2.56E-08	mitotically associated long non coding RNA
ENSG00000204959	ARHGEF34P	5.65	3.94E-08	Rho guanine nucleotide exchange factor 34, pseudogene
ENSG00000148180	GSN	-4.86	1.16E-07	Calcium-regulated, actin-modulating protein, promotes actin nucleation and capping
ENSG00000274423	SEC22B2P	5.07	1.33E-06	SEC22 homolog B2, pseudogene
ENSG00000038219	BOD1L1	5.03	1.72E-06	Protects stalled DNA replication forks
ENSG00000185515	BRCC3	5.21	2.35E-06	Metalloprotease that specifically cleaves 'Lys-63'-linked polyubiquitin chains
ENSG00000230551	LnRNA, overlaps with CSNK1A1	4.70	8.58E-06	CSNK1A1 preferentially phosphorylates acidic proteins
ENSG00000142765	SYTL1	5.02	8.59E-06	Binds phosphatidylinositol 3,4,5-trisphosphate (found in inner leaflet of plasma membrane), Rab27a effector
ENSG00000224116	INHBA antisense RNA 1	4.73	1.87E-05	INHBA has many functions, including steroidogenesis, can act in a paracrine, autocrine, and/or endocrine manner
ENSG00000143226	FCGR2C	4.52	7.51E-05	Fc fragment of IgG receptor IIc (gene/pseudogene)
ENSG00000143226	FCGR2A	4.52	7.51E-05	Binds to the Fc region of immunoglobulins gamma
ENSG00000143727	ACP1	-3.97	1.62E-4	Low molecular weight phosphotyrosine protein phosphatase, substrates are tyrosine phosphorylated proteins, low-MW aryl phosphates and acyl phosphates
ENSG00000273247	LnRNA, antisense to RAB33B	4.34	2.32E-4	RAB33B is involved in autophagy
ENSG00000163602	RYBP	3.96	7.56E-4	Acts as part of complex, required to maintain the transcriptionally repressive state of many genes
ENSG00000116754	SRSF11	-3.43	8.64E-4	May function in pre-mRNA splicing
ENSG00000122641	INHBA	3.27	1.73E-3	INHBA has many functions, including steroidogenesis, can act in a paracrine, autocrine, and/or endocrine manner
ENSG00000123106	CCDC91	3.67	2.11E-3	Involved in sorting of hydrolases to lysosomes

Ensembl gene ID	SYMBOL	Log2 FC	FDR	Molecular Function
ENSG00000255146	LnRNA, overlaps partially with OR9Q1	-2.90	8.67E-2	Odorant receptor
ENSG00000004766	VPS50	2.59	2.37E-1	Functions as part of a complex, involved in endocytic recycling (bringing protein receptors back to the plasma membrane)

Within this constrained list are several genes with protein products involved in various types of protein trafficking, protein degradation and modulation of cytokine responses, which are described in further detail below.

5.3.4.1. Differentially represented genes with gene products involved in protein trafficking

Synaptotagmin-like protein 1 (SYTL1) is involved in vesicle trafficking to the cell surface (exocytosis) in neutrophils (Kurz et al., 2016), cytotoxic T lymphocytes (Holt et al., 2008) and prostate cancer cells (Johnson et al., 2005) as well as protein recycling away from the uropod (trailing end of migrating cell) (Ramadass et al., 2019). SYTL1 has two binding partners RAB8 and RAB27A (the former of which is involved in macropinosome trafficking) (Hattula et al., 2006). It also associates with nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, a multi-subunit enzyme responsible for producing ROS (Berkowitz et al., 2001). Syndetin (VPS50) functions as part of a tethering complex known as endosome-associated recycling protein (EARP) located at recycling endosomes (Gershlick et al., 2016; Schindler et al., 2015). Synaptosomal-associated protein 25 (SNAP25) plays an important role in membrane fusion, for a review see Han et al., (2017). Interestingly it was found to be important for *Salmonella*-containing vacuole (SCV) fusion with infection-associated macropinosomes (required for stability of the SCV/ replication niche of STM) (Stévenin et al., 2019). Ras-related protein Rab-33B (RAB33B) functions as part of a Golgi-retrograde transport pathway (Starr et al., 2010). It was found in an LC-MS screen pulled down by Autophagy-related protein 16-1 (ATG16L1) (Behrends et al., 2010) and is ubiquitinated by *Legionella pneumophila* effectors during infection (Qiu et al., 2017). Increased expression of *RAB33B* is associated with greater uropathogenic *E. coli* clearance from the urinary tract (Wang et al., 2018). The mature form of Pro-cathepsin H (CTSH), cathepsin H has an important role in protein degradation of lysosomes as the only lysosome-associated aminopeptidase (Kirschke, 2013). It is also important for recognition of viral RNA (Garcia-Cattaneo et al., 2012).

5.3.4.2. Differentially represented genes with gene products involved in immune response regulation

Tumor necrosis factor (TNF) alpha-induced protein 3 (TNFAIP3)-interacting protein 3 (TNIP3) inhibits NF-kappa-B activation induced by tumor necrosis factor, Toll-like receptor 4 (TLR4), interleukin-1 and PMA (Rath et al., 2018; Wullaert et al., 2007). It is also associated with ulcerative colitis (a chronic inflammatory bowel disease) (Zhang et al., 2020) and downregulated in long term ulcerative colitis (Low et al., 2019). Gelsolin (GSN) has roles in many cellular processes, including actin filament severing and capping, apoptosis, phagocytosis see Li et al., (2012) for a review. It also functions as part of a negative feedback loop with LPS and LPS-induced cytokines, reducing LPS-induced cytotoxicity (Cheng et al., 2017); as well as inhibiting macrophage death in *Streptococcus* infection (Fettucciari et al., 2015).

### 5.3.5. Gene ontology analysis of the *F. tularensis* U937 screen

Gene ontology enrichment analysis was performed for the over- and under- represented trapped genes in the *F. tularensis* U937 screen using the same settings and format applied to generate the results in Section 5.2.4. No GO terms were identified as enriched after adjusting for multiple testing (FDR threshold of 0.1) and thus no overlaps were found with the STM U937 screen.

### 5.3.6. Protein interaction networks from combined query gene list

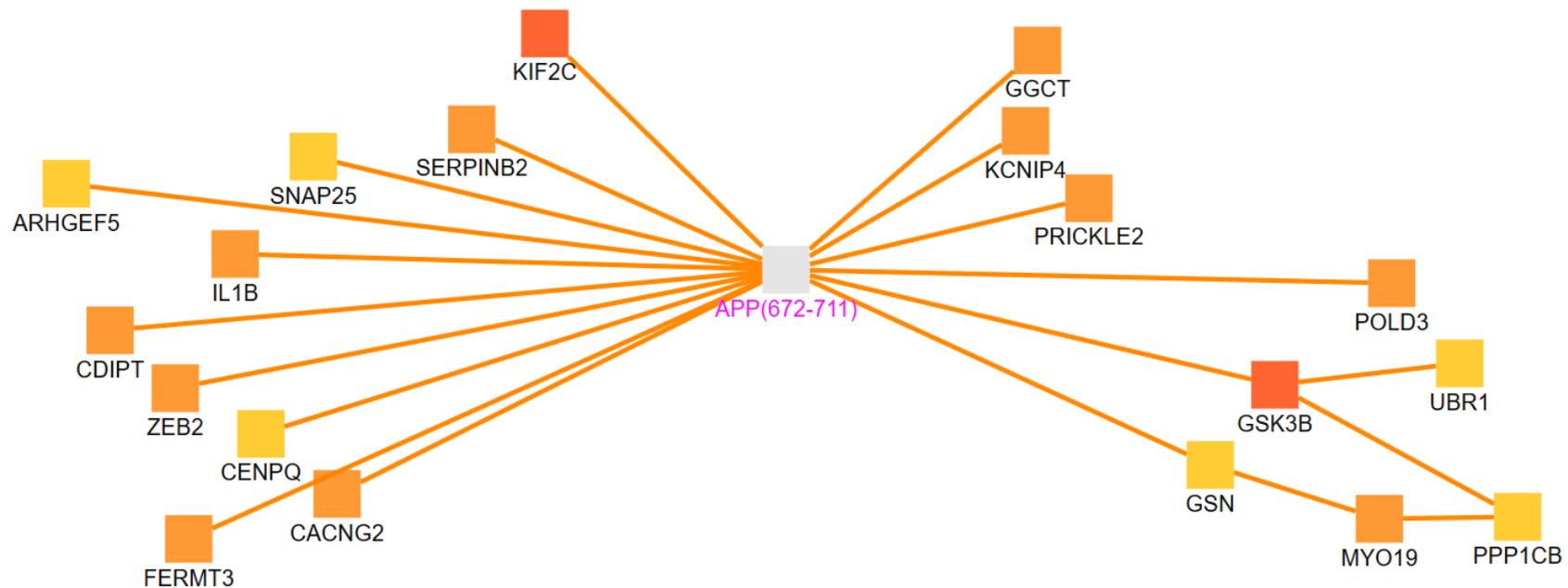
To define the scope of the respective gene trap screens performed, the number of distinct Ensembl gene IDs found (represented at least once in one or more differentiated U937 samples) were counted - the *F. tularensis* U937 screen had 13,721 and the STM U937 screen had 14,703. Within the latter list, 1,580 were pseudogenes, 1,218 were non-coding RNAs and the remaining (11,905) were protein coding genes, antisense transcripts, or transcript readthroughs. Based on these figures, the mutant library used in the STM screen has a coverage of 58-60% of protein coding genes in the human genome, as the human genome has between 19,901-20,376 protein coding genes (depending on the database queried) (Salzberg, 2018). As a proportion of the genome could not be assessed using this gene trap library, potential links between genes within and between the screens may be missing. A bioinformatics method of identifying potential 'missing links' is the use of protein interaction databases to generate induced interaction networks. A comparative analysis of 21 molecular network databases found that Search Tool for Recurring Instances of Neighbouring Genes (STRING) database and Consensus Pathway Database (CPDB) identified the most members of literature defined networks when subsets of these networks were queried (J. K. Huang et al., 2018). Therefore, these databases were chosen for the purpose of identifying potential missing links. As the aim of this project was to identify host infection factors common (or even universal) to different bacterial species, a decision was made to input the list of genes identified in both U937 screens (specifically those in Table 5.2 and Table 5.5) (as well as the 'disrupted' class in the secondary STM screen analysis, **Table 5.4**) into the above protein interaction databases together.

#### 5.3.6.1. Consensus Pathway Database

The CPDB ([www.ConsensusPathDB.org](http://www.ConsensusPathDB.org)) (Herwig et al., 2016; Kamburov et al., 2013) is an online resource that creates visual networks of interactions by integrating protein interaction data from many distinct and overlapping network databases. Importantly, the resource can add in 'missing links' (a.k.a. intermediate nodes) to better connect the network. Only medium and high confidence protein interactions (as defined by CPDB) were allowed to connect nodes. A region of interest within this network is a hub of 19 direct or indirect protein-protein interactions with the Amyloid Precursor Protein (APP) (Figure 5.15), which is at the heart of Alzheimer's disease pathology. An explanation for this unexpected observation might be a simple over-representation of APP-interacting proteins in the database, brought about by the intense research scrutiny of this protein over the past thirty years. If this were the case, one would see more induced network modules containing APP as an intermediate node than would be expected by chance. To assess this likelihood, 5 random gene sets from the human genome containing the same number of query proteins were entered into the induced network module with the same settings with the addition of low confidence protein interactions. None of these induced network modules included APP as an intermediate interactor (data not shown). In support of this being a genuine interaction profile, a more direct connection between Alzheimer's disease and infection has been recently suggested – both in terms of infection as an inflammatory trigger for neurodegenerative processes, but also, more fundamentally, the theory that APP might play a direct role in macrophage infection regulation (H. Li et al., 2018).

Macrophages express APP (Vehmas et al., 2004) and the gene's deletion seems to result in an altered macrophage inflammatory response (Puig et al., 2017). More directly, amyloid- $\beta$  protein, a modified form of APP, is protective in *in vitro* and *in vivo* models of STM and *Candida albicans* infection (Kumar et al., 2016; Soscia et al., 2010).

Furthermore, these interactions were replicated in InnateDB (<https://www.innatedb.com/>) (Breuer et al., 2013) when constrained to InnateDB curated interactions. APP and the interacting query genes in Figure 5.15 were entered into GeneMania, another interactive network amalgamator (Warde-Farley et al., 2010), which identified high confidence protein-protein interactions between APP, FERMT3, KIF2C, CACGN2, GGCT, CDIPT, PRICKLE2, ARHGEF5, SERPINB2, IL1B, CENPQ, POLD3, KCNIP4, SNAP25 and ZEB2.



**Figure 5.15 Protein-protein interactions between both screen gene products converge on Amyloid-beta precursor protein (APP).** The combined list of differentially represented genes in the original analysis of the STM screen (orange squares), the disrupted genes from the de novo analysis (red squares) and the differentially represented genes in the F. tularensis LVS screen (yellow squares) was entered into the CPDB <http://cpdb.molgen.mpg.de/CPDB> induced network modules analysis (Kamburov et al., 2013). Each of these interactions were identified based on high throughput protein interaction experiments, such as (Hein et al., 2015; St-Denis et al., 2016; Vinayagam et al., 2011). Medium and high confidence protein interactions were allowed in the network analysis as well as intermediate nodes (grey squares) not included in the query list. The network presented here is a truncated version of that generated to highlight a subset of the interactions- all intermediate nodes with fewer than four interactions were hidden as well as all unconnected nodes resulting from this.

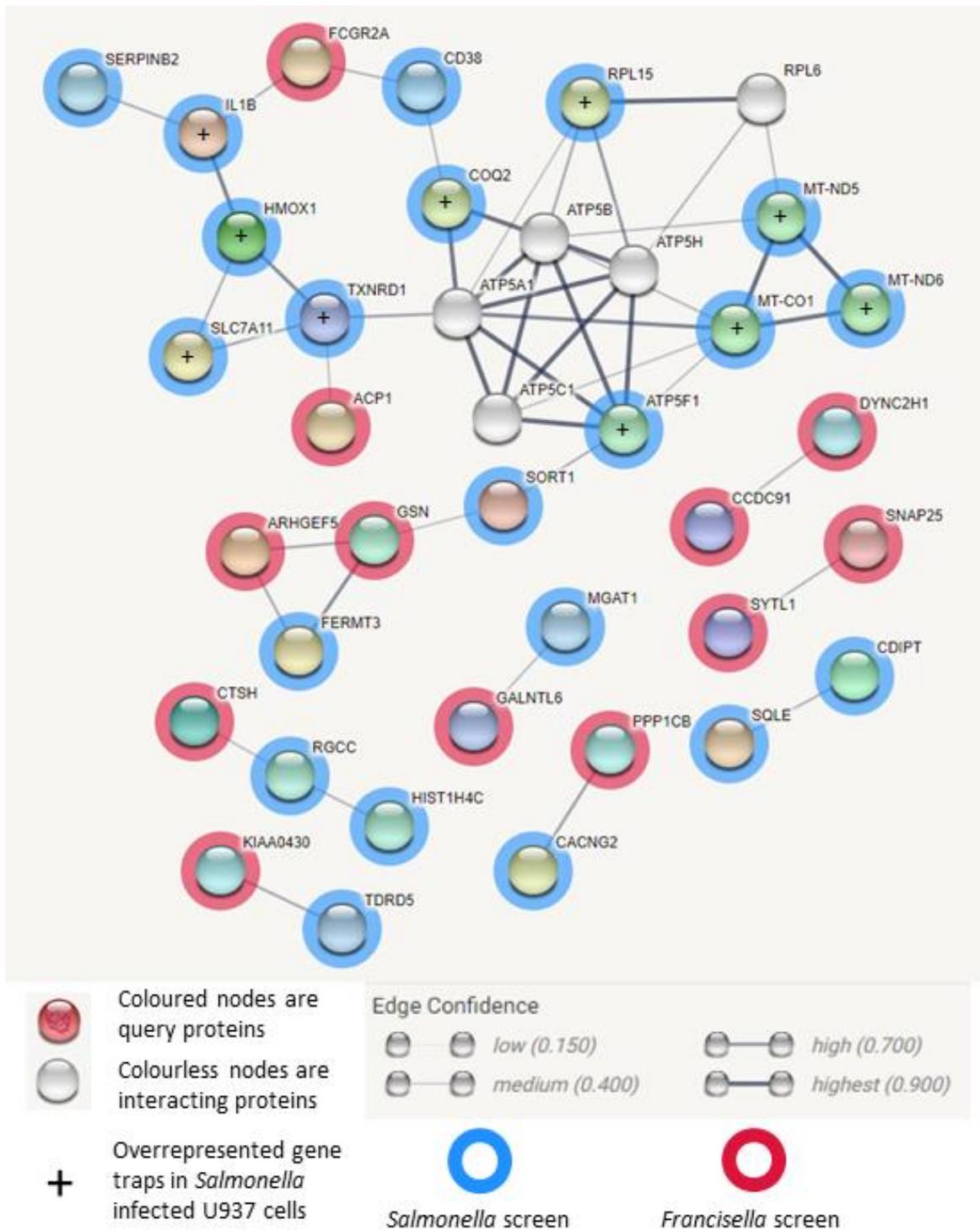
Besides the APP 'network hub', other missing link network hubs (each interacting with 3 query proteins) include Myosin light chain kinase, smooth muscle (MYLK), PH domain leucine-rich repeat-containing protein phosphatase 1 (PHLPP1), Golgi-associated, gamma adaptin ear containing, ARF binding protein 2 (GGA2) and Polycomb group RING finger protein 5 (PCGF5). Besides physical interactions with the query proteins, these network hubs have additional connection(s) to infection or the immune response. MYLK serves as a protein scaffold to allow NF- $\kappa$ B inflammatory signalling (Tauseef et al., 2012) and promotes asthmatic inflammation in a mouse model (Ting Wang et al., 2014). PHLPP1 regulates LPS-stimulated immune responses in macrophages (Cohen-Katsenelson et al., 2019) and knockdown reduces lysosome activity resulting in increased STM burden in bone marrow-derived macrophages (BMDMs) (Fischer et al., 2019). An *Ehrlichia chaffeensis* (an intracellular bacterial pathogen) effector depletes Polycomb group RING finger 5 (PCGF5) during macrophage infection (Zhu et al., 2017) and *PCGF5* knockdown increases *E. chaffeensis* burden (Mitra et al., 2018). Polycomb group RING finger proteins function as part of complexes to transcriptionally silence genes, (review in Connelly and Dykhuizen (2017)). Golgi-associated, gamma adaptin ear containing, ARF binding proteins are involved in protein and vesicle transport between the Golgi and endosomes (Uemura and Waguri (2020) for review) and GGA2 is required for appropriate lysosome protein transport (Hida et al., 2007).

The identification of APP as a missing link network hub between many identified genes suggests its importance as a host infection factor in each of the U937 screens. If it is an important host infection factor, why did it fail to emerge from either of the screens? Interestingly, *APP* was represented in the STM screen at a very low level (mean fragment count of 1.1), as well as two of the other missing links, *MYLK* (count of 2 in 1/12 samples) and *PHLPP1* (counts of 3, 7, 4 in 3/12 samples). *APP* and *PHLPP1* were also represented at low levels in the final samples of the *F. tularensis* screen. Extremely low fragment counts attributed to a given gene may be misidentified matches (during genome alignment by HISAT2). If the fragment counts are genuine, it is likely that the representation of *APP* KO GT cells in the differentiated population was extremely low before selection. Bacterial infection is an inherently noisy form of selection. Strain based acid and heat resistance profiles and bistable flagellar and virulence gene expression mean that *Salmonella* species exhibit various stochastic behaviours (Aspidou et al., 2019; Bailly-Bechet et al., 2011; Koutsoumanis and Lianou, 2013; Stewart et al., 2011) which contribute to the noise of selection. Extremely low representation of a given GT mutation within a mutant library prior to selection is likely to lead to random loss or retention of the host cell(s) harbouring the mutation. Any filter used to remove extremely low representation genes in differential gene expression data will be somewhat arbitrary, as it is unknown when an experimental signal could be distinguished from the (stochastic or bioinformatics) noise. Nevertheless, a filter based on the gene count and representation across differentiated U937 cell samples would give an impression of the coverage of selectable mutations in protein coding genes. For this purpose, all Ensembl gene IDs identified within any differentiated U937 sample from the STM screen were filtered for those that had gene counts of more than 5 in at least 2 samples. Of the Ensembl gene IDs retained, only 373 (4% of retained gene IDs) referred to protein coding genes, which is 0.04-1.8% of the protein coding genome.

#### 5.3.6.2. Search Tool for Recurring Instances of Neighbouring Genes (STRING) database

The combined gene list was submitted to a second database of interaction data known as the Search Tool for Recurring Instances of Neighbouring Genes (STRING) database (<https://string-db.org/>) (Szklarczyk et al., 2019). All genes from Table 5.2, Table 5.4, and Table 5.5 with reviewed Uniprot accession IDs were used in the query (79 distinct genes) and the network was limited to 5 intermediate nodes (Figure 5.16). Potential sources of functional interactions were limited to

experimental data, gene fusions (gene orthologues that have merged to generate a single protein coding gene), databases (pathway and gene ontology), co-expression data and Pubmed texts, which could be combined to make a minimum interaction score of 0.4. Query nodes without any interactions were hidden to simplify the network presented (Figure 5.16). In contrast to CPDB, the STRING database presents functional associations rather than molecular interactions, such as influencing the transcription, translation or modification of an associated protein, jointly participating in a specific bodily function as well as direct physical binding. Each piece of evidence receives a confidence score, which are then combined to generate a total score per interaction. STRING also predicts associations based on homology from other species, considering the evolutionary distance and presence of any other homologues in the proteome. STRING defines nodes as proteins but does not distinguish between different isoforms or post-translationally modified versions.



**Figure 5.16 Products of genes identified in STM and LVS differentiated U937 GT screens may have functional associations.** *Protein interactions between products of genes from the STM screen including disrupted genes from the secondary analysis as well as genes in the constrained list from the F. tularensis LVS screen. Protein interactions identified using the STRING database, <https://string-db.org>, all genes with reviewed Uniprot accession IDs were used in the query (79 distinct genes, coloured circles). Potential interactions were mined from experiments, gene fusions, databases, co-expression data and Pubmed text with a total minimum interaction score of 0.4. A maximum of 5 non-query interacting proteins were allowed (grey circles). Blue and red halos indicate products of genes identified in the STM and F. tularensis LVS screens, respectively. KIAA0430 is MARF1. ATP5F1A is a duplicate of ATP5A1.*

To provide an example of the scoring system STRING uses, the interaction between mitochondrial 4-hydroxybenzoate polyprenyltransferase, (COQ2) (a query node) and mitochondrial ATP synthase subunit alpha (ATP5A1, an intermediate node) is described. Firstly, COQ2 and ATP5A1 are in the curated Reactome pathway 'mitochondrial protein import' (score 0.900); specifically, COQ2 is a candidate 'Cargo of TIMM23 SORT' (*Cargo of TIMM23 SORT [mitochondrial inner membrane]*, n.d.) and ATP5A1 is a candidate 'Cargo of TOMM40' (May, 2011). Both TIMM23 SORT (a presequence translocase protein complex) and Mitochondrial import receptor subunit TOM40 homolog (TOMM40) manage protein organisation within the mitochondria (Mick et al., 2012). Secondly, orthologues of COQ2 and ATP5A1 are consistently coexpressed in *Mus musculus* (score 0.215), *Plasmodium falciparum* (0.200) and *Schizosaccharomyces pombe* (0.052) under multiple conditions (overall coexpression score 0.113). Thirdly, orthologues of COQ2 and ATP5A1 were mentioned together in 4 PubMed abstracts (score 0.062); generating a final score of 0.909.

The ten nodes (RPL15, RPL6, MT-ND5, MT-ND6, MT-CO1, ATP5F1, ATP5C1, ATP5A1, COQ2, ATP5B, ATP5H) in the middle with the strongest interaction scores may function towards a common host cellular goal during STM infection. Indeed, three of the genes encoding these proteins (MT-ND5, MT-ND6, MT-CO1) were identified as a part of the enriched gene ontology process 'electron transport chain' in Section 5.2.4. Mitochondrial ATP synthase subunit beta (ATP5B), B1 (ATP5F1), D (ATP5H), gamma (ATP5C1) and ATP5A1 are all subunits of complex V of the electron transport chain, which catalyses the phosphorylation of adenosine diphosphate (ADP). COQ2, an enzyme involved in the synthesis of ubiquinone, has been identified in this STRING network as well as part of the gene ontology enriched component 'organelle inner membrane'. If the highlighted proteins function towards a common goal, the trapped versions of the query nodes would be expected to share a common represented direction in the STM U937 screen. Gene trap *RPL15*, *MT-ND5*, *MT-ND6*, *MT-CO1*, *ATP5F1* and *COQ2* are all over-represented in the STM-infected samples (Table 5.2, Figure 5.16). To assess the reliability of these functional associations, the query genes within this 11-node subnetwork were tested in CPDB and GeneMania induced network modules. Both Database networks were able to connect the query genes but required biochemical interactions (CPDB) or low confidence co-expression and pathway data (GeneMania). These loose connections demonstrate the differences in the type of interactions presented by CPDB and STRING.

Of note APP was not presented as an intermediate node in the STRING network generated from the combined gene list. Even when APP was additionally submitted as a query gene, the confidence score lowered to 0.15 and the number of non-query interactors increased to 100; the APP network hub in Figure 5.15 could not be replicated by STRING.

## 5.4. Assessment of phenotype in an independent mutant model

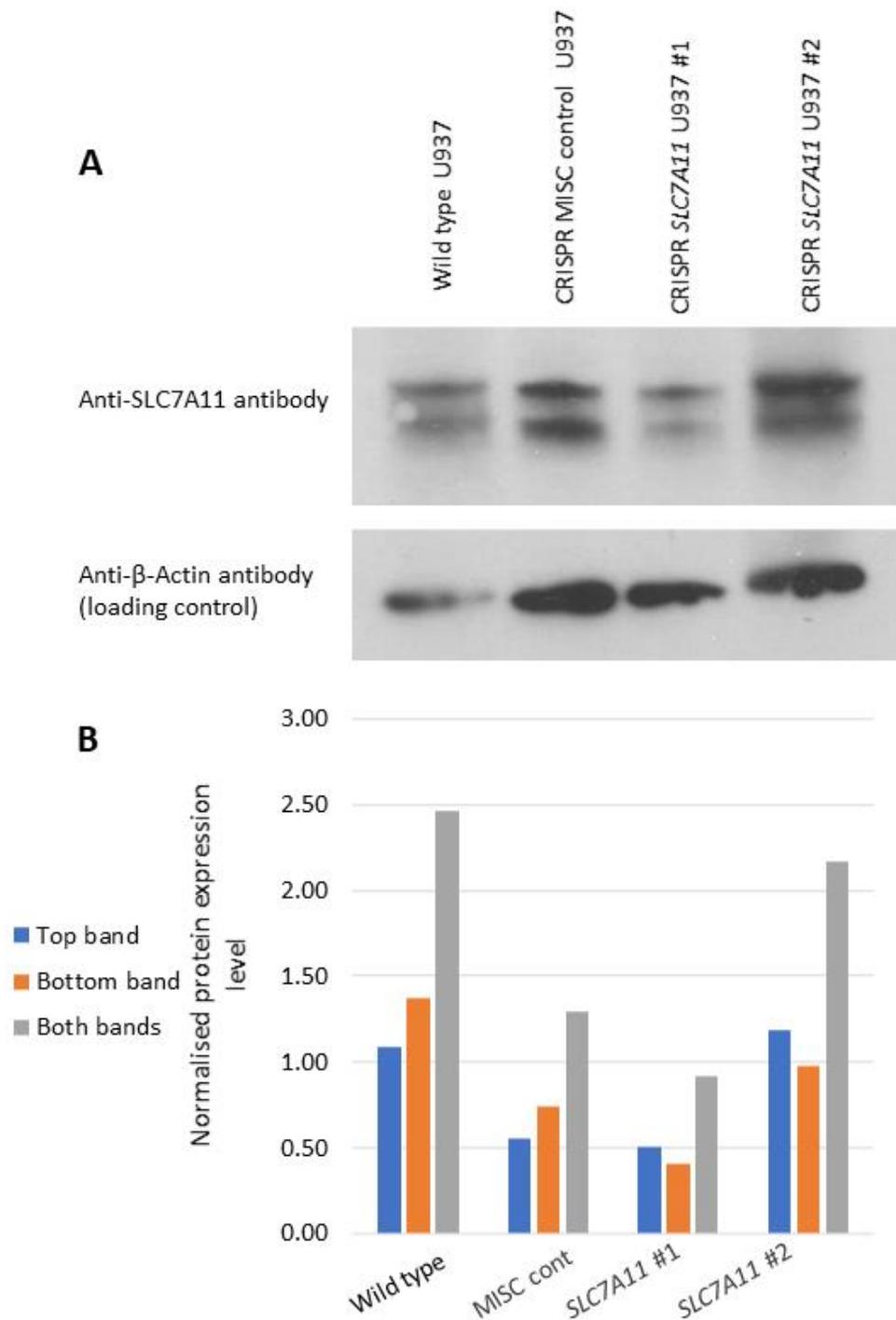
### 5.4.1. Creation of CRISPR constructs and transfection to create U937 CRISPR mutant pools

To validate the U937 GT library STM screen as a method to identify genes involved in the response to infection or processes essential for pathogen survival; an example gene hit was chosen. *SLC7A11* was chosen for this purpose as it was a gene hit with a high average log<sub>2</sub>FC in the STM screen at 0.1 FDR threshold in which the mutant version promoted host survival. Additionally, the drug inhibitor Sulfasalazine, (already on the market to treat rheumatoid arthritis) is available for the protein.

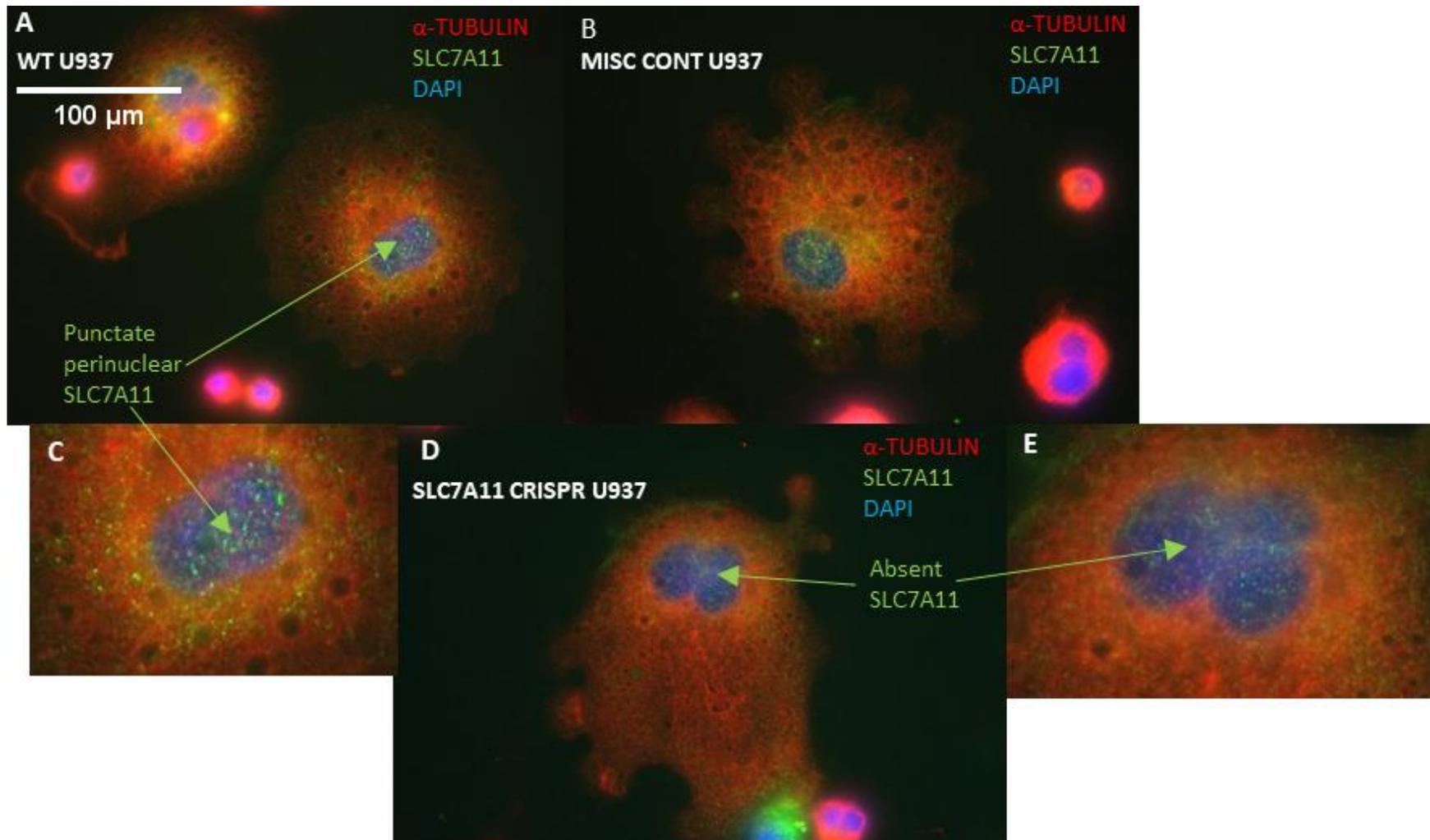
To this end, CRISPR mutant pools were made in U937 cells targeted at *SLC7A11* as well as a different type of CRISPR control (generated with an sgRNA containing three intentional point mutations). The sgRNA for the CRISPR control was based on the sgRNA designed for another gene hit from the U937 GT library STM screen, *CUL1*. The design of this CRISPR control (known as 'Miscullin' onwards) is expected to produce off-target effects similar to what would be found in the CRISPR KO *CUL1* pool, but without the on-target (*CUL1*) specific effects. CRISPR mutant pools instead of isolated lines were decided on as the model to be used due to lack of time and difficulty growing U937 cell culture at low densities. To enhance the effect of the CRISPR treatment, two gRNAs were designed for *SLC7A11* and the CRISPR control. The gRNAs were designed using CHOP CHOP, <https://chopchop.cbu.uib.no/> (Labun et al., 2019) and the modified sgRNA sequence for the CRISPR control was compared to the human genome using the NCBI nucleotide Blat database to ensure no homology with any other sequences. The chosen sgRNAs were then transfected as part of a pair of a CRISPR Cas9-containing plasmids per gene into a single batch of U937 cells.

Assessing the mutation status of the CRISPR target loci by sequencing gDNA from the whole CRISPR mutant pool (as was performed for the HEK-293 CRISPR KO cell lines in Chapter 3) would be labour intensive and difficult to draw conclusions from due to the likely highly heterogeneous nature of the pool. Instead, an anti-human SLC7A11 antibody was used to measure the comparative presence of protein by Western blot. Total protein was extracted from U937 CRISPR mutant *SLC7A11* pools #1, #2 (generated at different times), CRISPR control 'Miscullin' (Misc), and wild type cells, which was used to compare protein expression levels of SLC7A11 (Figure 5.17A). Ideally the  $\beta$ -actin loading control would have similar levels of staining across every sample to allow comparison of the protein expression of interest. As the loading control staining is very different between samples, the intensities of the SLC7A11 protein bands have been normalised against the  $\beta$ -actin loading control staining per sample before plotting them (Figure 5.17B). Based on the normalised expression level for both SLC7A11 protein bands the CRISPR KO *SLC7A11* #1 pool appears to have the lowest SLC7A11 protein expression.

Similarly, microscopy was used to confirm any reduction in cellular expression as well as to assess potential changes in subcellular protein distribution (Figure 5.18). SLC7A11 expression is localised to small dotted perinuclear regions in wild type cocktail-differentiated U937 cells (Figure 5.18C), perhaps indicating localisation to intracellular vesicles such as recycling endosomes, or in small discrete regions of the cell surface. SLC7A11 protein expression is barely visible or absent in cocktail differentiated CRISPR KO *SLC7A11* #1 pool cells (Figure 5.18E), confirming the Western blot results. Therefore, CRISPR KO *SLC7A11* #1 pool cells were taken forward to validate the infection phenotype discovered in the STM U937 GT screen.



**Figure 5.17 U937 CRISPR KO SLC7A11 pool #1 has much less SLC7A11 protein expression than wild type or Miscullin control. SLC7A11 protein expression in U937 Wild type, CRISPR Miscullin control (MISC cont), CRISPR KO SLC7A11 pool #1 and #2 measured by Western blot, A. Western blot anti-SLC7A11 band intensities measured by ImageJ and normalised based on the intensities of each anti-β-Actin band, B. Experiment performed once.**



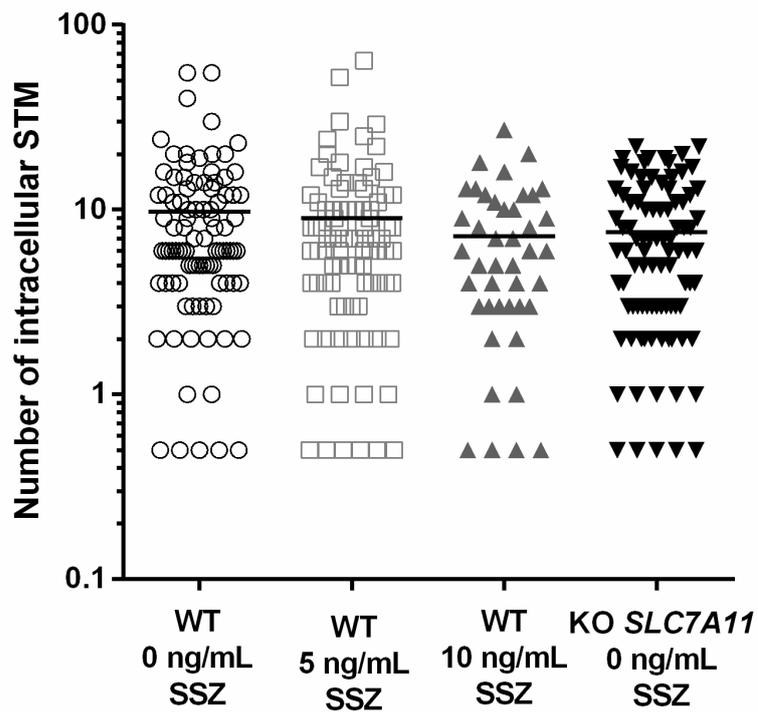
**Figure 5.18 U937 CRISPR KO SLC7A11 #1 pool has reduced SLC7A11 staining.** *Immunofluorescent micrographs of U937 wild type (WT) A, C, Miscullin CRISPR control pool (MISC) B and CRISPR KO SLC7A11 pool #1 D, E.  $\alpha$ -Tubulin is stained red with mouse anti-human  $\alpha$ -Tubulin primary antibody and Alexa Fluor 594 anti-mouse antibody. SLC7A11 is stained with rabbit anti-human SLC7A11 primary antibody and Alexa Fluor 488 anti-rabbit antibody. C and E are digitally magnified images of A and D, respectively. All images were taken on an upright epifluorescence microscope at a total magnification of 1000x. One experiment was performed.*

#### 5.4.2. Effects of genetic ablation or pharmacological inhibition of SLC7A11 on STM burden

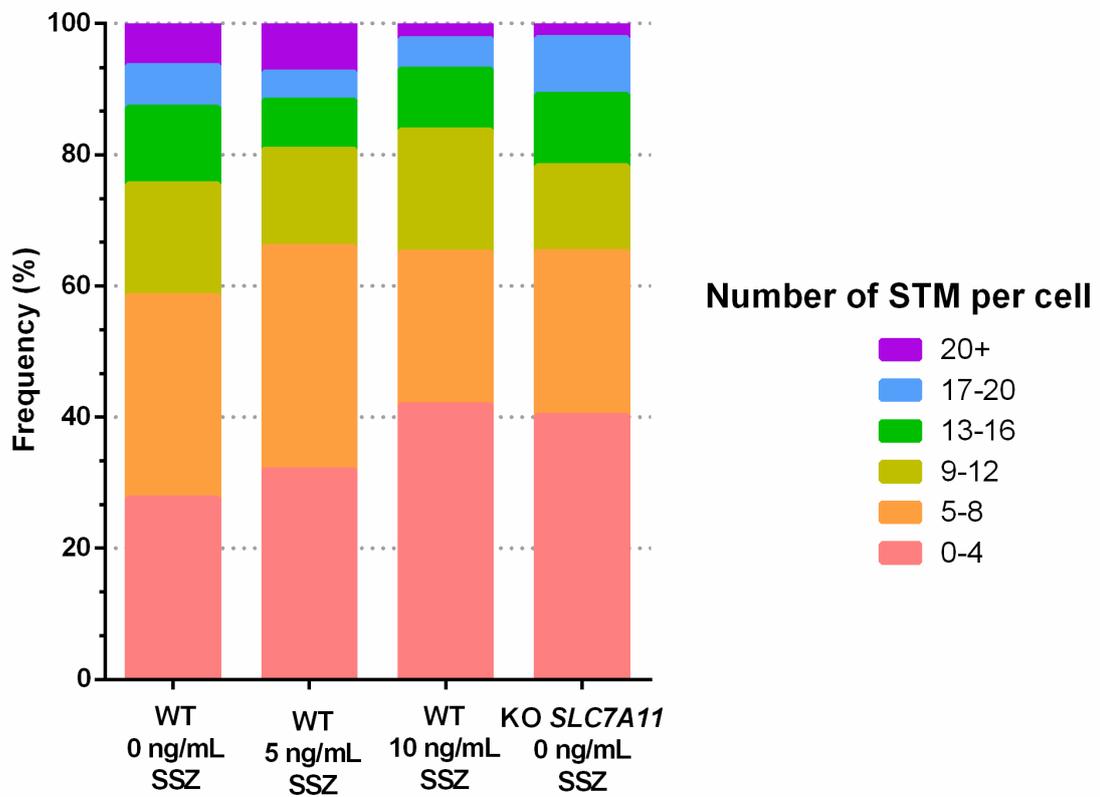
Differentiated U937 cells of various genotypes were inoculated with green fluorescent protein-expressing STM bacteria and, in parallel, a wild type subset of these differentiated U937 cells were treated with different concentrations of the SLC7A11 inhibitor, Sulfasalazine. The availability of a drug inhibitor for one of the candidate proteins identified in the screen is useful, as it allows interrogation of the infection phenotype using a non-genetic approach. Sulfasalazine (SSZ), a drug that inhibits SLC7A11, is predicted to have bacteriostatic activity in the intestine (Belluzzi et al., 2010; Pinczowski et al., 1994). This is a potential confounder for our experimental assay because we wish to examine drug action via host cell protection as opposed to direct antibacterial action. To determine an appropriate concentration lacking such activity for use in infection experiments, bactericidal and bacteriostatic assays were carried out. Concentrations of sulfasalazine (0 ng/mL, 6.25 ng/mL, 12.5 ng/mL, 25 ng/mL, 50 ng/mL), together with equivalent vehicle controls, were added to paper disks and these were placed on both growing STM lawns on LB-agar plates (bacteriostatic and bactericidal action) and previously grown STM lawns (bactericidal action). No bacteriostatic or bactericidal activity was seen at any of the concentrations (data not shown).

To determine the effect of SLC7A11 inhibition on the STM burden in U937 cells, 4 experimental groups were used, all of which were treated with the differentiation drug cocktail and inoculated with  $4 \times 10^7$  GFP expressing STM /well on day 6. This is 10-fold less than the number of bacteria /well to be commensurate with the STM screen. Fewer bacteria were added as it was thought that  $4 \times 10^8$  STM /well might overwhelm any microscopy images. The experimental groups were vehicle control treated CRISPR KO SLC7A11 pool #1 cells, vehicle control treated wild type U937 cells and wild type U937 cells treated with 5 or 10 ng/mL of sulfasalazine (to inhibit SLC7A11). Gentamicin (30  $\mu$ g/mL) was added to all experimental groups at 1-hour post inoculation, at which point the sulfasalazine or vehicle control (DMSO) were added as appropriate. At 24 hours post inoculation, cells were washed, fixed, and stained with mouse anti- $\alpha$ -tubulin primary antibody, red fluorescent anti-mouse secondary antibody before mounting on slides with DAPI mountant. The genotype/drug treatment details written on each slide were hidden from the microscopist using tape. A reference letter was written on the tape and the 'code' only broken after completion of counting. Flattened/ low light diffracting differentiated cells were chosen for analysis as these show a morphology, which we believed to be closest to macrophages of interest (as previously indicated, Chapter 4). Additionally, the circular adherent U937 cells tend to cluster together, making it difficult to count intracellular bacteria. The number of intracellular bacteria per flattened differentiated cell was assessed by microscopy (as described in Chapter 2) and plotted (Figure 5.19). Flattened CRISPR KO SLC7A11 pool #1 cells and wild type cells treated with 10 ng/mL of sulfasalazine tended to have less intracellular STM than wild type cells treated with DMSO, though this trend failed to reach statistical significance (Figure 5.19A). Intracellular bacterial counts were additionally sorted into frequency bins (Figure

5.19B) to get a better impression of the spread of the data. This data restructuring suggests that inhibition of SLC7A11 either through genetic ablation or pharmacological inhibition increases the proportion of host cells with few or no bacteria. Further replicate experiments are required but were not possible during my studies due to the Covid-19 outbreak. Of particular importance would be to fix cells at 6 rather than 24 hours post-inoculation. This is because we saw much evidence of GFP (bacterial cell) degradation that obscured counting.



Cell line and sulfasalazine (SSZ) concentration



Cell line and sulfasalazine (SSZ) concentration

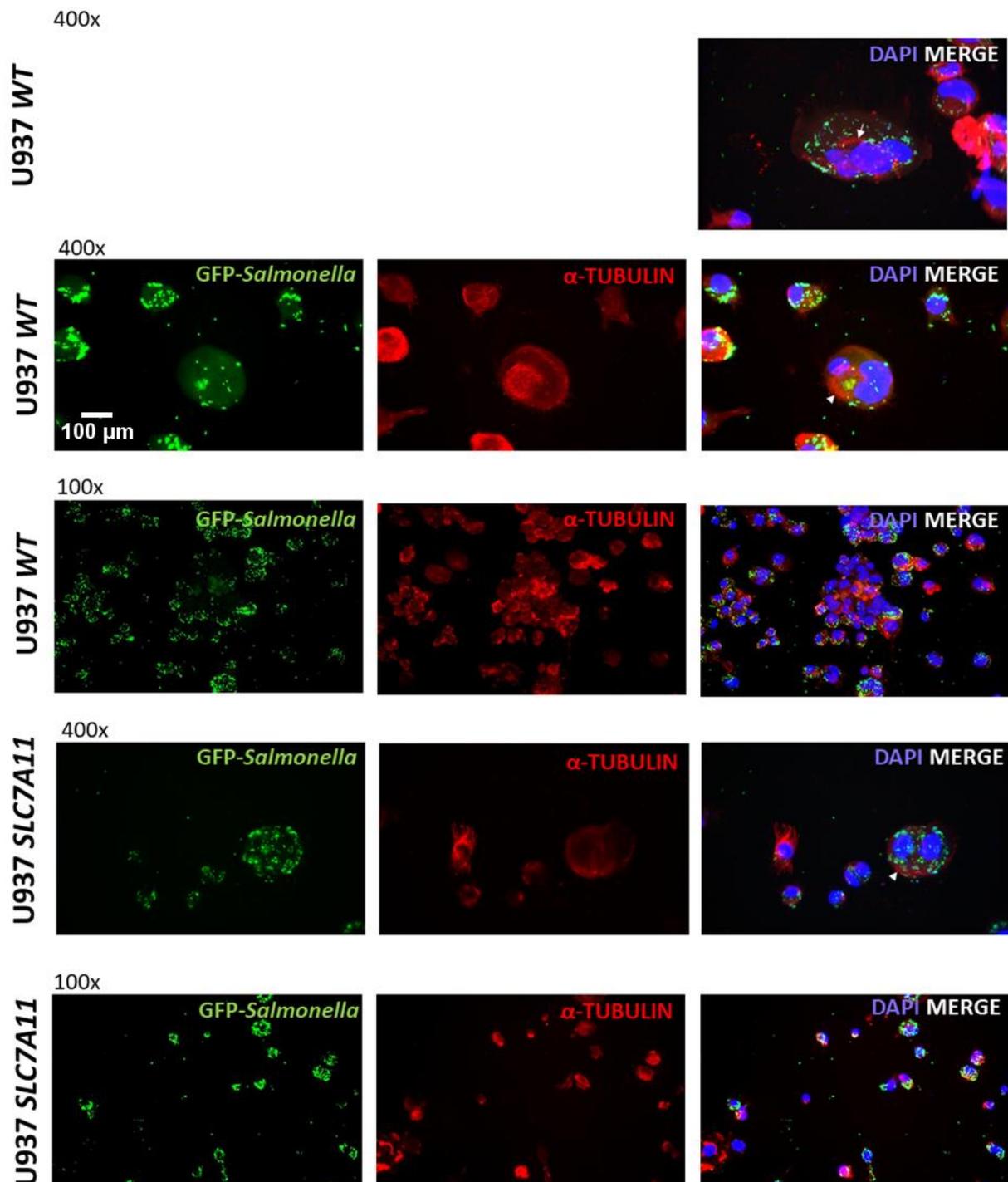
**Figure 5.19 Flattened cocktail differentiated CRISPR KO *SLC7A11* pool #1 U937 cells and wild type cells treated with 10 ng/mL sulfasalazine tend to have fewer intracellular STM than wild type cells.** The number of intracellular bacteria per flattened differentiated cell was enumerated and plotted (A) on a log scale axis, or (B) as a stacked frequency plot. Experiment performed once. (A) Data was statistically analysed with one-way ANOVA, but no statistical difference was found, so post-hoc tests were not performed.

#### 5.4.2.1. Microtubule changes

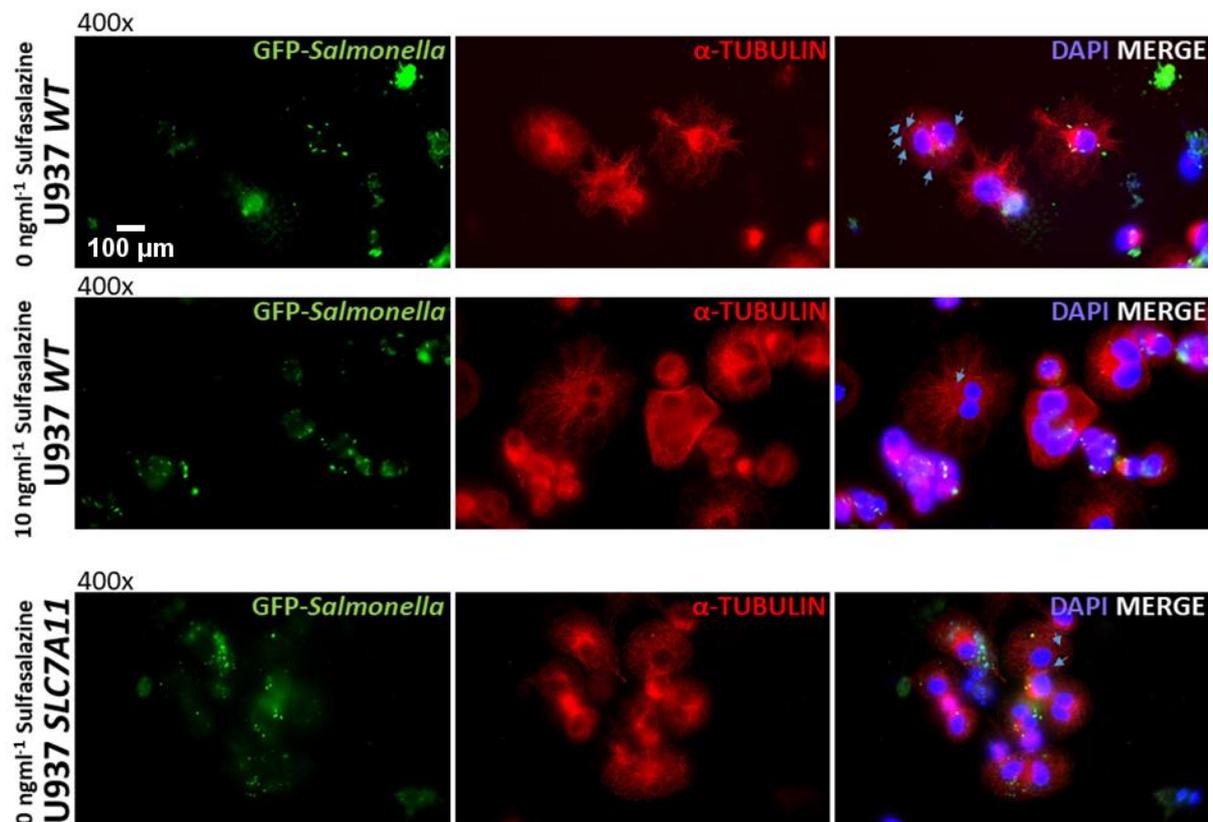
Images taken of cells from the same experiment as Figure 5.19 at 2- and 24- (same time point as that counted) hours post inoculation are also presented (Figure 5.20 and Figure 5.21, respectively). The characteristic web-like  $\alpha$ -tubulin cytoskeletal structures in the large flattened 'pancake-like' U937 cells (like those in Figure 5.21) were not found in any cells at 2-hours post inoculation (Figure 5.20). Some clustering is localised around GFP<sup>+</sup> STM (e.g. arrow heads in Figure 5.20). A single nucleotide polymorphism found in humans increases microtubule instability through increased *TUBB6* expression and consequently promotes resistance to pyroptosis (Salinas et al., 2013). If U937 cells contain this genetic variant (or otherwise have higher expression levels of *TUBB6*), it would explain the (temporary) lack of characteristic  $\alpha$ -tubulin structures and the unusual resistance to *Salmonella*-induced cell death discussed in Chapter 4.

Some cells have long dotted tubulin structures associated with GFP<sup>+</sup> STM at one end (see arrows in Figure 5.20), like those seen in infected NIH3T3 cells (Brawn et al., 2007). These may be connected to (unlabelled) *Salmonella*-induced filaments (aka *Salmonella*-induced tubules), elongated structures associated with microtubules suggested to facilitate nutrient procurement for vacuolar bacteria (Jennings et al., 2017).

At 24 hours post inoculation some cells contain small DAPI-stained rods (arrows)- indicating STM lacking GFP<sup>+</sup> stain. All STM found independent of host cells are GFP<sup>+</sup> throughout the experiment, indicating that the GFP is uniformly expressed within the bacterial population and that it is sufficiently long lasting. GFP fluorescence does not function well at low pH and is susceptible to lysosomal enzymes (Shinoda et al., 2018); suggesting that the DAPI rods are STM within phago- or auto- lysosomes. It was not determined whether any of the microtubule changes described are differentially present in CRISPR KO *SLC7A11* pool U937 cells compared to wild type cells.



**Figure 5.20 STM infection causes microtubule remodelling of differentiated U937 cells at 2-hours post inoculation.** Wild type and CRISPR KO SLC7A11 pool drug cocktail differentiated U937 cells were inoculated with  $4.09 \times 10^7$  GFP expressing STM /well. Cells were washed with PBS, fixed with ice-cold PFA 2 hours-post inoculation; before staining with anti- $\alpha$ -tubulin antibody (red) and DAPI mountant. Images taken with an epifluorescence microscope at a 400x or 100x total magnification. Experiment was performed once.



**Figure 5.21** Example microscopy images to indicate what was seen during data collection at 24 hours post-inoculation for Figure 5.19. Wild type and CRISPR KO *SLC7A11* pool drug cocktail differentiated U937 cells were inoculated with  $4.09 \times 10^7$  GFP expressing STM /well. Sulfasalazine at stated concentrations or DMSO vehicle control was added with 30  $\mu\text{g}/\text{mL}$  gentamicin one hour later. Cells were washed with PBS, fixed with ice-cold PFA 24 hours-post inoculation; before staining with anti- $\alpha$ -tubulin antibody (red) and DAPI mountant. Images taken with an epifluorescence microscope at a 400x total magnification. Experiment was performed once.

## 5.5. Discussion

### 5.5.1. Summary/Overview

This study has taken a hypothesis through a discovery screen, identified differentially represented genes, and began validation with an independent CRISPR KO pool, through to a pharmacological inhibition of the identified protein in an infection setting. An initial validation experiment has been performed using an independent CRISPR KO pool of an example gene *SLC7A11* and a pharmacological inhibitor (sulfasalazine). This experiment was inconclusive but demonstrated a trend towards reduced STM burden with decreased *SLC7A11* function.

The following genes (identified as genes of interest in the STM U937 screen) have associations with infection susceptibility described in the literature, though not much is known about the mechanisms for the associations. *CDH13* (Connell et al., 2013; Hong et al., 2017; Mackinnon et al., 2016), *CLEC19A* (Offenbacher et al., 2016), *VPS13B* (Duplomb et al., 2019), *LGALS12* (Asiamah et al., 2019; Wan et al.,

2016), *SERPINB2* (Darnell et al., 2006; Major et al., 2013), *SORT1* (Gaviria-Agudelo et al., 2014; Ma et al., 2015; Vázquez et al., 2016). Therefore, these previously published associations with infection susceptibility confirm that the STM U937 screen is identifying genes with a genuine connection to infection susceptibility.

Besides this, many genes identified in the STM U937 screen were found with published and inferred links to infection processes and macrophage function. These included M2 polarisation of STM-infected macrophages (*GSK3B*) (Gibbs et al., 2020; Panagi et al., 2020), regulation of inflammasome activation (*CUL1*) (Wan et al., 2019), regulation of oxidative stress (*SLC7A11*) (Cai et al., 2016), regulation of vesicle acidification (*WDR7*) (Kissing et al., 2018; Li et al., 2020), retrograde transport (*SNX6*) (Bonifacino and Hurley, 2008; Shi et al., 2009), regulation of inflammatory cytokines (*MARKS1*) (Lee et al., 2015; Micheva-Viteva et al., 2017) and induction of the stimulator of interferon response cGAMP interactor 1 (*STING1*) immune response (*IL-1B*) (Aarreberg et al., 2019).

In the *F. tularensis* U937 GT screen, many genes were found associated with protein trafficking (*SYTL1* - exocytosis, retrograde transport; *VPS50* - endosome recycling; *RAB33B* - Golgi-associated transport) (Gershlick et al., 2016; Holt et al., 2008; Johnson et al., 2005; Kurz et al., 2016; Ramadass et al., 2019; Schindler et al., 2015; Starr et al., 2010), lysosome-mediated protein degradation (*CTSH*) (Kirschke, 2013) and negative regulation of inflammatory cytokine responses (*TNIP3*, *GSN*) (Cheng et al., 2017; Rath et al., 2018; Wullaert et al., 2007). These general processes make logical sense as steps essential for successful infection/ host cell survival.

Of particular interest, a trapped version of the *SNAP25* gene was statistically overrepresented in the *F. tularensis* U937 screen. *SNAP25* is a known host infection factor for *Salmonella*, as it is important in maintaining the stability of the SCV, and thus, efficient bacterial replication (Stévenin et al., 2019). Unlike *Salmonella*, *Francisella* escapes the phagosome and replicates in the cytosol; so *SNAP25* is likely important at an earlier infection time point. *SNAP25* is a well-known mediator of vesicle fusion with the plasma membrane for exocytosis; however, a study found that its inhibition also prevents the initiation of endocytosis (Xu et al., 2013). This coupling of exo- and endocytosis helps maintain membrane homeostasis, see Wu et al., (2014) for review. The inability of a *SNAP25* gene trap mutant to endocytose may limit *F. tularensis* entry. Indeed, macropinocytosis may be included as part of the entry arsenal of *F. tularensis* LVS (Bradburne et al., 2013). As *SNAP25* is a previously published host infection factor, this demonstrates the efficacy of the *F. tularensis* LVS screen.

### 5.5.2. Involvement of the Electron Transport Chain

The trapped versions of several genes associated with the electron transport chain were identified as overrepresented in the STM U937 screen. These genes were further identified as part of an enriched gene ontology process labelled electron transport within the overrepresented subset of genes from the screen. In a predicted functional protein interaction network these genes had strong evidence scores and pulled out further 'missing links', other mitochondrial ATP synthase subunits. This analysis suggests that a process important to the cellular survival of STM infection has been identified. As stated in Section 5.2.4.1, genetic mutation or inhibition of these genes causes increased mROS production. STM are sensitive to ROS but effectively avoid phagosomal ROS in macrophages. Therefore, mROS production provides an alternative means of macrophage-mediated bactericidal activity. Additionally, reduced levels of complex-I as a result of GT mutagenesis would promote electron transfer to ubiquinone through complex-II, producing a suitable environment for an M1 polarised dysfunctional TCA cycle (as described in Chapter 1).

STM grow faster in macrophages metabolising fatty acids (Eisele et al., 2013) (which requires a functional TCA cycle and electron transport chain) with an M2(IL-4) like phenotype (Lathrop et al., 2015). The reduction in functional Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase, Mitochondrially Encoded Cytochrome C Oxidase I or Mitochondrial ATP synthase inhibits this metabolic activity causing reduced bacterial growth even without the additional antibacterial activity. mROS can stimulate host cell death pathways (Platnich et al., 2018; Y. Wang et al., 2019), so the survival advantage garnered by these GT mutations must be sufficient to overcome the potential background cell death caused by aberrant mROS.

### 5.5.3. Novel gene associations with STM and or *F. tularensis* infection

The two sets of genes identified as differentially represented in the STM and LVS U937 GT library screens do not directly overlap, as was originally anticipated – in order to identify ‘universal’ protective factors. However, some protein interactors create links between the screens. Protein network analysis has predicted inter- and intra-screen protein interactions or functional associations based on high throughput protein interaction studies (CPDB), low throughput protein interaction studies, databases, co-expression data and PubMed text (STRING, all). APP, one of the ‘missing link’ network-hubs identified by CPDB connected 16 query nodes (including proteins from both screens). There are many potential reasons for the lack of an important host resistance/ susceptibility gene in the original screen- lack of introns, essential gene function (producing a lethal phenotype), or the scope of the mutant library. APP contains many introns (found on UCSC genome browser), knock out mice are available (showing that an APP KO is not lethal) and the gene was found trapped in the U937 GT library, though it was lowly represented. Factors associated with extremely low gene representation (bioinformatics artefact, random loss or gain of representation due to stochastic noise) were previously mentioned in the results. Additionally, a caveat of molecular network analysis is that any given interaction found may simply not be relevant in the context studied; nevertheless, this type of analysis provides the opportunity to increase the discovery space or to focus on particularly important query genes.

CPDB and STRING predicted seemingly incongruous sets of protein networks from the same query set and were only able to replicate the other network in a limited fashion, or not at all. To discover why these network databases produce completely distinct outputs, the sources and methods were compared. Briefly, STRING derives physical protein interaction data from a reduced selection of databases compared to CPDB - which includes primary interaction databases such as the Comprehensive Resource of Mammalian Protein Complexes (CORUM) (Giurgiu et al., 2019; Herwig et al., 2016). Additionally, the network databases integrate non-query interacting proteins in different ways, which can be observed by adjusting the non-query interacting protein limit. There are many other differences besides; for example, STRING stores data on thousands of organisms and incorporates interactions between predicted orthologues into the query protein network (Szkłarczyk et al., 2019), while CPDB only houses data on 3 organisms (Herwig et al., 2016). The differences in method design are the result of differing intentions for the functions of the network databases and the outputs should be viewed in this light. Other authors using the induced network module of CPDB to analyse molecular networks between differentially expressed genes (the query nodes) and intermediate nodes were able to find intermediate node ‘hubs’ – nodes with many query node interactions (Huang et al., 2013; C. Li et al., 2018). In contrast, authors looking to narrow their focus on important genes within large differentially expressed gene sets used STRING to identify ‘hub’ genes within PPI networks (Chen et al., 2019; Ke et al., 2019; Zhou et al., 2019). Thus, combining the interactions found from CPDB and STRING may identify additional genes of interest for further study.

#### 5.5.4. Lack of overlap with published *Salmonella* infection screens?

Systematic interrogation of host responses to bacterial infection is a relatively understudied area of research. Since RNAi technology became available many studies targeting host infection factors were published before the extent of technical caveats were fully understood and adjusted for. Five genome wide host factor studies of *Salmonella* infection were found in the literature, of these, three were RNAi screens performed prior to an understanding of the impact of RNAi sequences on the microRNA space (Misselwitz et al., 2011; Thornbrough et al., 2016, 2012). A fourth screen used RNAi with at least four oligos per gene (which can be used to average the phenotype), but it is unclear whether other requisite measures have been applied to ensure the sources of false positives are managed (Andritschke et al., 2016). Additionally, these screens were performed in non-professional immune cells (e.g. MCF-7, HEK-293T). Together these factors are likely the biggest contributors to the lack of overlap between these screens and the STM U937 screen.

Finally, the fifth screen was performed in CRISPR-sgRNA lentiviral transduced differentiated THP-1 cells (a human monocyte-like cell line) and selected for host cells with low STM burden after 30 minutes of infection (Yeung et al., 2019). They found a lot of genes involved in phagocytosis and plasma membrane PRR responses, for example, FCGR1A and TLR2. The difference in infection time (30 minutes compared to ~48 hours) is likely the biggest contributor to the differences between the Yeung et al., (2019) screen and the STM U937 screen. The low scope of the differentiated GT library likely contributed in small part to the lack of overlap between these screens.

A previously identified *Salmonella* infection susceptibility gene Squalene epoxidase (*SQLE*) was identified in the STM U937 screen. *SQLE* catalyses a rate limiting step of the cholesterol biosynthesis pathway, yet in contrast to the literature described (Yu et al., 2009), the trapped version of the gene encoding it was underrepresented in STM-infected U937 cells; indicating susceptibility to STM infection. Lipid rafts, (of which cholesterol is a key component), are important for organisation of host membrane protein signalling organisation, for a review see Santos and Preta (2018). Perhaps host-beneficial signalling through a plasma membrane protein during STM macrophage infection is disrupted in a diminished cholesterol environment, generating a susceptible phenotype.

#### 5.5.5. Comparison to other infection or immune response-based genetic screens:

Host infection screens utilising bacteria other than *Salmonella* as an agent of selection have also been used for comparison to both U937 screens. Additionally, a human genetic association study using naturally diverse isolated primary cells to interrogate the human genome has been included in the list of studies compared to the U937 screens. Compared studies sharing gene hits with the screens described in here are listed in Table 5.6.

**Table 5.6 Comparison of STM and *F. tularensis* U937 GT screens with other infection or immune response based genetic screens. Genome wide or high throughput screens with hits matching those found in the STM and *F. tularensis* U937 GT screens. Only RNAi screens deemed to have sufficient controls are included.**

Human Gene	Mutation	Study type	Screen	Phenotype	Cell	Paper	Screen comparison
<i>SERPINB2</i>	SNP allele	Genome-wide association study	HapMap host derived cells	STM-induced host cell death	lymphoblastoid cells	(Ko et al., 2009)	GT version overrepresented in STM-infected cells
<i>PAQR5</i>	CRISPR KO	Genome wide	LPS and cholera toxin B-subunit treatment	Host cell survival	RAW264.7 cells	(Napier et al., 2016)	GT version underrepresented in STM-infected cells
<i>FERMT3</i>	RNAi knockdown	Genome wide, high throughput validation	<i>Brucella abortus</i> infection	Increased infection (GFP bacteria measured using automated high throughput imaging analysis)	HeLa cells	(Casanova et al., 2019)	GT version overrepresented in STM-infected cells
<i>COQ2</i>	RNAi knockdown	High throughput, high throughput validation using additional siRNAs	<i>Yersinia pestis</i> infection	Reduced bacterial growth (bioluminescence)	RAW264.7 cells	(Connor et al., 2018)	GT was overrepresented in STM-infected cells
<i>PKC-η</i>	RNAi knockdown	Kinome, high throughput and low throughput validation using alternative genetic approaches and measurements	<i>B. thailandensis</i> infection	Reduced bacterial growth via loss of MARCKS phosphorylation	THP-1 cells	(Micheva-Viteva et al., 2017)	GT version of <i>MARCKS</i> overrepresented in STM-infected cells

### 5.5.6. Why does the screen have a comparatively low gene coverage?

The number of protein coding genes with counts higher than 5 in at least 2 samples in the differentiated GT library was much smaller (0.04-1.8% of the protein coding genome) than the scope originally estimated. Another pooled high throughput (CRISPR) KO screen in U937 cells using RNA-Seq to measure sgRNA abundancies reported as genome wide had 16,700 genes with at least 1 gene count in 1 sample; of those, 8,200 genes had at least 5 counts in 1 sample (Haney et al., 2018). Studies using RNA-Seq to measure differential gene expression have similar patterns of highly and lowly represented genes; the latter becoming more prominent with increasing sequencing depth. The majority of lowly represented transcripts found in the STM and *F. tularensis* U937 GT screens are likely to be genuinely present in the sequenced samples (rather than bioinformatics artefacts).

### 5.5.7. Technical critique of U937 screen

Some of the trapped genes identified in the STM U937 screen are putative regulators of cell proliferation or regulators of cell death (*SHC4*) (Ahmed et al., 2019; Turco et al., 2012). Overrepresented trapped genes associated with cell death (or its positive regulation) were also identified in the LVS screen (*RYBP*) (Zhan et al., 2018). The presence of these confounding genes are likely due to the nature of the screen.

As the analysis of this screen incorporated methods not previously used for random GT library screening, some optimisation was needed. In addition to method optimisation, time to familiarise oneself with the analysis programmes was required, therefore there are many areas for potential improvement and these are described.

The most common (and vastly cheaper) method of identifying gene trap mutations after random mutagenesis is through clonal cell isolation before bacterial cloning and Sanger sequencing. There appears to be only one other published paper documenting the use of RNA-Seq for gene trap identification after random library mutagenesis (the paper described an unrelated pharmacological screen) (Mayor-Ruiz et al., 2017). Mayor-Ruiz et al., (2017) isolated surviving clonal host cells on which to perform RNA-Seq to elucidate the transcriptomic changes in response to gene traps identified through parallel post-sequencing analysis. Therefore, this is the first study to use RNA-Seq to delineate gene trap identities and proportions from pooled cell mutants. The following sections describe critiques and potential improvements that could be implemented as future work to the post-sequencing analysis. The mentioned programmes, their uses and availability are summarised in Table 5.7.

#### 5.5.7.1. Trimming

If time had allowed, the bioinformatics analysis might have benefited from a more thorough investigation and optimisation of the available trimming options, as described in Williams et al., (2016). Many of the available trimming programmes have been compared using a variety of parameters, but most authors agree that trimming should be optimised for the specific dataset (Fabbro et al., 2013; MacManes, 2014; Payá-Milans et al., 2018). Furthermore, one paper claims no trimming may be better (dependent on the dataset and aligner) (Liao and Shi, 2019). Potential benefits include greater retention of reads, greater mappability and more concordant alignments (though not necessarily with the same settings).

### 5.5.7.2. Genome Alignment

A comparison of genome alignment programmes of differential gene expression data from formalin-fixed, paraffin-embedded (FFPE) cancer tissue found that HISAT2 aligned reads to more pseudogenes than (RNA) STAR, including genes identified as the top 50 hits (Raplee et al., 2019). It is unclear whether these findings would apply to the data analysed here, as the authors do not provide quality assessment data but state that the sample type is prone to RNA degradation. Indeed, other comparative analyses of RNA quality from FFPE tissue suggest that the sequencing data is likely to be error prone, resulting in more complex alignment programme requirements (Esteve-Codina et al., 2017; Wehmas et al., 2019). Within the list of total trapped genes from the STM U937 GT screen, pseudogenes accounted for 1.2% of ~15,500 genes but only contributed an average of 0.2% of gene counts to total gene counts per sample. Additionally, four pseudogenes contributed to the differentially represented gene list at an FDR threshold of 0.3. It would be of interest to compare the alignment results of HISAT2 and RNA STAR to identify if one produces better results with these data.

### 5.5.7.3. *De novo* vs reference-based assembly vs an alternative

A different (and overlapping) set of differentially represented genes was identified in two methods of differential gene representation analysis of the STM U937 gene trap screen; the result of two major sources of difference in the method. Firstly, *de novo* transcriptome assembly is a difficult process and the transcript assembly is normally less accurate as a result (Conesa et al., 2016; Pertea et al., 2015; Smith-Unna et al., 2016). Secondly, the nature of transcripts derived from trapped genes will generate many re-assembled transcripts that do not fit the mould provided by an official reference transcriptome.

Two alternative methods that might improve the accuracy of the differential representation analysis may be possible. Firstly, use of long read sequencing technologies, e.g. that provided by Pacific Bioscience or Oxford Nanopore, which can produce read lengths of ~ 15 kilobases or ≥ 30 kilobases, respectively, would enable complete sequencing of the majority of transcript lengths. A big disadvantage of these technologies is that they are low throughput, causing sequencing of a similar depth and sample number to be prohibitively expensive (Byrne et al., 2019; PacBio, 2020). Secondly, read pairs of which one read contains a partial GT vector exon could be solely used for gapped genome alignment, which would bypass the requirement to identify transcripts. Each read pair containing a GT vector exon is equal to a single replicon of a GT transcript. Reads derived from fragments containing only a small portion of human mRNA would likely not map uniquely, neither would reads mapping to repetitive regions of the genome, or those matching high similarity homologues (e.g. *FCGR2A* and *FCGR2C*). These factors would likely result in a smaller number of genes reported overall but would likely result in a more accurate gene count and therefore, more accurate differential representation. This method would however require more sophisticated script writing skills.

### 5.5.7.4. *F. tularensis* U937 screen: low statistical power

The *F. tularensis* U937 GT screen differential representation analysis suffers from a lack of biological replicates (3 replicates were carried through the analysis compared to 6 for the STM U937 GT screen); as the paired samples from experiment A were not included. An alternative comparison of the gene count results between samples to that implemented here (such as a principal component plot) might provide a better overall impression of how different the counts from Experiment A actually are. If the overall impression is that the samples from each experiment (e.g. infected vs mock-infected) are equally different, this would strongly suggest that Experiment A samples are appropriate to include in the final analysis, thus improving the statistical power.

#### 5.5.7.5. Repeated or independent measures designs for statistical analysis?

A repeated measures design for statistical analysis may fit the data better than the statistical analysis (independent measures) performed with limma voom for both screens. The idea is to cut out the inter-experimental variation, which would potentially make the analysis more sensitive to subtler differences. It is possible to perform such an analysis within limma voom (Hoffman and Roussos, 2020; Ritchie et al., 2015), but it requires use of the command line.

**Table 5.7 Alternative or additional programmes suggested as options to improve upon the current analyses.** *The described use is the purpose to which the programme would be put for alternative or future analysis of the presented datasets.*

Programme	Use	Availability
sortmeRNA	identify and remove (if present) ribosomal and mitochondrial RNA	Galaxy Europe
FastQ screen	identify and remove (if present) sequences matching the Salmonella genome	Downloadable, requires Linux, runs Perl script
Exploration of available trimming programmes	To optimise trimming prior to alignment	Preferably those available on a public Galaxy instance
RNA STAR	Alternative read-to-genome aligner	Galaxy Main/ Europe
Principal component analysis programme	Determine how similar Experiment A is from the other samples in the <i>F. tularensis</i> U937 GT screen	Preferably those available on a public Galaxy instance
Dream	Repeated measures design of limma	R command line/ Bioconductor

#### 5.5.8. Future work

##### 5.5.8.1. Experimental

The host survival phenotype of all CRISPR-generated mutants of interest should be assessed using LDH activity assay at multiple time points. The LDH activity in uninfected (as well as infected) wild type cells should be measured to incorporate any potential background cell death.

SLC7A11 was previously shown to limit macrophage ROS activity in *M. tuberculosis* infection and *slc7a11*<sup>-/-</sup> mice had reduced intracellular *M. tuberculosis* replication than wild type counterparts (Cai et al., 2016). However, the role of SLC7A11 has not yet been explored in STM infection. Under the conditions used here, the differences in bacterial burden observed between differentiated wild type and CRISPR KO *SLC7A11* U937 cells did not reach significance. The MOI used was roughly an order of magnitude lower and the cells infected for half the time used in the screen. Nevertheless, based on previous observations a six-hour post inoculation time point is estimated to show a distinguishing phenotype. To complete this series of experiments the bacterial burden within drug cocktail differentiated small circular adherent U937 cells should be compared to that in the larger flattened cells found in the same differentiation treatment.

A recent paper reports the ability to efficiently knock out genes in human MDMs using CRISPR Cas9 (Freund et al., 2020), therefore, this protocol could be applied to specific genes identified in this work, for example *SLC7A11*.

The potential of APP as a target could also be investigated, beginning with KO host cells *in vitro* – and go on to use KO APP mice for *in vivo* infection experiments if successful.

#### 5.5.8.2. Computational

A principal component analysis plot on the *F. tularensis* U937 screen samples (including Experiment A samples) should be performed to identify how related the different samples are to each other. This would provide a better indication of the difference (if any) between the passage control samples and the samples included in the current analysis. If this more holistic comparison demonstrates that the patterns in Experiment A are comparable to the other experiment samples, a differential representation analysis including all the experiments for the *F. tularensis* U937 screen could be implemented.

The alternative bioinformatics processes described in sections 5.5.7.1, 5.5.7.2, 5.5.7.4, 5.5.7.5 and 5.5.7.3 should be performed and compared to the results described here.

A gene ontology analysis of differentially represented genes (data not shown) identified between the original library (suspension) and uninfected differentiated U937 GT cells should be performed to determine the processes potentially important to differentiation under the protocol used. Processes identified as having enriched genes may indicate the polarity of the macrophage phenotype produced. This would indicate the initial phenotype of the macrophage model, which might help comparisons with the *Salmonella*-macrophage infection knowledge base.

#### 5.5.9. Conclusion

The two U937 screens do not appear to have any direct overlap, but they share general themes. Broadly similar processes were identified, such as protein trafficking and immune response regulation. These are common intracellular infection requirements for host defence and pathogen survival, though the specific proteins appear to be pathogen specific. The STM screen overlaps to a greater degree with other infection literature, which is expected due to the larger list of differentially represented genes as well as the fact that it is a better studied pathogen.

# Chapter 6

## Discussion

## 6. DISCUSSION

### 6.1. Summary

The original objective of this work was to identify innate responses to intracellular pathogens or host processes essential for pathogen survival and replication. Ultimately, it was envisioned that these would represent targets that could be modulated, without detriment to the host, by host-directed therapeutics. Such therapeutics would offer an alternative route to the current, failing antibiotic strategies.

The genomic space of two human cell lines (HEK-293 cells and differentiated U937 cells as examples of 'non-professional immune cells,' or 'professional immune cells', respectively) were interrogated through random mutagenesis. The intention was to identify host factors common to many, if not all, bacterial infections; so, two different bacterial pathogens were chosen as the selective pressure in two respective screens.

The infection phenotypes of resistance mutations previously identified in host susceptibility genes in the HEK-293 screen were investigated in independent CRISPR KO cell lines generated for the purpose (Chapter 3). Bacterial burden assays demonstrated the presence of host cell resistance to STM and *S. sonnei* infection in the CRISPR KO *PCDH7* cell line.

Subsequently, a differentiation protocol was created to generate U937 cells with macrophage-like morphology, which possessed the ability to phagocytose bacteria and expressed the general differentiation marker CD11b. A U937 gene trap (GT) library was created and screened with two important pathogens, one of global medical significance (*S. enterica* ser. Typhimurium), and the other of importance as a potential bioterror threat (*F. tularensis*) (Chapter 4). The sequencing output from each screen was bioinformatically processed, statistically analysed and examined with pathway analysis software (Chapter 5). These U937 infection screens identified a multitude of genes as potential infection resistance or susceptibility factors. Of these, a number have established support in the literature for roles in promoting resistance or susceptibility to infection. Many genes were also isolated that have no previous association with infection. A potential protein interaction network was identified from overrepresented trapped genes from the STM U937 screen that might help to define important host processes during infection. Finally, a U937 CRISPR KO pool for the representative candidate gene *SLC7A11* was generated and characterised before its use in an initial pilot infection phenotype experiment (Chapter 5). This preliminary validation of *SLC7A11* as a host susceptibility factor fell short of generating statistically significant results but offers insight into a strategy to expand the validation of these new candidates.

## 6.2. Assessing the success of the functional screening approach through the biological actions of the identified host processes

The STM and *F. tularensis* U937 GT screens identified many genes. *SLC7A11* and *SERPINB2* are two examples that partially fulfil the criteria set out in the introduction as potential inhibitor drug targets in that, when mutated, both appear to improve host cell survival during STM infection. All that needs to be confirmed in order to completely fulfil the criteria is a validated connection between mutation of these genes and a reduction in bacterial burden. Knocking out *SLC7A11* increases antimycobacterial activity in macrophages (Cai et al., 2016), though the activity against STM is yet to be confirmed. Wild type *SERPINB2* may indirectly promote STM proliferation by creating a more suitable intracellular environment (Pham et al., 2020; Saliba et al., 2016; Schroder et al., 2019a; Stapels et al., 2018; Zhao et al., 2013).

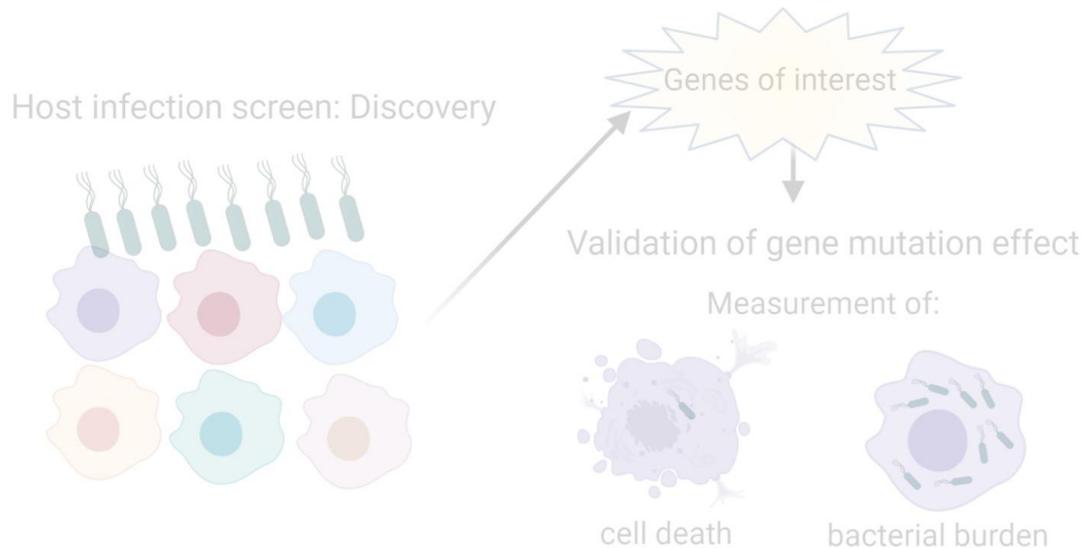
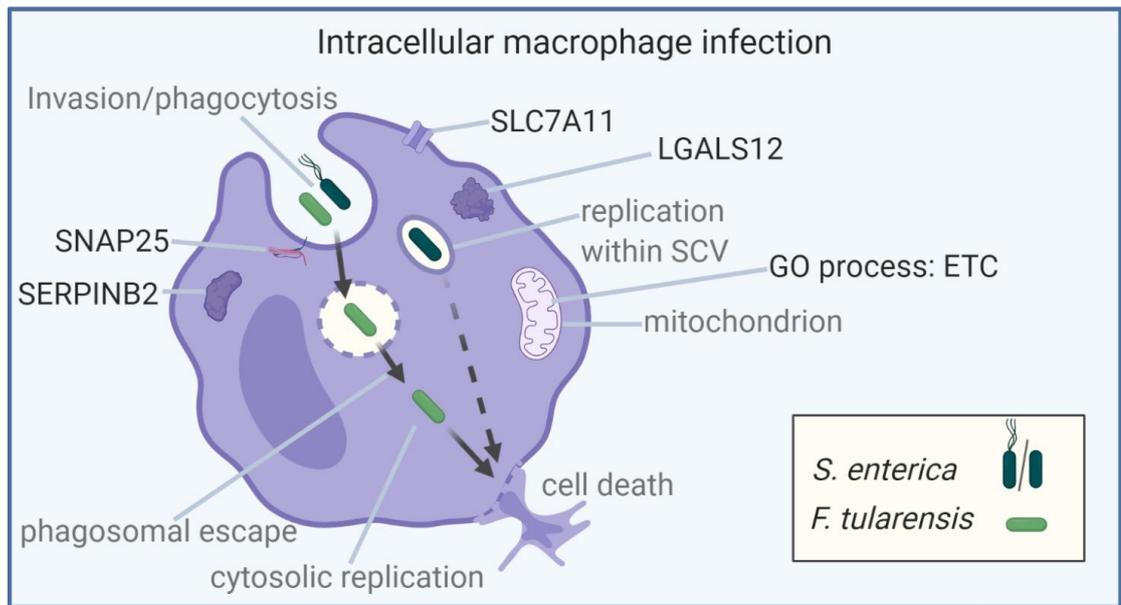
Several of the other identified genes (*CDH13*, *CLEC19A*, *VPS13B*, *LGALS12*, and *SORT1*) identified in the STM screen are already associated with infection susceptibility (Asiamah et al., 2019; Connell et al., 2013; Duplomb et al., 2019; Gaviria-Agudelo et al., 2014; Hong et al., 2017; Ma et al., 2015; Mackinnon et al., 2016; Offenbacher et al., 2016; Vázquez et al., 2016; Wan et al., 2016). The *F. tularensis* screen identified *SNAP25*, a previously identified host infection factor in STM infection of human epithelial cell lines (Stévenin et al., 2019); as well as other gene products involved in lysosome-mediated protein degradation – *CTSH*, (Kirschke, 2013) – and regulation of inflammatory cytokine responses – *TNIP3* (Rath et al., 2018; Wullaert et al., 2007) and *GSN* (Cheng et al., 2017). These connections with relevant examples from the literature suggest that the screens were successful in identifying genuine host infection factors.

The GO processes ‘electron transport chain’, ‘response to oxidative stress’, GO component ‘organelle inner membrane’ and the STRING protein functional interaction network all indicated an overlapping cluster of genes or gene products identified as overrepresented (protective when trapped) in the STM U937 screen. The overlapping trapped genes were Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Core- Subunit 5 (*MT-ND5*), Subunit 6 (*MT-ND6*) (subunits of complex I of the electron transport chain, ETC) and Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1*) (subunit of complex IV of the ETC). The effect of trapping or inhibiting these genes or gene products will be discussed in more detail in Section 6.3.2. Thus, a pathway promoting oxidative stress was identified as a credible host cell infection response important for survival during STM infection. This is a promising result, as too much oxidative stress would typically be thought to induce host cell death, therefore, an optimal balance between bactericidal action and host cell death must be maintained. Indeed, a small amount of oxidative stress for multicellular eukaryotes during development appears to be necessary for developing stress resistance and increasing life span (Bazopoulou et al., 2019).

Both the STM HEK-293 and U937 host cell survival screens also identified genes encoding proteins in the cadherin superfamily (*PCDH7* – HEK-293 screen, *CDH13* – U937 screen), which aligns with established ideas of host cell entry of bacteria via adhesion factors

(Everman et al., 2019; Kochi et al., 2019; Li et al., 2016; Pizarro-Cerdá and Cossart, 2006; Sousa et al., 2007; Watters and Palmenberg, 2018). *PCDH7* as a potential host infection factor is discussed further in Sections 6.3 and 6.5.2.

In summary, many promising potential host infection factors and pathways were identified, some of which may be targetable for the treatment of bacterial infections. Thus, the primary aim of this work has been achieved. Figure 6.1 reworks the original image presented in the Introduction Chapter in the light of our screen results – there are several overlapping aspects of pathogen infection and host response that have been exposed by the screens detailed within this thesis.



**Figure 6.1 Most promising potential host infection factors identified through genetic macrophage screens utilising two evolutionarily distant bacterial pathogens.** *SLC7A11*, *SERPINB2*, *LGALS12* and genes sharing the gene ontology (GO) process 'electron transport chain' (ETC) were identified as potential host infection factors in the *S. enterica* ser. Typhimurium (STM) U937 genetic screen. *SNAP25* was identified as a potential host infection factor in the *F. tularensis* LVS U937 genetic screen. *S. enterica*, and *F. tularensis* each have distinct molecular mechanisms of achieving entry, survival and replication within macrophages. Two macrophage host infection screens were employed to identify host infection factors that could be pharmacologically modulated to reduce bacterial burden, thus, improving morbidity.

### 6.3. A universal infection drug target – does it exist?

An initial goal of this work was to identify an infection, or host response, drug target common to a broad range of intracellular bacterial infections. One instance where that could be the case is human *PCDH7*, which when knocked out, promotes resistance to infection with either of two bacteria (*S. sonnei* and STM) belonging to different taxonomic genera (Chapter 3). Other potential host infection factors shared between the STM U937 screen and more distantly related pathogens identified in the literature (such as *Yersinia pestis*, in the same order, or *Brucella abortus* in the same phylum) include *COQ2* and *FERMT3* (Casanova et al., 2019; Connor et al., 2018). The quality of the data for the *FERMT3* connection is relatively low, as *FERMT3* was number 66 in the STM screen and number 292 in the screen by Casanova et al., (2019). The quality of the *COQ2* association with reduced *Y. pestis* growth was also relatively low, as the number of independent siRNAs used to silence it were effectively 3. Nevertheless, these comparisons indicate that host infection factors may span across different families, or even different classes, of bacteria.

#### 6.3.1. A gene-centric approach

Two screens were performed in the U937 GT library created in Chapter 4, using two genetically dissimilar bacterial pathogens (from different taxonomic orders, Thiotrichales and Enterobacteriales). No shared differentially represented genes were identified between these screens despite a potential 13,721 mutated transcribed sequences upon which selection took place and the respective bacteria appearing to be more evolutionarily related than *B. abortus* is to either. A total of 136 differentially represented trapped genes were identified between the two screens (Chapter 5), though the majority of these genes were identified within the STM screen. The lack of identified shared host infection factors between STM and *F. tularensis* may result from a combination of technical limitations (such as the limited genetic scope), and the dissimilarity of survival strategies between the two bacteria. This latter issue would reduce the number of common host infection factors, whereas the former reduces the chances of finding them.

#### 6.3.2. A pathway-centric approach

Individual genes required for host survival against many evolutionarily distinct bacterial pathogens are difficult to find. Therefore, a host pathway common to many infections may instead be a better target. Intracellular pathogens have to survive many of the same host defences, so they often converge on similar evolutionary solutions, though they may not interact with the same protein. This suggests a host infection network that is common to many distantly related intracellular bacterial pathogens. Many authors are discussing the concept of testing the effect of drugs on the interactome (Aulner et al., 2019; Bang et al., 2019; Chiang et al., 2018; Fotis et al., 2018; H. Li et al., 2012) rather than on *in vitro* protein activity as has been done in the past. An example of a host infection response network, identified in the STM U937 screen, is the generation (or retention) of mROS, as a likely mechanism of controlling bacterial burden. As described in Section 6.2, genes encoding subunits of ETC complex I and IV were indicated as part of this host infection response network. Reducing the number of electrons transported through complex I and complex IV

(e.g. through genetic ablation of one allele of a subunit) causes excess mitochondrial ROS generation (Chouchani et al., 2014; Guarás et al., 2016; Jin et al., 2014; Lapuente-Brun et al., 2013; Lopez-Fabuel et al., 2016), which is either directly, or indirectly, antibacterial (Roca and Ramakrishnan, 2013; West et al., 2011). Reducing the activity of other candidate proteins (*PINK1*, *SLC7A11*, *TXNRD1*) within each of these categories is also known to contribute to greater mROS production (Morais et al., 2014; Nalvarte et al., 2004).

Patients lacking subunits of phagosomal NADPH oxidase (required to generate ROS within the phagosome) are susceptible to infections with opportunistic bacteria - review - Roos and Boer (2014) therefore inherent ROS-mediated defences are normally sufficient to clear such bacteria. Intracellular pathogens adapted to human physiology either evade ROS, are highly resistant to ROS, or some degree of both (e.g. *Burkholderia pseudomallei*, *Mycobacterium tuberculosis*) (Burtneck et al., 2008; Köster et al., 2017; Loprasert et al., 2003; Mehta and Singh, 2019). *M. tuberculosis* classified within the Actinobacteriota phylum is still susceptible to ROS inside host cells if exposed to a sufficient concentration (Cai et al., 2016; Roca and Ramakrishnan, 2013). Therefore, evolutionarily distant bacteria may be targeted through the generation of ROS.

### 6.3.3. Exploration of one such pathway; different survival strategies to intraphagosomal ROS

Given the emergence of ROS as a common denominator of many of the genes identified in the STM U937 screen, it is important to explore how bacteria respond to this intracellular environmental change. STM and *F. tularensis* LVS (the strain used in the *F. tularensis* U937 screen) may have different sensitivities to oxidative stress, which are described in some detail below.

The lack of the phagosomal NADPH oxidase does not appear to affect *F. tularensis* LVS growth *in vivo* or *in vitro* (Ma et al., 2016; Zhuo Ma et al., 2019), though this may be as much the result of efficient escape from the phagosome as resistance to oxidative stress. *In vitro* evidence produced using strong oxidising agents on agar indicate that ROS reduce *F. tularensis* LVS growth if given at sufficient concentrations (Alqahtani et al., 2018; Ma et al., 2016; Zhuo Ma et al., 2019); though it is unclear whether such conditions are attainable within host cells.

*Salmonella* also has some resistance to oxidative environments resulting from ROS (Bogomolnaya et al., 2013; Fu et al., 2017), but can still be affected at higher concentrations (West et al., 2011). As part of its intracellular oxidative stress survival strategy, *Salmonella* delays the co-localisation of phagosomal NADPH oxidase with the SCV in DCs (Gogoi et al., 2018) and macrophages (Heijden et al., 2015; Suvarnapunya and Stein, 2005; Vazquez-Torres et al., 2000). This allows replication inside the SCV without substantial exposure to ROS induced oxidative stress. However, a small percentage of wild type *Salmonella* exit the SCV (Perrin et al., 2004) and encounter the oxidative environment of the cytosol (Heijden et al., 2015; West et al., 2011). Therefore, if screen mutants accentuated the cytosolic oxidative environment, it would be expected to improve the host anti-*Salmonella* response. Indeed, a group of trapped genes involved in the host oxidative stress response were overrepresented during STM infection, indicating that U937 cells lacking these oxidative

stress modulating enzymes or enzyme subunits have a survival advantage. None of these genes were identified in the *F. tularensis* LVS screen, perhaps either due to a difference in sensitivity to ROS, or the reduced number of replicates in the screen.

#### 6.3.4. Differences in pathogen responses to differing intracellular environments within professional and non-professional phagocytic cells

There are several potential reasons for the lack of shared genes between the HEK-293 and U937 STM screens, indeed, genes relating to professional phagocytic or macrophage specific responses were expected to contribute to the distinct macrophage gene set. Additionally, STM can reside within various host cell types in different organelle contexts (Castanheira and García-del Portillo, 2017) – for example free cytosolic bacteria are found within epithelial cells (Knodler, 2015; Laughlin et al., 2014) and macrophages (Meunier and Broz, 2015; Thurston et al., 2016) but not in fibroblasts (López-Montero et al., 2016). As an uncharacterised cell line (HEK-293) was used in the initial screens, it is unclear whether free cytosolic bacteria would feature as part of the intracellular bacterial population.

#### 6.3.5. Stochastic variation in the screening process – detecting the signal through the noise

This is a technology and assay dependent project, as such, the results are somewhat dependent on the methods used. Many aspects of this project operated at the interface between the limits of sensitivity of the experimental approaches and the complexities of two-species biological processes. For example, stochastic variation in U937 cell proliferation, differentiation and cell death responses will have affected the trapped gene representation within individual biological replicates. Stochastic variation in bacterial gene expression and replication would also have been present between individual bacteria within a single bacterial culture, which would also have affected the trapped gene representation in the U937 host cells. Evidence of variation is clear between biological replicates, but the number of replicates in the STM U937 GT screen provides sufficient statistical power to discern signals within the data. To confirm the stringency of the selection pressure, iterative rounds of selection with the same host cells would have been ideal – as this would have provided multiple opportunities to sample the surviving host population. However, the use of differentiated U937 cells precluded the opportunity to perform multiple successive rounds of infection selection (discussed in Chapter 4).

#### 6.3.6. Lack of shared host infection factors between HEK-293 and U937 cells – what technical limitations were at play?

The sequencing of clones in the initial HEK-293 survival screen reduced the potential to identify genes common to both survival screens in two ways. Firstly, the screen generated a comparatively low number of genomic regions of interest – 19 HEK-293 GT colonies were isolated, from which amplified GT transcripts were sequenced. Of those, 25% matched multiple regions of the genome and another 25% matched lncRNAs or were left unannotated. The remaining 9 sequences matched genes, which therefore, could be

compared with the U937 screen. Secondly, sequencing of clones precludes the identification of trapped genes, which confer a reduced survival advantage, as the absence of these in the isolated GT population cannot be quantified. In contrast, the STM U937 screen identified 50 significantly underrepresented genes (thought to confer a reduced survival advantage) through the use of pooled sequencing.

The U937 STM screen required a different MOI and infection end point (to the HEK-293 screen) to generate a suitable stringency of selection. The differences in experimental condition likely led to disparate host cell death resistance requirements.

The U937 GT library used in the survival screen was made from two sub-libraries, each of which had 1,000-10,000 mutational events (Chapter 4), at best this could have meant on average a single mutation for each protein coding gene in the human genome. However, only 15,449 open reading frames were identified from the STM U937 screen (Chapter 5). Of these 11,905 were protein coding genes, antisense transcripts, or transcript readthroughs (a coverage of 58-60% of protein coding genes). As noted in Chapter 5, many of these protein coding genes were represented at extremely low levels, such that 373 protein coding genes (0.04-1.8% of the protein coding genome) had gene counts of more than 5 in at least two samples. The small number of trapped protein coding genes well represented within the U937 GT library likely had a considerable contribution to the lack of genes of interest in common with the HEK-293 screen.

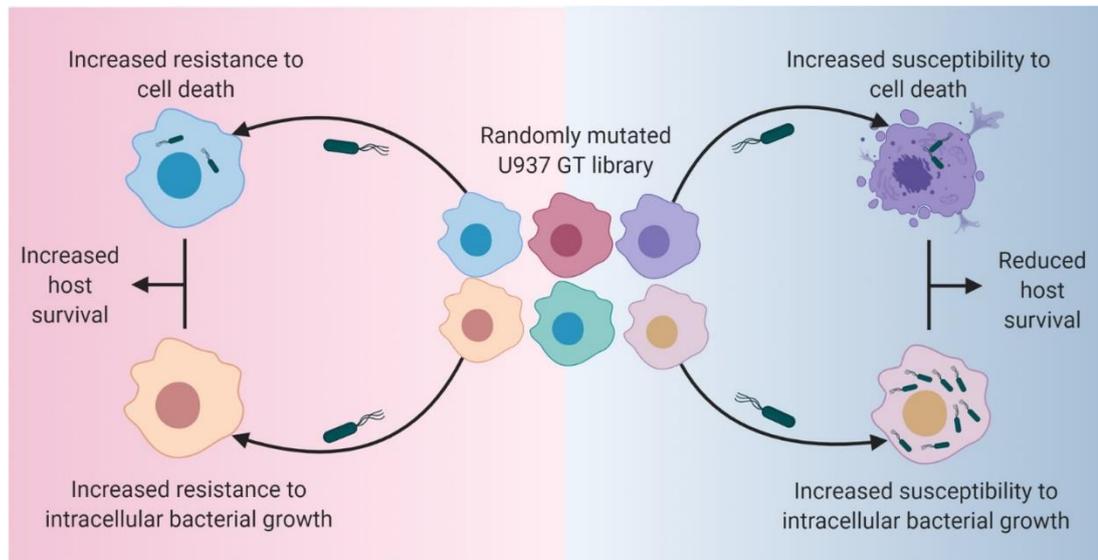
### 6.3.7. What technical limitations prevented identification of host infection factors common to STM and *F. tularensis*?

The STM and *F. tularensis* U937 screens had technical failings that reduced the potential to find common host infection factors. As discussed in Chapter 5, the *F. tularensis* screen lacked biological replicates, thus reducing the discovery rate. Additionally, up to 94% of trapped protein coding genes were extremely lowly represented (less than 5 transcripts in 90% of samples) within the STM screen (used as an example). As mentioned in Chapter 5, it is unknown how well represented a GT cell (and thus a GT transcript) within the GT library must be for a selection advantage or disadvantage to be distinguished from stochastic noise. Nevertheless, this information indicates that the library scope does not fully cover the protein coding genome. As proteins are currently much easier to target with drugs, this is the category of most interest. These technical caveats demonstrate that the conclusions drawn about the potential shared host infection factors in the human genome are predicated on a small subset of potentially targetable genes. So, what about comparisons with the broader literature base of host infection screens?

## 6.4. How could discovery screens be optimised?

### 6.4.1. Identification of infection resistance or susceptibility genes through host cell survival selection requires context – host cell survival is not conclusive of simple infection resistance

Programmed cell death (PCD) of infected host cells is an important part of the immune response arsenal to prevent increased pathogen growth for a wide spectrum of bacterial pathogens (as described in Chapter 1) (Jorgensen et al., 2017). For example, genetic inhibition of pyroptosis, necroptosis and apoptosis in mice results in dramatic increases in STM burden during *in vivo* infection (Doerflinger et al., 2020). PCD is so important that the signalling pathways between pyroptosis, necroptosis and apoptosis are well interconnected, greatly hindering pathogens from completely obstructing cell death (Christgen et al., 2020; Doerflinger et al., 2020; Jorgensen et al., 2017). Therefore, directly improving host cell survival during infection would often not be a good treatment strategy. However, improved host cell survival during an infection screen may often be the result of increased resistance to bacterial invasion, or bacteriostatic, or bactericidal activity (as indicated on the left hand side of Figure 6.2). Secondary analyses subsequent to host cell survival screens should be performed to distinguish between these underlying phenotypes. This can start with combining current literature knowledge with pathway analysis to narrow down potential genes of interest (as performed in Chapter 5). Ultimately, the phenotypes should be interrogated experimentally through genetic and or pharmacological means to determine the effect on bacterial burden and the underlying mechanism.



**Figure 6.2 Increased host cell survival may be the result of a resistance to cell death or a resistance to bacterial entry/growth.** Genetic mutations within a randomly mutated cell line library may predispose the host cell to a greater (on the left) or lesser (on the right) likelihood of cell death during a bacterial infection. This final phenotype may be the result of a difference in susceptibility to intracellular bacterial growth (bottom half), which entails a greater or lesser level of cellular damage/stress; thus the difference in propensity for cell death is merely a consequence of this. On the other hand, the difference in the likelihood of cell death due to the mutation may instead be the result of a difference in responsiveness to stressors or cell death cues; meaning that the infection is merely tangential.

#### 6.4.2. Host cell survival or Bacterial burden – methods of selection of infection resistance

Intracellular bacterial burden is a direct measure of the combined abilities of bacteria to enter, grow and survive inside the host cell; so why was bacterial burden not used to distinguish between resistant and susceptible host cell populations? Discovery screening of a randomly mutated host library requires subsequent isolation of the host cells to identify the mutations present in the selected population(s). The only methods of determining intracellular bacterial load that are not terminal are FACS of cells infected with fluorescent protein expressing bacteria, or live cell imaging of the same. Microscopy based screening at any reasonable scale requires automated robotic assistance; furthermore, subsequent isolation of host cell colonies of interest would be difficult to impossible. Microscopy-based infection screens have been performed by multiple labs but are restricted to an arrayed mutant library format, for example Anand et al., (2020), Casanova et al., (2019), or Jeng et al., (2019). The focus of discussion will therefore be on host cell survival and FACS-based bacterial burden screening.

Host cell survival and bacterial burden, when used to facilitate gene discovery each have technique-associated selection biases that may preclude the interrogation of certain mutants or act synergistically with the infection conditions, thus obscuring the real effect. Host cell survival selection is biased towards mutants that promote faster proliferation, or are hypo-responsive to cell death-, senescence-, or quiescence- cues in the overrepresented

trapped gene population; as well as the inverse in the underrepresented population. FACS-based selection is biased towards mutations promoting hardier cells (to withstand the high pressure encountered during sorting); indeed, many technical FACS-related factors can bias the screen based on cell stress responses. Some cell types are more fragile than others, cells in poor condition (e.g. after infection) are more susceptible to the effects of sorting. Furthermore, BSL2 pathogen infected cells require pre-sorting treatment to maintain safety standards, leading to greater clumping (Alles et al., 2017; Lanier and Warner, 1981; Westermann and Vogel, 2018). The technique-specific stressors would also obscure identification of processes important to infection susceptibility and resistance (Binek et al., 2019; Llufrío et al., 2018; Richardson et al., 2015). Besides these factors, further cells may be lost during FACS due to electronic aborts (occurring when multiple cells pass through the laser beam at a time) and sorting conflicts (when two or more cells are too close to each other to allow sorting). The stringency of the sorting selection can be modulated based on user requirements, but a balance has to be reached between retaining enough sorted cells and maintaining results of suitable quality. Indeed, as noted in Chapter 3, GT *PRKG1* HEK-293 cells may have produced an artificially high GFP signal due to clustering. The biases from each method can be adjusted for through comparison of the infected population with a mock-infected population, though mutants unable to withstand the experimental process regardless of infection would not be included.

#### 6.4.3. Mutant library generation – GT or CRISPR?

The GT technology used to generate the U937 randomly mutated library in Chapter 4 is cost effective but limited to affecting genes containing introns and the overall coverage cannot be controlled. An alternative technology for KO cell library generation mentioned in Chapter 1 is CRISPR Cas9. Use of CRISPR Cas9 for library generation has the distinct advantage of targeted gene deletion even at scale but comes at extremely high cost. A whole genome library ( $\geq 19,000,000$  genes – protein coding only) is the best possible coverage, but is this cost effective? Approximately a quarter of the human genome is considered druggable (Griffith et al., 2013; Wang et al., 2020), so reducing a CRISPR library to only include druggable targets would reduce the price while retaining the targetable genes. However, this reduces the available context of the novel drug targets identified in the screen; reducing the ability to determine the common features in the broad infection processes between pathogen screens.

A second factor is the feasibility of generating a CRISPR KO library with available resources. Use of an arrayed CRISPR KO library in combination with an arrayed selection format would entail knowledge retention of the identity of the host mutants throughout the selection process. This would obviate the need for costly NGS services, but the scale of the screen would be limited without costly robotic assistance. Additionally, prepared arrayed transfection technologies like viral vectors or lipofectamine are not effective in macrophages (Guo et al., 2019; Kajaste-Rudnitski and Naldini, 2015). Nucleofection is much more effective, but the accessible equipment does not allow scale (Maeß et al., 2014b). Based on these factors, the mutant cell library must be pooled, but the combination of buying a CRISPR library and NGS sequencing is potentially too costly.

#### 6.4.4. What is the optimal route to identify GTs of interest in a screen – clonal isolation of cells, or NGS of pooled populations?

Two methods of gene trap mutation identification have been explored here, the first (used in the HEK-293 GT screen, Chapter 1) requires clonal isolation of surviving host cells and subsequent Sanger sequencing of the RACE-PCR products. The second (used in two U937 screens, Chapter 4 and 5) allows sequencing of RACE-PCR products from the pooled selected population(s), but also requires NGS and subsequent bioinformatics analysis. NGS of pooled selected cells provides a quantitative analysis of differentially represented GT transcripts (and thus GT cells) in selected, compared to unselected, cells. It also provides an in-depth overview of the GT transcripts present in both populations, allowing the analyst to set a threshold at which the trapped genes are considered significant. In contrast, individual sequencing of clonal GT populations provides a list of genes (and other genomic sites) that have been mutated and which have allowed the survival of the host cell. This method does not provide any information about the size of the survival advantage conferred by the mutation, which can be inferred from the fold change and variation present between replicates in the former method.

Clonal isolation and sequencing is labour-intensive at large-scale, thus compromises must be made between the intensity of selection and the arbitrary picking of colonies for sequencing. Greater selection intensities will lose more GT cells of potential interest, but will in theory, lead to a smaller number of colonies containing mutants with greater survival advantages (thus avoiding the need to choose arbitrarily). Additionally, clonal isolation and sequencing contains many stages at which GT transcripts of interest can be lost, during clonal isolation and culturing or transduction of the plasmid containing the RACE-PCR product and unclear sequence signals, to name a few. However, as long as the sequence signals are clear, in most cases a sequence of sufficient length will be produced that the genomic location can be correctly (and uniquely) identified. Sequences with multiple hits (present in 5 of the isolated colonies in the HEK-293 screen) are likely from repetitive regions of the genome, non-canonical GT insertions, or highly similar paralogous genes. The sequence data further shows specifically where the gene trap insertion took place. The precise and accurate location information would aid further experimental analysis of the effect of non-canonical GT mutations as well as discrimination between predicted lncRNAs or anti-sense transcripts and nearby genes (Gow et al., 2013).

In contrast, sequencing from pooled cells requires a comparatively large initial time commitment to learn and optimise the bioinformatics analysis of the resulting sequence data, but this does not increase noticeably with increased scale. This allows a much greater number of different GT cells (and GT transcripts) to be analysed, allowing greater freedom of selection stringency. The greater potential scale of sequencing from pooled cells, allows analysis (e.g. pathway analysis) of the broader processes that are important to the survival of the host cell during infection, as well as a larger scope for comparison between screens. The use of well documented differential gene expression bioinformatics processing methods provides quantitative gene count information at the expense of information of precise gene trap insertion locations. An alternative bioinformatics processing method has been suggested (Chapter 5) which may provide more precise gene trap insertion locations

but at the expense of the accurate identification of some trapped genes or non-coding sequences.

Finally, sequencing from clonal isolates does not allow the identification of susceptibility mutations after destructive selection events, whereas sequencing from pooled cells facilitates identification of susceptibility mutations through comparison with a mock-treated population. The appropriate method depends on the usage requirements and experimental aims; in this case, the greater quantity of genes of interest and breadth of information provided by sequencing from pooled cells is advantageous to the identification of shared host infection factors.

#### 6.4.5. A solution for iterative rounds of infection selection

As discussed in Chapter 4, differentiated U937 cells do not proliferate, obfuscating the option to perform iterative rounds of infection selection. Iterative rounds of infection selection would have provided reassurance of the reliability of the genes identified as improved/reduced survival factors.

Undifferentiated U937 cells could be used as a host cell in an infection discovery screen (as performed by Jeng et al., (2019)) to allow iterative rounds of infection and sampling for pooled sequencing. Undifferentiated U937 cells may be more phenotypically similar to naïve monocytes, therefore, they will likely produce a different, overlapping set of differentially represented genes compared to the differentiated U937 screens described in Chapters 4 and 5.

### 6.5. Future work

#### 6.5.1. Further development of a pathway-centric approach through screening with a broader range of pathogens

A pathway-centric approach for drug target identification was suggested as an option (in Section 6.3) to broaden the applicability of host-targeted therapy for infection. This approach could work well but relies on a knowledge base of host interactions for a range of evolutionary distant bacteria. Such a knowledge base could be created from additional host infection screens with more evolutionarily distant intracellular bacteria. These host infection screens must be sufficiently large-scale to capture the majority of genes in a given functional network and of sufficient replication to distinguish the experimental signals through the noise. The genes identified in those screens could then be sampled within pathway analysis databases like STRING, CPDB and GOrilla to identify shared host infection networks.

#### 6.5.2. Taking the most promising host infection factors forward

Of the candidate genes that could be taken forward without further empirical or bioinformatics discovery, *PCDH7* (Chapter 3) is the most promising based on the current evidence. Additional candidates of interest requiring further investigation are *PRKG1* (Chapter 3) and *SLC7A11* (Chapter 5). To further confirm these genes as STM (and *S. sonnei*)

host infection factors, independent genetic mutants should be made in other well-characterised cell lines and primary cells. Where possible, gene products should be pharmacologically modulated and assessed for resistance to STM infection through measurement of bacterial burden and host cell death. If the infection resistance phenotypes withstand further testing in these physiological models, the resistance mechanism should be investigated, and the infection resistance should be tested *in vivo*.

### 6.5.3. Potential druggability of identified proteins and pathways

A major process indicated in increasing infected macrophage cell survival in the STM U937 screen was the inhibition of the oxidative stress response (Chapter 5). As the inhibition of enzymes involved in this response generates increased mROS or prevents their reduction, the cytosol becomes more oxidising. An oxidising cytosol has bactericidal activity against STM (Heijden et al., 2015), demonstrating that this process is a potential pharmacological target. Several drugs clinically approved to treat cancer increase ROS production but as a side effect, or as a means to induce cell death (Poprac et al., 2017). In contrast, a few drugs are described below that increase the oxidising intracellular environment without inducing cell death.

SLC7A11, half of a cystine-glutamate antiporter (xCT), is important for glutathione production (an antioxidant), without which the intracellular redox balance is more sensitive to increased ROS production. The *SLC7A11* gene was identified as a potential candidate in the STM screen (Chapter 5). Sulfasalazine, an inhibitor of SLC7A11, reduces macrophage intracellular *M. tuberculosis* burden (without generating host cell death), by reducing glutathione production (Cai et al., 2016). Sulfasalazine was employed to inhibit SLC7A11 during STM infection (described in Chapter 5), potentially at subinhibitory concentrations.

*TXNRD1*, encoding an enzyme that can recycle key electron carriers (Nalvarte et al., 2004), was also identified as an oxidative stress response gene of potential importance in STM infection (Chapter 5). *TXNRD1* is inhibited by Auranofin, an approved anti-rheumatic agent. Auranofin is currently in clinical trials as an add-on therapy (in combination with antibiotics) for pulmonary tuberculosis (The Aurum Institute NPC, 2019) or as a single drug therapy for amoebiasis (National Institute of Allergy and Infectious Diseases (NIAID), 2020, 2017). At 1  $\mu$ M, Auranofin-treated primary human neutrophils generate greater levels of superoxide and have greater *in vitro* survival (Liu et al., 2000). Auranofin also promotes ROS production in variety of other cell lines (Chen et al., 2016; You and Park, 2016). Additionally, mouse macrophages exposed to LPS in combination with ATP or Nigericin are less likely to undergo lytic cell death and produce less IL-1 $\beta$  when treated with Auranofin (Isakov et al., 2014; H. Y. Kim et al., 2019). Therefore, low concentrations of Auranofin both increase ROS production and inhibit lytic host cell death.

SERPINB2 is a well-established inhibitor of urokinase-type plasminogen activator, but the protein clearly has other intracellular functions (see Section 5.2.3.1 for detail). A trapped version of the gene was highly overrepresented in the STM U937 screen (Chapter 5), suggesting conferment of a protective effect, thus making SERPINB2 a desirable target. Two investigational protein drugs (Lanoteplase, Tenecteplase) based on human tissue-type plasminogen activator are suggested to interact with SERPINB2. These might be expected

to function as competitive inhibitors of SERPINB2, as they are inactivated forms of its binding partner.

SNAP25, a t-SNARE membrane fusion protein, is a potential host infection factor for STM (Stévenin et al., 2019) and *F. tularensis* infection, making it of particular interest. Currently, no drugs are available to target SNAP25 as the vast majority of available drugs and drug-like molecules are small enzyme agonists or antagonists. A small number of cytokines and receptors are currently targeted by monoclonal antibodies (Goulet and Atkins, 2020; Su and Shuai, 2020), but these are very expensive to produce. Two emerging types of drug – antisense oligonucleotides and proteolysis-targeting chimaeras – promise to expand the potentially druggable genome (Churcher, 2018; Crooke et al., 2018; Lai and Crews, 2017). Therefore, proteins like SNAP25 may one day become suitable drug targets.

Pharmacologically enhancing PCD itself to improve bacterial clearance may be an effective way to target a broad spectrum of bacterial infections; but is risky, as overactivation can generate excessive inflammation. Indeed, in respiratory tularaemia, an infection characterised by lethal cytokine storm (Mares et al., 2008), host cell death and the related signalling pathways are an important source of inflammation (Pulavendran et al., 2020; Singh et al., 2017).

## 6.6. Concluding remarks

In conclusion, this project has improved the knowledge-base of potential host infection factors available for potential pharmacological modulation to treat human STM, *S. sonnei* and *F. tularensis* infection. Pharmacological inhibition of host infection factors has the potential to function as a substitute for antibiotics and can potentially target conditions caused by closely related bacteria in the same genus and perhaps even those in different classes. This may counteract the need for bacterial strain typing to confirm antibiotic resistance/ susceptibility, as well as reducing the risk of the generation of antimicrobial resistance. The subsequent steps required to take this information closer to a drug treatment, are the validation of the infection phenotypes of the most promising host infection factors, their mutants, and their inhibitors.

## REFERENCES

- Aarreberg, L.D., Esser-Nobis, K., Driscoll, C., Shuvarikov, A., Roby, J.A., Gale, M., 2019. Interleukin-1 $\beta$  Induces mtDNA Release to Activate Innate Immune Signaling via cGAS-STING. *Molecular Cell* 74, 801-815.e6. <https://doi.org/10.1016/j.molcel.2019.02.038>
- Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Röhl, I., Hopfner, K.-P., Ludwig, J., Hornung, V., 2013. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* 498, 380-384. <https://doi.org/10.1038/nature12306>
- Acosta-Gutiérrez, S., Ferrara, L., Pathania, M., Masi, M., Wang, J., Bodrenko, I., Zahn, M., Winterhalter, M., Stavenger, R.A., Pagès, J.-M., Naismith, J.H., van den Berg, B., Page, M.G.P., Ceccarelli, M., 2018. Getting Drugs into Gram-Negative Bacteria: Rational Rules for Permeation through General Porins. *ACS Infect. Dis.* 4, 1487-1498. <https://doi.org/10.1021/acsinfecdis.8b00108>
- Adalja, A.A., Toner, E., Inglesby, T.V., 2015. Clinical Management of Potential Bioterrorism-Related Conditions [WWW Document]. <http://dx.doi.org/10.1056/NEJMra1409755>. <https://doi.org/10.1056/NEJMra1409755>
- Adey, A., Burton, J.N., Kitzman, J.O., Hiatt, J.B., Lewis, A.P., Martin, B.K., Qiu, R., Lee, C., Shendure, J., 2013. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* 500, 207-211. <https://doi.org/10.1038/nature12064>
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46, W537-W544. <https://doi.org/10.1093/nar/gky379>
- Ahmed, S.B.M., Amer, S., Emad, M., Rahmani, M., Prigent, S.A., 2019. Studying the ShcD and ERK interaction under acute oxidative stress conditions in melanoma cells. *The International Journal of Biochemistry & Cell Biology* 112, 123-133. <https://doi.org/10.1016/j.biocel.2019.05.009>
- Ahner, A., Gong, X., Schmidt, B.Z., Peters, K.W., Rabeh, W.M., Thibodeau, P.H., Lukacs, G.L., Frizzell, R.A., 2013. Small heat shock proteins target mutant cystic fibrosis transmembrane conductance regulator for degradation via a small ubiquitin-like modifier-dependent pathway. *Mol. Biol. Cell* 24, 74-84. <https://doi.org/10.1091/mbc.E12-09-0678>
- Akimana, C., Al-Khodor, S., Kwaik, Y.A., 2010. Host Factors Required for Modulation of Phagosome Biogenesis and Proliferation of *Francisella tularensis* within the Cytosol. *PLOS ONE* 5, e11025. <https://doi.org/10.1371/journal.pone.0011025>
- Alam, J., Cai, J., Smith, A., 1994. Isolation and characterization of the mouse heme oxygenase-1 gene. Distal 5' sequences are required for induction by heme or heavy metals. *J Biol Chem* 269, 1001-1009.
- Alam, J., Stewart, D., Touchard, C., Boinapally, S., Choi, A.M.K., Cook, J.L., 1999. Nrf2, a Cap'n'Collar Transcription Factor, Regulates Induction of the Heme Oxygenase-1 Gene. *J. Biol. Chem.* 274, 26071-26078. <https://doi.org/10.1074/jbc.274.37.26071>

- Aldo, P.B., Craveiro, V., Guller, S., Mor, G., 2013. Effect of Culture Conditions on the Phenotype of THP-1 Monocyte Cell Line. *American Journal of Reproductive Immunology* 70, 80–86. <https://doi.org/10.1111/aji.12129>
- Allen, H.K., Trachsel, J., Looft, T., Casey, T.A., 2014. Finding alternatives to antibiotics. *Ann. N.Y. Acad. Sci.* 1323, 91–100. <https://doi.org/10.1111/nyas.12468>
- Alles, J., Karaikos, N., Praktijnjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., Zinzen, R., Kocks, C., Rajewsky, N., 2017. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology* 15, 44. <https://doi.org/10.1186/s12915-017-0383-5>
- Alqahtani, M., Ma, Z., Ketkar, H., Suresh, R.V., Malik, M., Bakshi, C.S., 2018. Characterization of a Unique Outer Membrane Protein Required for Oxidative Stress Resistance and Virulence of *Francisella tularensis*. *Journal of Bacteriology* 200. <https://doi.org/10.1128/JB.00693-17>
- Ameen, N., Apodaca, G., 2007. Defective CFTR Apical Endocytosis and Enterocyte Brush Border in Myosin VI-Deficient Mice. *Traffic* 8, 998–1006. <https://doi.org/10.1111/j.1600-0854.2007.00587.x>
- Anand, I.S., Choi, W., Isberg, R.R., 2020. Components of the endocytic and recycling trafficking pathways interfere with the integrity of the Legionella-containing vacuole. *Cellular Microbiology* 22, e13151. <https://doi.org/10.1111/cmi.13151>
- Andreu, N., Phelan, J., de Sessions, P.F., Cliff, J.M., Clark, T.G., Hibberd, M.L., 2017. Primary macrophages and J774 cells respond differently to infection with *Mycobacterium tuberculosis*. *Sci Rep* 7. <https://doi.org/10.1038/srep42225>
- Andrews, S., 2019. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document]. Babraham Bioinformatics. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 8.14.19).
- Andritschke, D., Dilling, S., Emmenlauer, M., Welz, T., Schmich, F., Misselwitz, B., Rämö, P., Rottner, K., Kerkhoff, E., Wada, T., Penninger, J.M., Beerenwinkel, N., Horvath, P., Dehio, C., Hardt, W.-D., 2016. A Genome-Wide siRNA Screen Implicates Spire1/2 in SipA-Driven *Salmonella Typhimurium* Host Cell Invasion. *PLOS ONE* 11, e0161965. <https://doi.org/10.1371/journal.pone.0161965>
- Antl, M., von Brühl, M.-L., Eiglsperger, C., Werner, M., Konrad, I., Kocher, T., Wilm, M., Hofmann, F., Massberg, S., Schlossmann, J., 2007. IRAG mediates NO/cGMP-dependent inhibition of platelet aggregation and thrombus formation. *Blood* 109, 552–559. <https://doi.org/10.1182/blood-2005-10-026294>
- Arena, G., Cissé, M.Y., Pyrdziak, S., Chatre, L., Riscal, R., Fuentes, M., Arnold, J.J., Kastner, M., Gayte, L., Bertrand-Gaday, C., Nay, K., Angebault-Prouteau, C., Murray, K., Chabi, B., Koechlin-Ramonatxo, C., Orsetti, B., Vincent, C., Casas, F., Marine, J.-C., Etienne-Manneville, S., Bernex, F., Lombès, A., Cameron, C.E., Dubouchaud, H., Ricchetti, M., Linares, L.K., Cam, L.L., 2018. Mitochondrial MDM2 Regulates Respiratory Complex I Activity Independently of p53. *Molecular Cell* 69, 594–609.e8. <https://doi.org/10.1016/j.molcel.2018.01.023>
- Arizmendi, O., Picking, W.D., Picking, W.L., 2016. Macrophage Apoptosis Triggered by IpaD from *Shigella flexneri*. *Infect. Immun.* 84, 1857–1865. <https://doi.org/10.1128/IAI.01483-15>
- Asea, A., Kraeft, S.-K., Kurt-Jones, E.A., Stevenson, M.A., Chen, L.B., Finberg, R.W., Koo, G.C., Calderwood, S.K., 2000. HSP70 stimulates cytokine production through a CD14-dependant pathway, demonstrating its dual role as a chaperone and cytokine. *Nature Medicine* 6, 435–442. <https://doi.org/10.1038/74697>

- Ashida, H., Nakano, H., Sasakawa, C., 2013. Shigella IpaH0722 E3 Ubiquitin Ligase Effector Targets TRAF2 to Inhibit PKC–NF- $\kappa$ B Activity in Invaded Epithelial Cells. *PLOS Pathogens* 9, e1003409. <https://doi.org/10.1371/journal.ppat.1003409>
- Ashida, H., Ogawa, M., Mimuro, H., Kobayashi, T., Sanada, T., Sasakawa, C., 2011. Shigella are versatile mucosal pathogens that circumvent the host innate immune system. *Current Opinion in Immunology, Host pathogens/Immune senescence* 23, 448–455. <https://doi.org/10.1016/j.coi.2011.06.001>
- Ashley, C.L., Abendroth, A., McSharry, B.P., Slobedman, B., 2019. Interferon-Independent Upregulation of Interferon-Stimulated Genes during Human Cytomegalovirus Infection is Dependent on IRF3 Expression. *Viruses* 11, 246. <https://doi.org/10.3390/v11030246>
- Ashokkumar, B., Vaziri, N.D., Said, H.M., 2006. Thiamin uptake by the human-derived renal epithelial (HEK-293) cells: cellular and molecular mechanisms. *American Journal of Physiology-Renal Physiology* 291, F796–F805. <https://doi.org/10.1152/ajprenal.00078.2006>
- Asiamah, E.K., Ekwemalor, K., Adjei-Fremah, S., Osei, B., Newman, R., Worku, M., 2019. Natural and synthetic pathogen associated molecular patterns modulate galectin expression in cow blood. *J Anim Sci Technol* 61, 245–253. <https://doi.org/10.5187/jast.2019.61.5.245>
- Aspidou, Z., Balomenos, A., Tsakanikas, P., Manolakos, E., Koutsoumanis, K., 2019. Heterogeneity of single cell inactivation: Assessment of the individual cell time to death and implications in population behavior. *Food Microbiology* 80, 85–92. <https://doi.org/10.1016/j.fm.2018.12.011>
- Atianand, M.K., Duffy, E.B., Shah, A., Kar, S., Malik, M., Harton, J.A., 2011. Francisella tularensis Reveals a Disparity between Human and Mouse NLRP3 Inflammasome Activation. *J. Biol. Chem.* 286, 39033–39042. <https://doi.org/10.1074/jbc.M111.244079>
- Aulner, N., Danckaert, A., Ihm, J., Shum, D., Shorte, S.L., 2019. Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases. *Trends in Parasitology* 35, 559–570. <https://doi.org/10.1016/j.pt.2019.05.004>
- Babina, M., Henz, B.M., 2003. All-trans retinoic acid down-regulates expression and function of beta2 integrins by human monocytes: opposite effects on monocytic cell lines. *Eur. J. Immunol.* 33, 616–625. <https://doi.org/10.1002/eji.200323367>
- Bailly-Bechet, M., Benecke, A., Hardt, W.D., Lanza, V., Sturm, A., Zecchina, R., 2011. An externally modulated, noise-driven switch for the regulation of SPI1 in Salmonella enterica serovar Typhimurium. *J. Math. Biol.* 63, 637–662. <https://doi.org/10.1007/s00285-010-0385-1>
- Bajpai, G., Schneider, C., Wong, N., Bredemeyer, A., Hulsmans, M., Nahrendorf, M., Epelman, S., Kreisel, D., Liu, Y., Itoh, A., Shankar, T.S., Selzman, C.H., Drakos, S.G., Lavine, K.J., 2018. The human heart contains distinct macrophage subsets with divergent origins and functions. *Nat Med* 24, 1234–1245. <https://doi.org/10.1038/s41591-018-0059-x>
- Baker, P.J., Boucher, D., Bierschenk, D., Tebartz, C., Whitney, P.G., D’Silva, D.B., Tanzer, M.C., Monteleone, M., Robertson, A.A.B., Cooper, M.A., Alvarez-Diaz, S., Herold, M.J., Bedoui, S., Schroder, K., Masters, S.L., 2015. NLRP3 inflammasome activation downstream of cytoplasmic LPS recognition by both caspase-4 and caspase-5. *European Journal of Immunology* 45, 2918–2926. <https://doi.org/10.1002/eji.201545655>

- Bakkeren, E., Huisman, J.S., Fattinger, S.A., Hausmann, A., Furter, M., Egli, A., Slack, E., Sellin, M.E., Bonhoeffer, S., Regoes, R.R., Diard, M., Hardt, W.-D., 2019. Salmonella persists promote the spread of antibiotic resistance plasmids in the gut. *Nature* 573, 276–280. <https://doi.org/10.1038/s41586-019-1521-8>
- Balali-Mood, M., Moshiri, M., Etemad, L., 2013. Medical aspects of bio-terrorism. *Toxicon, The 17th World Congress of the International Society on Toxinology and Venom Week 2012* 69, 131–142. <https://doi.org/10.1016/j.toxicon.2013.01.005>
- Bang, S., Son, S., Kim, S., Shin, H., 2019. Disease Pathway Cut for Multi-Target drugs. *BMC Bioinformatics* 20, 74. <https://doi.org/10.1186/s12859-019-2638-3>
- Bannai, S., Sato, H., Ishii, T., Sugita, Y., 1989. Induction of cystine transport activity in human fibroblasts by oxygen. *J. Biol. Chem.* 264, 18480–18484.
- Barnich, N., Aguirre, J.E., Reinecker, H.-C., Xavier, R., Podolsky, D.K., 2005. Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor- $\kappa$ B activation in muramyl dipeptide recognition. *J Cell Biol* 170, 21–26. <https://doi.org/10.1083/jcb.200502153>
- Baruzzo, G., Hayer, K.E., Kim, E.J., Di Camillo, B., FitzGerald, G.A., Grant, G.R., 2017. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods* 14, 135–139. <https://doi.org/10.1038/nmeth.4106>
- Basoni, C., Nobles, M., Grimshaw, A., Desgranges, C., Davies, D., Perretti, M., Kramer, I.J.M., Genot, E., 2005. Inhibitory control of TGF- $\beta$ 1 on the activation of Rap1, CD11b, and transendothelial migration of leukocytes. *The FASEB Journal* 19, 822–824. <https://doi.org/10.1096/fj.04-3085fje>
- Bazopoulou, D., Knoefler, D., Zheng, Y., Ulrich, K., Oleson, B.J., Xie, L., Kim, M., Kaufmann, A., Lee, Y.-T., Dou, Y., Chen, Y., Quan, S., Jakob, U., 2019. Developmental ROS individualizes organismal stress resistance and lifespan. *Nature* 576, 301–305. <https://doi.org/10.1038/s41586-019-1814-y>
- Beckwith, K.S., Beckwith, M.S., Ullmann, S., Sætra, R.S., Kim, H., Marstad, A., Åsberg, S.E., Strand, T.A., Haug, M., Niederweis, M., Stenmark, H.A., Flo, T.H., 2020. Plasma membrane damage causes NLRP3 activation and pyroptosis during Mycobacterium tuberculosis infection. *Nature Communications* 11, 2270. <https://doi.org/10.1038/s41467-020-16143-6>
- Behrends, C., Sowa, M.E., Gygi, S.P., Harper, J.W., 2010. Network organization of the human autophagy system. *Nature* 466, 68–76. <https://doi.org/10.1038/nature09204>
- Belluzzi, A., Serrani, M., Roda, G., Bianchi, M.L., Castellani, L., Grazia, M., Rosati, G., Ugolini, G., Roda, E., 2010. Pilot study: the use of sulfasalazine for the treatment of acute pouchitis. *Alimentary Pharmacology & Therapeutics* 31, 228–232. <https://doi.org/10.1111/j.1365-2036.2009.04163.x>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bergounioux, J., Elisee, R., Prunier, A.-L., Donnadieu, F., Sperandio, B., Sansonetti, P., Arbibe, L., 2012. Calpain Activation by the Shigella flexneri Effector VirA Regulates Key Steps in the Formation and Life of the Bacterium's Epithelial Niche. *Cell Host & Microbe* 11, 240–252. <https://doi.org/10.1016/j.chom.2012.01.013>
- Berkowitz, J.K.M., Catz, S.D., Johnson, J.L., Ruedi, J.M., Thon, V., Babior, B.M., 2001. JFC1, a Novel Tandem C2 Domain-containing Protein Associated with the Leukocyte NADPH Oxidase. *J. Biol. Chem.* 276, 18855–18862. <https://doi.org/10.1074/jbc.M011167200>

- Bertram, C., von Neuhoff, N., Skawran, B., Steinemann, D., Schlegelberger, B., Hass, R., 2008. The differentiation/retrodifferentiation program of human U937 leukemia cells is accompanied by changes of VCP/p97. *BMC Cell Biology* 9, 12. <https://doi.org/10.1186/1471-2121-9-12>
- Bierschenk, D., Boucher, D., Schroder, K., 2017. Salmonella-induced inflammasome activation in humans. *Molecular Immunology, Inflammasomes* 86, 38–43. <https://doi.org/10.1016/j.molimm.2016.11.009>
- Bierschenk, D., Monteleone, M., Moghaddas, F., Baker, P.J., Masters, S.L., Boucher, D., Schroder, K., 2019. The Salmonella pathogenicity island-2 subverts human NLRP3 and NLRC4 inflammasome responses. *Journal of Leukocyte Biology* 105, 401–410. <https://doi.org/10.1002/JLB.MA0318-112RR>
- Billing, A.M., Fack, F., Turner, J.D., Muller, C.P., 2011. Cortisol is a potent modulator of lipopolysaccharide-induced interferon signaling in macrophages. *Innate Immun* 17, 302–320. <https://doi.org/10.1177/1753425910369269>
- Binda, O., Roy, J.-S., Branton, P.E., 2006. RBP1 Family Proteins Exhibit SUMOylation-Dependent Transcriptional Repression and Induce Cell Growth Inhibition Reminiscent of Senescence. *Mol. Cell. Biol.* 26, 1917–1931. <https://doi.org/10.1128/MCB.26.5.1917-1931.2006>
- Binek, A., Rojo, D., Godzien, J., Rupérez, F.J., Nuñez, V., Jorge, I., Ricote, M., Vázquez, J., Barbas, C., 2019. Flow Cytometry Has a Significant Impact on the Cellular Metabolome. *J. Proteome Res.* 18, 169–181. <https://doi.org/10.1021/acs.jproteome.8b00472>
- Birger, A., Ottolenghi, M., Perez, L., Reubinoff, B., Behar, O., 2018. ALS-related human cortical and motor neurons survival is differentially affected by Sema3A. *Cell Death & Disease* 9, 1–9. <https://doi.org/10.1038/s41419-018-0294-6>
- Birmingham, C.L., Smith, A.C., Bakowski, M.A., Yoshimori, T., Brumell, J.H., 2006. Autophagy Controls Salmonella Infection in Response to Damage to the Salmonella-containing Vacuole. *J. Biol. Chem.* 281, 11374–11383. <https://doi.org/10.1074/jbc.M509157200>
- Blander, J.M., Medzhitov, R., 2004. Regulation of Phagosome Maturation by Signals from Toll-Like Receptors. *Science* 304, 1014–1018. <https://doi.org/10.1126/science.1096158>
- Bogdan, C., 2015. Nitric oxide synthase in innate and adaptive immunity: an update. *Trends in Immunology* 36, 161–178. <https://doi.org/10.1016/j.it.2015.01.003>
- Bogomolnaya, L.M., Andrews, K.D., Talamantes, M., Maple, A., Ragoza, Y., Vazquez-Torres, A., Andrews-Polymenis, H., 2013. The ABC-Type Efflux Pump MacAB Protects *Salmonella enterica* serovar Typhimurium from Oxidative Stress. *mBio* 4. <https://doi.org/10.1128/mBio.00630-13>
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonifacino, J.S., Hurley, J.H., 2008. Retromer. *Current Opinion in Cell Biology* 20, 427–436. <https://doi.org/10.1016/j.ceb.2008.03.009>
- Booker, M., Samsonova, A.A., Kwon, Y., Flockhart, I., Mohr, S.E., Perrimon, N., 2011. False negative rates in *Drosophila* cell-based RNAi screens: a case study. *BMC Genomics* 12, 50. <https://doi.org/10.1186/1471-2164-12-50>
- Bosurgi, L., Cao, Y.G., Cabeza-Cabrerizo, M., Tucci, A., Hughes, L.D., Kong, Y., Weinstein, J.S., Licona-Limon, P., Schmid, E.T., Pelorosso, F., Gagliani, N., Craft, J.E., Flavell, R.A., Ghosh, S., Rothlin, C.V., 2017. Macrophage function in tissue repair and remodeling

- requires IL-4 or IL-13 with apoptotic cells. *Science* 356, 1072–1076.  
<https://doi.org/10.1126/science.aai8132>
- Boukes, G.J., van de Venter, M., 2012. Rooperol as an antioxidant and its role in the innate immune system: An in vitro study. *Journal of Ethnopharmacology* 144, 692–699.  
<https://doi.org/10.1016/j.jep.2012.10.014>
- Bournazou, I., Pound, J.D., Duffin, R., Bournazos, S., Melville, L.A., Brown, S.B., Rossi, A.G., Gregory, C.D., 2009. Apoptotic human cells inhibit migration of granulocytes via release of lactoferrin. *J Clin Invest* 119, 20–32. <https://doi.org/10.1172/JCI36226>
- Bradburne, C.E., Verhoeven, A.B., Manyam, G.C., Chaudhry, S.A., Chang, E.L., Thach, D.C., Bailey, C.L., Hoek, M.L. van, 2013. Temporal Transcriptional Response during Infection of Type II Alveolar Epithelial Cells with *Francisella tularensis* Live Vaccine Strain (LVS) Supports a General Host Suppression and Bacterial Uptake by Macropinocytosis. *J. Biol. Chem.* 288, 10780–10791.  
<https://doi.org/10.1074/jbc.M112.362178>
- Braun, J.E., Huntzinger, E., Fauser, M., Izaurrealde, E., 2011. GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. *Molecular Cell* 44, 120–133. <https://doi.org/10.1016/j.molcel.2011.09.007>
- Bravo, D., Blondel, C.J., Hoare, A., Leyton, L., Valvano, M.A., Contreras, I., 2011. Type IVB pili are required for invasion but not for adhesion of *Salmonella enterica* serovar Typhi into BHK epithelial cells in a cystic fibrosis transmembrane conductance regulator-independent manner. *Microbial Pathogenesis* 51, 373–377.  
<https://doi.org/10.1016/j.micpath.2011.07.005>
- Brawn, L.C., Hayward, R.D., Koronakis, V., 2007. *Salmonella* SPI1 Effector SipA Persists after Entry and Cooperates with a SPI2 Effector to Regulate Phagosome Maturation and Intracellular Replication. *Cell Host & Microbe* 1, 63–75.  
<https://doi.org/10.1016/j.chom.2007.02.001>
- Breuer, K., Ferooshani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E.W., Brinkman, F.S.L., Lynn, D.J., 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* 41, D1228–1233. <https://doi.org/10.1093/nar/gks1147>
- Britzen-Laurent, N., Bauer, M., Berton, V., Fischer, N., Syguda, A., Reipschläger, S., Naschberger, E., Herrmann, C., Stürzl, M., 2010. Intracellular Trafficking of Guanylate-Binding Proteins Is Regulated by Heterodimerization in a Hierarchical Manner. *PLOS ONE* 5, e14246. <https://doi.org/10.1371/journal.pone.0014246>
- Brodmann, M., Dreier, R.F., Broz, P., Basler, M., 2017. *Francisella* requires dynamic type VI secretion system and ClpB to deliver effectors for phagosomal escape. *Nat Commun* 8. <https://doi.org/10.1038/ncomms15853>
- Brown, D., 2015. Antibiotic resistance breakers: can repurposed drugs fill the antibiotic discovery void? *Nat Rev Drug Discov* 14, 821–832. <https://doi.org/10.1038/nrd4675>
- Bujko, A., Atlasy, N., Landsverk, O.J.B., Richter, L., Yaqub, S., Horneland, R., Øyen, O., Aandahl, E.M., Aabakken, L., Stunnenberg, H.G., Bækkevold, E.S., Jahnsen, F.L., 2017. Transcriptional and functional profiling defines human small intestinal macrophage subsets. *Journal of Experimental Medicine* 215, 441–458.  
<https://doi.org/10.1084/jem.20170057>
- Bulgin, R., Raymond, B., Garnett, J.A., Frankel, G., Crepin, V.F., Berger, C.N., Arbeloa, A., 2010. Bacterial Guanine Nucleotide Exchange Factors SopE-Like and WxxxE Effectors. *Infection and Immunity* 78, 1417–1425.  
<https://doi.org/10.1128/IAI.01250-09>

- Bulua, A.C., Simon, A., Maddipati, R., Pelletier, M., Park, H., Kim, K.-Y., Sack, M.N., Kastner, D.L., Siegel, R.M., 2011. Mitochondrial reactive oxygen species promote production of proinflammatory cytokines and are elevated in TNFR1-associated periodic syndrome (TRAPS). *J Exp Med* 208, 519–533.  
<https://doi.org/10.1084/jem.20102049>
- Bürckstümmer, T., Baumann, C., Blüml, S., Dixit, E., Dürnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., Bennett, K.L., Superti-Furga, G., 2009. An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nature Immunology* 10, 266–272. <https://doi.org/10.1038/ni.1702>
- Burrows, A.C., Prokop, J., Summers, M.K., 2012. Skp1-Cul1-F-box Ubiquitin Ligase (SCF $\beta$ TrCP)-mediated Destruction of the Ubiquitin-specific Protease USP37 during G2-phase Promotes Mitotic Entry. *J. Biol. Chem.* 287, 39021–39029.  
<https://doi.org/10.1074/jbc.M112.390328>
- Burtnick, M.N., Brett, P.J., Nair, V., Warawa, J.M., Woods, D.E., Gherardini, F.C., 2008. Burkholderia pseudomallei Type III Secretion System Mutants Exhibit Delayed Vacuolar Escape Phenotypes in RAW 264.7 Murine Macrophages. *Infection and Immunity* 76, 2991–3000. <https://doi.org/10.1128/IAI.00263-08>
- Butchar, J.P., Cremer, T.J., Clay, C.D., Gavrilin, M.A., Wewers, M.D., Marsh, C.B., Schlesinger, L.S., Tridandapani, S., 2008. Microarray Analysis of Human Monocytes Infected with Francisella tularensis Identifies New Targets of Host Response Subversion. *PLOS ONE* 3, e2924. <https://doi.org/10.1371/journal.pone.0002924>
- Buwitt-Beckmann, U., Heine, H., Wiesmüller, K.-H., Jung, G., Brock, R., Akira, S., Ulmer, A.J., 2006. TLR1- and TLR6-independent Recognition of Bacterial Lipopeptides. *J. Biol. Chem.* 281, 9049–9057. <https://doi.org/10.1074/jbc.M512525200>
- Byrne, A., Cole, C., Volden, R., Vollmers, C., 2019. Realizing the potential of full-length transcriptome sequencing. *Philos Trans R Soc Lond B Biol Sci* 374.  
<https://doi.org/10.1098/rstb.2019.0097>
- Cai, Y., Yang, Q., Liao, M., Wang, H., Zhang, C., Nambi, S., Wang, W., Zhang, M., Wu, J., Deng, G., Deng, Q., Liu, H., Zhou, B., Jin, Q., Feng, C.G., Sasseti, C.M., Wang, F., Chen, X., 2016. xCT increases tuberculosis susceptibility by regulating antimicrobial function and inflammation. *Oncotarget* 7, 31001–31013.  
<https://doi.org/10.18632/oncotarget.9052>
- Callebaut, I., Hoffmann, B., Lehn, P., Mornon, J.-P., 2017. Molecular modelling and molecular dynamics of CFTR. *Cell. Mol. Life Sci.* 74, 3–22.  
<https://doi.org/10.1007/s00018-016-2385-9>
- Canton, J., Khezri, R., Glogauer, M., Grinstein, S., 2014. Contrasting phagosome pH regulation and maturation in human M1 and M2 macrophages. *MBoC* 25, 3330–3341. <https://doi.org/10.1091/mbc.e14-05-0967>
- Cargo of TIMM23 SORT [mitochondrial inner membrane] (No. Reactome release 72), n.d. . Reactome.
- Carneiro, L.A.M., Travassos, L.H., Soares, F., Tattoli, I., Magalhaes, J.G., Bozza, M.T., Plotkowski, M.C., Sansonetti, P.J., Molkentin, J.D., Philpott, D.J., Girardin, S.E., 2009. Shigella Induces Mitochondrial Dysfunction and Cell Death in Nonmyeloid Cells. *Cell Host & Microbe* 5, 123–136. <https://doi.org/10.1016/j.chom.2008.12.011>
- Casanova, A., Low, S.H., Québatte, M., Sedzicki, J., Tschon, T., Ketterer, M., Smith, K., Emmenlauer, M., Ben-Tekaya, H., Dehio, C., 2019. A Role for the VPS Retromer in Brucella Intracellular Replication Revealed by Genomewide siRNA Screening. *mSphere* 4. <https://doi.org/10.1128/mSphere.00380-19>

- Casden, N., Behar, O., 2019. An approach for accelerated isolation of genetically manipulated cell clones with reduced clonal variability. *J Cell Sci* 132. <https://doi.org/10.1242/jcs.217661>
- Case, E.D.R., Chong, A., Wehrly, T.D., Hansen, B., Child, R., Hwang, S., Virgin, H.W., Celli, J., 2014. The Francisella O-antigen mediates survival in the macrophage cytosol via autophagy avoidance. *Cellular Microbiology* 16, 862–877. <https://doi.org/10.1111/cmi.12246>
- Casson, C.N., Yu, J., Reyes, V.M., Taschuk, F.O., Yadav, A., Copenhaver, A.M., Nguyen, H.T., Collman, R.G., Shin, S., 2015. Human caspase-4 mediates noncanonical inflammasome activation against gram-negative bacterial pathogens. *PNAS* 112, 6688–6693. <https://doi.org/10.1073/pnas.1421699112>
- Casson, J., McKenna, M., High, S., 2016. On the road to nowhere: cross-talk between post-translational protein targeting and cytosolic quality control. *Biochemical Society Transactions* 44, 796–801. <https://doi.org/10.1042/BST20160045>
- Castanheira, S., García-del Portillo, F., 2017. Salmonella Populations inside Host Cells. *Front Cell Infect Microbiol* 7. <https://doi.org/10.3389/fcimb.2017.00432>
- Casteel, D.E., Zhuang, S., Gudi, T., Tang, J., Vuica, M., Desiderio, S., Pilz, R.B., 2002. cGMP-dependent Protein Kinase I $\beta$  Physically and Functionally Interacts with the Transcriptional Regulator TFII-I. *J. Biol. Chem.* 277, 32003–32014. <https://doi.org/10.1074/jbc.M112332200>
- Celli, J., Zahrt, T.C., 2013. Mechanisms of Francisella tularensis Intracellular Pathogenesis. *Cold Spring Harb Perspect Med* 3. <https://doi.org/10.1101/cshperspect.a010314>
- Center for Computational Biology at Johns Hopkins University, 2016. TopHat: A spliced read mapper for RNA-Seq [WWW Document]. Johns Hopkins University Center for Computational Biology. URL <https://ccb.jhu.edu/software/tophat/index.shtml>
- Centers for Disease Control and Prevention, 2018. Bioterrorism Agents/Diseases | Emergency Preparedness & Response [WWW Document]. URL <https://emergency.cdc.gov/agent/agentlist-category.asp> (accessed 9.28.18).
- Centers for Disease Control and Prevention, 2000. Biological and Chemical Terrorism: Strategic Plan for Preparedness and Response (No. 49 (RR04)), Morbidity and Mortality Weekly Report. Centres for Disease Control and Prevention.
- Chaitidis, P., O'Donnell, V., Kuban, R.J., Bermudez-Fajardo, A., Ungethuen, U., Kühn, H., 2005. Gene expression alterations of human peripheral blood monocytes induced by medium-term treatment with the TH2-cytokines interleukin-4 and -13. *Cytokine* 30, 366–377. <https://doi.org/10.1016/j.cyto.2005.02.004>
- Chamaillard, M., Hashimoto, M., Horie, Y., Masumoto, J., Qiu, S., Saab, L., Ogura, Y., Kawasaki, A., Fukase, K., Kusumoto, S., Valvano, M.A., Foster, S.J., Mak, T.W., Nuñez, G., Inohara, N., 2003. An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunology* 4, 702–707. <https://doi.org/10.1038/ni945>
- Chanput, W., Mes, J.J., Wichers, H.J., 2014. THP-1 cell line: An in vitro cell model for immune modulation approach. *International Immunopharmacology* 23, 37–45. <https://doi.org/10.1016/j.intimp.2014.08.002>
- Chanput, W., Peters, V., Wichers, H., 2015. THP-1 and U937 Cells, in: Verhoeckx, K., Cotter, P., López-Expósito, I., Kleiveland, C., Lea, T., Mackie, A., Requena, T., Swiatecka, D., Wichers, H. (Eds.), *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models*. Springer, Cham (CH).
- Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., Filipowicz, W., 2011. miRNA repression involves GW182-mediated recruitment of CCR4–NOT through

- conserved W-containing motifs. *Nature Structural & Molecular Biology* 18, 1218–1226. <https://doi.org/10.1038/nsmb.2166>
- Chen, H.-F., Ma, R.-R., He, J.-Y., Zhang, H., Liu, X.-L., Guo, X.-Y., Gao, P., 2017. Protocadherin 7 inhibits cell migration and invasion through E-cadherin in gastric cancer. *Tumour Biol.* 39, 1010428317697551. <https://doi.org/10.1177/1010428317697551>
- Chen, L., Lu, D., Sun, K., Xu, Y., Hu, P., Li, X., Xu, F., 2019. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 692, 119–125. <https://doi.org/10.1016/j.gene.2019.01.001>
- Chen, X., Tzekov, R., Su, M., Hong, H., Min, W., Han, A., Li, W., 2016. Auranofin Inhibits Retinal Pigment Epithelium Cell Survival through Reactive Oxygen Species-Dependent Epidermal Growth Factor Receptor/ Mitogen-Activated Protein Kinase Signaling Pathway. *PLOS ONE* 11, e0166386. <https://doi.org/10.1371/journal.pone.0166386>
- Chen, Y.-K., Chen, C.-Y., Hu, H.-T., Hsueh, Y.-P., 2012. CTTNBP2, but not CTTNBP2NL, regulates dendritic spinogenesis and synaptic distribution of the striatin–PP2A complex. *MBoC* 23, 4383–4392. <https://doi.org/10.1091/mbc.e12-05-0365>
- Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O’Riordan, C.R., Smith, A.E., 1990. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* 63, 827–834. [https://doi.org/10.1016/0092-8674\(90\)90148-8](https://doi.org/10.1016/0092-8674(90)90148-8)
- Cheng, Y., Guo, X., Gong, Y., Ding, X., Yu, Y., 2015. Sentrin/small ubiquitin-like modifier-specific protease 5 protects oral cancer cells from oxidative stress-induced apoptosis. *Molecular Medicine Reports* 12, 2009–2014. <https://doi.org/10.3892/mmr.2015.3662>
- Cheng, Y., Hu, X., Liu, C., Chen, M., Wang, J., Wang, M., Gao, F., Han, J., Sun, D., Zhang, C., Min, R., 2017. Gelsolin Inhibits the Inflammatory Process Induced by LPS. *CPB* 41, 205–212. <https://doi.org/10.1159/000456043>
- Cheng, Y.S., Colonna, R.J., Yin, F.H., 1983. Interferon induction of fibroblast proteins with guanylate binding activity. *J. Biol. Chem.* 258, 7746–7750.
- Chiang, C.-Y., Uzoma, I., Moore, R.T., Gilbert, M., Duplantier, A.J., Panchal, R.G., 2018. Mitigating the Impact of Antibacterial Drug Resistance through Host-Directed Therapies: Current Progress, Outlook, and Challenges. *mBio* 9. <https://doi.org/10.1128/mBio.01932-17>
- Choi, A.M., Alam, J., 1996. Heme oxygenase-1: function, regulation, and implication of a novel stress-inducible protein in oxidant-induced lung injury. *Am J Respir Cell Mol Biol* 15, 9–19. <https://doi.org/10.1165/ajrcmb.15.1.8679227>
- Chong, A., Wehrly, T.D., Nair, V., Fischer, E.R., Barker, J.R., Klose, K.E., Celli, J., 2008. The Early Phagosomal Stage of *Francisella tularensis* Determines Optimal Phagosomal Escape and *Francisella* Pathogenicity Island Protein Expression. *Infection and Immunity* 76, 5488–5499. <https://doi.org/10.1128/IAI.00682-08>
- Choubey, D., Walter, S., Geng, Y., Xin, H., 2000. Cytoplasmic localization of the interferon-inducible protein that is encoded by the AIM2 (absent in melanoma) gene from the 200-gene family. *FEBS Letters* 474, 38–42. [https://doi.org/10.1016/S0014-5793\(00\)01571-4](https://doi.org/10.1016/S0014-5793(00)01571-4)
- Chouchani, E.T., Pell, V.R., Gaude, E., Aksentijević, D., Sundier, S.Y., Robb, E.L., Logan, A., Nadtochiy, S.M., Ord, E.N.J., Smith, A.C., Eyassu, F., Shirley, R., Hu, C.-H., Dare, A.J., James, A.M., Rogatti, S., Hartley, R.C., Eaton, S., Costa, A.S.H., Brookes, P.S., Davidson, S.M., Duchon, M.R., Saeb-Parsy, K., Shattock, M.J., Robinson, A.J., Work,

- L.M., Frezza, C., Krieg, T., Murphy, M.P., 2014. Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* 515, 431–435. <https://doi.org/10.1038/nature13909>
- Choudhury, R., Bonacci, T., Wang, X., Truong, A., Arceci, A., Zhang, Y., Mills, C.A., Kernan, J.L., Liu, P., Emanuele, M.J., 2017. The E3 Ubiquitin Ligase SCF(Cyclin F) Transmits AKT Signaling to the Cell-Cycle Machinery. *Cell Reports* 20, 3212–3222. <https://doi.org/10.1016/j.celrep.2017.08.099>
- Christgen, S., Zheng, M., Kesavardhana, S., Karki, R., Malireddi, R.K.S., Banoth, B., Place, D.E., Briard, B., Sharma, B.R., Tuladhar, S., Samir, P., Burton, A., Kanneganti, T.-D., 2020. Identification of the PANoptosome: A Molecular Platform Triggering Pyroptosis, Apoptosis, and Necroptosis (PANoptosis). *Front. Cell. Infect. Microbiol.* 10. <https://doi.org/10.3389/fcimb.2020.00237>
- Chung The, H., Boinett, C., Pham Thanh, D., Jenkins, C., Weill, F.-X., Howden, B.P., Valcanis, M., De Lappe, N., Cormican, M., Wangchuk, S., Bodhidatta, L., Mason, C.J., Nguyen, T.N.T., Ha Thanh, T., Voong, V.P., Duong, V.T., Nguyen, P.H.L., Turner, P., Wick, R., Ceysens, P.-J., Thwaites, G., Holt, K.E., Thomson, N.R., Rabaa, M.A., Baker, S., 2019. Dissecting the molecular evolution of fluoroquinolone-resistant *Shigella sonnei*. *Nat Commun* 10, 4828. <https://doi.org/10.1038/s41467-019-12823-0>
- Chung The, H., Rabaa, M.A., Thanh, D.P., Lappe, N.D., Cormican, M., Valcanis, M., Howden, B.P., Wangchuk, S., Bodhidatta, L., Mason, C.J., Nguyen, T.N.T., Thuy, D.V., Thompson, C.N., Lan, N.P.H., Vinh, P.V., Thanh, T.H., Turner, P., Sar, P., Thwaites, G., Thomson, N.R., Holt, K.E., Baker, S., 2016. South Asia as a Reservoir for the Global Spread of Ciprofloxacin-Resistant *Shigella sonnei*: A Cross-Sectional Study. *PLOS Medicine* 13, e1002055. <https://doi.org/10.1371/journal.pmed.1002055>
- Churcher, I., 2018. PROTAC-Induced Protein Degradation in Drug Discovery: Breaking the Rules or Just Making New Ones? *J. Med. Chem.* 61, 444–452. <https://doi.org/10.1021/acs.jmedchem.7b01272>
- Citterio, L., Ferrandi, M., Carpini, S.D., Simonini, M., Kuznetsova, T., Molinari, I., Antonio, G.D., Lanzani, C., Merlino, L., Brioni, E., Staessen, J.A., Bianchi, G., Manunta, P., 2013. cGMP-Dependent Protein Kinase 1 Polymorphisms Underlie Renal Sodium Handling Impairment. *Hypertension* 62, 1027–1033. <https://doi.org/10.1161/HYPERTENSIONAHA.113.01628>
- Clemens, D.L., Lee, B.-Y., Horwitz, M.A., 2005. *Francisella tularensis* Enters Macrophages via a Novel Process Involving Pseudopod Loops. *Infect Immun* 73, 5892–5902. <https://doi.org/10.1128/IAI.73.9.5892-5902.2005>
- Cohen-Katsenelson, K., Stender, J.D., Kawashima, A.T., Lordén, G., Uchiyama, S., Nizet, V., Glass, C.K., Newton, A.C., 2019. PHLPP1 counter-regulates STAT1-mediated inflammatory signaling. *Elife* 8, e48609. <https://doi.org/10.7554/eLife.48609>
- Compan, V., Baroja-Mazo, A., López-Castejón, G., Gomez, A.I., Martínez, C.M., Angosto, D., Montero, M.T., Herranz, A.S., Bazán, E., Reimers, D., Mulero, V., Pelegrín, P., 2012. Cell Volume Regulation Modulates NLRP3 Inflammasome Activation. *Immunity* 37, 487–500. <https://doi.org/10.1016/j.immuni.2012.06.013>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* 17. <https://doi.org/10.1186/s13059-016-0881-8>
- Connell, S., Meade, K.G., Allan, B., Lloyd, A.T., Downing, T., O'Farrelly, C., Bradley, D.G., 2013. Genome-wide association analysis of avian resistance to *Campylobacter*

- jejuni colonization identifies risk locus spanning the CDH13 gene. *G3 (Bethesda)* 3, 881–890. <https://doi.org/10.1534/g3.113.006031>
- Connelly, K.E., Dykhuizen, E.C., 2017. Compositional and functional diversity of canonical PRC1 complexes in mammals. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1860, 233–245. <https://doi.org/10.1016/j.bbagr.2016.12.006>
- Connor, M.G., Pulsifer, A.R., Chung, D., Rouchka, E.C., Ceresa, B.K., Lawrenz, M.B., 2018. *Yersinia pestis* Targets the Host Endosome Recycling Pathway during the Biogenesis of the *Yersinia*-Containing Vacuole To Avoid Killing by Macrophages. *mBio* 9. <https://doi.org/10.1128/mBio.01800-17>
- Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., Ferguson, D.J.P., Campbell, B.J., Jewell, D., Simmons, A., 2010. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Medicine* 16, 90–97. <https://doi.org/10.1038/nm.2069>
- Cooper, G.M., 2000. *The cell: a molecular approach*, 2. ed. ed. ASM Press [u.a.], Washington, DC.
- Covarrubias, S., Robinson, E.K., Shapleigh, B., Vollmers, A., Katzman, S., Hanley, N., Fong, N., McManus, M.T., Carpenter, S., 2017. CRISPR/Cas-based screening of long non-coding RNAs (lncRNAs) in macrophages with an NF- $\kappa$ B reporter. *J. Biol. Chem.* 292, 20911–20920. <https://doi.org/10.1074/jbc.M117.799155>
- Crawford, D.R., Wang, Y., Schools, G.P., Kochheiser, J., Davies, K.J.A., 1997. Down-regulation of Mammalian Mitochondrial RNAs During Oxidative Stress. *Free Radical Biology and Medicine* 22, 551–559. [https://doi.org/10.1016/S0891-5849\(96\)00380-2](https://doi.org/10.1016/S0891-5849(96)00380-2)
- Criss, A.K., Ahlgren, D.M., Jou, T.S., McCormick, B.A., Casanova, J.E., 2001. The GTPase Rac1 selectively regulates *Salmonella* invasion at the apical plasma membrane of polarized epithelial cells. *Journal of Cell Science* 114, 1331–1341.
- Crooke, S.T., Witztum, J.L., Bennett, C.F., Baker, B.F., 2018. RNA-Targeted Therapeutics. *Cell Metabolism* 27, 714–739. <https://doi.org/10.1016/j.cmet.2018.03.004>
- Cros, J., Cagnard, N., Woollard, K., Patey, N., Zhang, S.-Y., Senechal, B., Puel, A., Biswas, S.K., Moshous, D., Picard, C., Jais, J.-P., D’Cruz, D., Casanova, J.-L., Trouillet, C., Geissmann, F., 2010. Human CD14<sup>dim</sup> Monocytes Patrol and Sense Nucleic Acids and Viruses via TLR7 and TLR8 Receptors. *Immunity* 33, 375–386. <https://doi.org/10.1016/j.immuni.2010.08.012>
- Cross, A.R., Baldwin, V.M., Roy, S., Essex-Lopresti, A.E., Prior, J.L., Harmer, N.J., 2019. Zoonoses under our noses. *Microbes and Infection* 21, 10–19. <https://doi.org/10.1016/j.micinf.2018.06.001>
- Cummings, B.S., Wills, L.P., Schnellmann, R.G., 2004. Measurement of Cell Death in Mammalian Cells. *Curr Protoc Pharmacol* 0 12. <https://doi.org/10.1002/0471141755.ph1208s25>
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K.J., Gall, A., Girón, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Marugán, J.C., Maurel, T., McMahon, A.C., Moore, B., Morales, J., Mudge, J.M., Nuhn, M., Ogeh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Yates,

- A.D., Zerbino, D.R., Flicek, P., 2019. Ensembl 2019. *Nucleic Acids Res* 47, D745–D751. <https://doi.org/10.1093/nar/gky1113>
- Cusick, J.K., Mustian, A., Goldberg, K., Reyland, M.E., 2010. RELT induces cellular death in HEK 293 epithelial cells. *Cellular Immunology* 261, 1–8. <https://doi.org/10.1016/j.cellimm.2009.10.013>
- Da Costa, J.P., Rodrigues, A.P.D., Farias, L.H.S., Frade, P.C.R., Da Silva, B.J.M., Do Nascimento, J.L.M., Silva, E.O., 2018. Biological effects of kojic acid on human monocytes in vitro. *Biomedicine & Pharmacotherapy* 101, 100–106. <https://doi.org/10.1016/j.biopha.2018.02.036>
- Daigneault, M., Preston, J.A., Marriott, H.M., Whyte, M.K.B., Dockrell, D.H., 2010. The Identification of Markers of Macrophage Differentiation in PMA-Stimulated THP-1 Cells and Monocyte-Derived Macrophages. *PLOS ONE* 5, e8668. <https://doi.org/10.1371/journal.pone.0008668>
- D’Arcy, M.S., 2019. Cell death: a review of the major forms of apoptosis, necrosis and autophagy. *Cell Biology International* 43, 582–592. <https://doi.org/10.1002/cbin.11137>
- Darnell, G.A., Schroder, W.A., Gardner, J., Harrich, D., Yu, H., Medcalf, R.L., Warrillow, D., Antalis, T.M., Sonza, S., Suhrbier, A., 2006. SerpinB2 Is an Inducible Host Factor Involved in Enhancing HIV-1 Transcription and Replication. *J. Biol. Chem.* 281, 31348–31358. <https://doi.org/10.1074/jbc.M604220200>
- David, D.J., Pagliuso, A., Radoshevich, L., Nahori, M.-A., Cossart, P., 2018. Lmo1656 is a secreted virulence factor of *Listeria monocytogenes* that interacts with the sorting nexin 6–BAR complex. *J. Biol. Chem.* 293, 9265–9276. <https://doi.org/10.1074/jbc.RA117.000365>
- David Powell, 2015. drpowell/degust v3.2.0. Zenodo. <https://doi.org/10.5281/zenodo.3258933>
- De Lappe, N., O’Connor, J., Garvey, P., McKeown, P., Cormican, M., 2015. Ciprofloxacin-Resistant *Shigella sonnei* Associated with Travel to India. *Emerg Infect Dis* 21, 894–896. <https://doi.org/10.3201/eid2105.141184>
- Defenouillère, Q., Fromont-Racine, M., 2017. The ribosome-bound quality control complex: from aberrant peptide clearance to proteostasis maintenance. *Curr Genet* 1–9. <https://doi.org/10.1007/s00294-017-0708-5>
- Del Prete, M.J., Robles, M.S., Guío, A., Martínez-A, C., Izquierdo, M., Garcia-Sanz, J.A., 2002. Degradation of cellular mRNA is a general early apoptosis-induced event. *The FASEB Journal* 16, 2003–2005. <https://doi.org/10.1096/fj.02-0392fje>
- D’Elia, R.V., Laws, T.R., Núñez, A., Taylor, C., Clark, G.C., 2015. Delayed presence of alternatively activated macrophages during a *Francisella tularensis* infection. *Microbial Pathogenesis* 78, 37–42. <https://doi.org/10.1016/j.micpath.2014.10.002>
- Deszo, E.L., Brake, D.K., Kelley, K.W., Freund, G.G., 2004. IL-4-dependent CD86 expression requires JAK/STAT6 activation and is negatively regulated by PKC $\delta$ . *Cellular Signalling* 16, 271–280. [https://doi.org/10.1016/S0898-6568\(03\)00137-2](https://doi.org/10.1016/S0898-6568(03)00137-2)
- Devaux, C.A., Mezouar, S., Mege, J.-L., 2019. The E-Cadherin Cleavage Associated to Pathogenic Bacteria Infections Can Favor Bacterial Invasion and Transmigration, Dysregulation of the Immune Response and Cancer Induction in Humans. *Front Microbiol* 10, 2598. <https://doi.org/10.3389/fmicb.2019.02598>
- Di Bacco, A., Gill, G., 2006. SUMO-specific proteases and the cell cycle. An essential role for SENP5 in cell proliferation. *Cell Cycle* 5, 2310–2313. <https://doi.org/10.4161/cc.5.20.3367>

- Di Bacco, A., Ouyang, J., Lee, H.-Y., Catic, A., Ploegh, H., Gill, G., 2006. The SUMO-specific protease SENP5 is required for cell division. *Mol. Cell. Biol.* 26, 4489–4498. <https://doi.org/10.1128/MCB.02301-05>
- Díaz-Villanueva, J.F., Díaz-Molina, R., García-González, V., 2015. Protein Folding and Mechanisms of Proteostasis. *Int J Mol Sci* 16, 17193–17230. <https://doi.org/10.3390/ijms160817193>
- Dick, M.S., Sborgi, L., Rühl, S., Hiller, S., Broz, P., 2016. ASC filament formation serves as a signal amplification mechanism for inflammasomes. *Nature Communications* 7, 11929. <https://doi.org/10.1038/ncomms11929>
- Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., Baker, C.N., Bower, L., Brown, J.M., Caddle, L.B., Chiani, F., Clary, D., Cleak, J., Daly, M.J., Denegre, J.M., Doe, B., Dolan, M.E., Edie, S.M., Fuchs, H., Gailus-Durner, V., Galli, A., Gambadoro, A., Gallegos, J., Guo, S., Horner, N.R., Hsu, C.-W., Johnson, S.J., Kalaga, S., Keith, L.C., Lanoue, L., Lawson, T.N., Lek, M., Mark, M., Marschall, S., Mason, J., McElwee, M.L., Newbigging, S., Nutter, L.M.J., Peterson, K.A., Ramirez-Solis, R., Rowland, D.J., Ryder, E., Samocha, K.E., Seavitt, J.R., Selloum, M., Szoke-Kovacs, Z., Tamura, M., Trainor, A.G., Tudose, I., Wakana, S., Warren, J., Wendling, O., West, D.B., Wong, L., Yoshiki, A., Wurst, W., MacArthur, D.G., Tocchini-Valentini, G.P., Gao, X., Flicek, P., Bradley, A., Skarnes, W.C., Justice, M.J., Parkinson, H.E., Moore, M., Wells, S., Braun, R.E., Svenson, K.L., Angelis, M.H. de, Herault, Y., Mohun, T., Mallon, A.-M., Henkelman, R.M., Brown, S.D.M., Adams, D.J., Lloyd, K.C.K., McKerlie, C., Beaudet, A.L., Bućan, M., Murray, S.A., 2016. High-throughput discovery of novel developmental phenotypes. *Nature* 537, 508–514. <https://doi.org/10.1038/nature19356>
- Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., Vance, R.E., 2013. The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *Cell Reports* 3, 1355–1361. <https://doi.org/10.1016/j.celrep.2013.05.009>
- Dlugosz, A., Muschiol, S., Zakikhany, K., Assadi, G., D'Amato, M., Lindberg, G., 2014. Human enteroendocrine cell responses to infection with *Chlamydia trachomatis*: a microarray study. *Gut Pathogens* 6, 24. <https://doi.org/10.1186/1757-4749-6-24>
- Doerflinger, M., Deng, Y., Whitney, P., Salvamoser, R., Engel, S., Kueh, A.J., Tai, L., Bachem, A., Gressier, E., Geoghegan, N.D., Wilcox, S., Rogers, K.L., Garnham, A.L., Dengler, M.A., Bader, S.M., Ebert, G., Pearson, J.S., Nardo, D.D., Wang, N., Yang, C., Pereira, M., Bryant, C.E., Strugnelli, R.A., Vince, J.E., Pellegrini, M., Strasser, A., Bedoui, S., Herold, M.J., 2020. Flexible Usage and Interconnectivity of Diverse Cell Death Pathways Protect against Intracellular Infection. *Immunity* 0. <https://doi.org/10.1016/j.immuni.2020.07.004>
- Doganay, M., Demiraslan, H., 2015. Human anthrax as a re-emerging disease. *Recent Pat Antiinfect Drug Discov* 10, 10–29. <https://doi.org/10.2174/1574891x10666150408162354>
- Drecktrah, D., Knodler, L.A., Ireland, R., Steele-Mortimer, O., 2006. The Mechanism of Salmonella Entry Determines the Vacuolar Environment and Intracellular Gene Expression. *Traffic* 7, 39–51. <https://doi.org/10.1111/j.1600-0854.2005.00360.x>
- Drevets, D.A., Canono, B.P., Leenen, P.J., Campbell, P.A., 1994. Gentamicin kills intracellular *Listeria monocytogenes*. *Infect. Immun.* 62, 2222.

- Du, J., Reeves, A.Z., Klein, J.A., Twedt, D.J., Knodler, L.A., Lesser, C.F., 2016. The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *PNAS* 113, 4794–4799. <https://doi.org/10.1073/pnas.1520699113>
- Du, K., Sharma, M., Lukacs, G.L., 2005. The  $\Delta$ F508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nat Struct Mol Biol* 12, 17–25. <https://doi.org/10.1038/nsmb882>
- Dun, Y., Mysona, B., Van Ells, T., Amarnath, L., Shamsul Ola, M., Ganapathy, V., Smith, S.B., 2006. Expression of the cystine-glutamate exchanger (xc<sup>-</sup>) in retinal ganglion cells and regulation by nitric oxide and oxidative stress. *Cell Tissue Res* 324, 189–202. <https://doi.org/10.1007/s00441-005-0116-x>
- Dunay, I.R., Gajurel, K., Dhakal, R., Liesenfeld, O., Montoya, J.G., 2018. Treatment of Toxoplasmosis: Historical Perspective, Animal Models, and Current Clinical Practice. *Clinical Microbiology Reviews* 31, e00057-17. <https://doi.org/10.1128/CMR.00057-17>
- Duplomb, L., Rivière, J., Jegou, G., Da Costa, R., Hammann, A., Racine, J., Schmitt, A., Droin, N., Capron, C., Gougerot-Pocidallo, M.-A., Dubrez, L., Aral, B., Lafon, A., Edery, P., Ghomid, J., Blair, E., El Chehadeh-Djebbar, S., Carmignac, V., Thevenon, J., Guy, J., Girodon, F., Bastie, J.-N., Delva, L., Faivre, L., Thauvin-Robinet, C., Solary, E., 2019. Serpin B1 defect and increased apoptosis of neutrophils in Cohen syndrome neutropenia. *J. Mol. Med.* 97, 633–645. <https://doi.org/10.1007/s00109-019-01754-4>
- Dupont, N., Lacas-Gervais, S., Bertout, J., Paz, I., Freche, B., Nhieu, G.T.V., Goot, F.G. van der, Sansonetti, P.J., Lafont, F., 2009. Shigella Phagocytic Vacuolar Membrane Remnants Participate in the Cellular Response to Pathogen Invasion and Are Regulated by Autophagy. *Cell Host & Microbe* 6, 137–149. <https://doi.org/10.1016/j.chom.2009.07.005>
- Eden, E., Lipson, D., Yogev, S., Yakhini, Z., 2007. Discovering Motifs in Ranked Lists of DNA Sequences. *PLOS Computational Biology* 3, e39. <https://doi.org/10.1371/journal.pcbi.0030039>
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48. <https://doi.org/10.1186/1471-2105-10-48>
- Edwards, A.M., Massey, R.C., 2011. Invasion of Human Cells by a Bacterial Pathogen. *J Vis Exp*. <https://doi.org/10.3791/2693>
- Edwards, J.A., Rockx-Brouwer, D., Nair, V., Celli, J., 2010. Restricted cytosolic growth of *Francisella tularensis* subsp. *tularensis* by IFN- $\gamma$  activation of macrophages. *Microbiology*, 156, 327–339. <https://doi.org/10.1099/mic.0.031716-0>
- Eisele, N.A., Ruby, T., Jacobson, A., Manzanillo, P.S., Cox, J.S., Lam, L., Mukundan, L., Chawla, A., Monack, D.M., 2013. Salmonella Require the Fatty Acid Regulator PPAR $\delta$  for the Establishment of a Metabolic Environment Essential for Long-Term Persistence. *Cell Host & Microbe* 14, 171–182. <https://doi.org/10.1016/j.chom.2013.07.010>
- El Mezayen, R., El Gazzar, M., Seeds, M.C., McCall, C.E., Dreskin, S.C., Nicolls, M.R., 2007. Endogenous signals released from necrotic cells augment inflammatory responses to bacterial endotoxin. *Immunology Letters* 111, 36–44. <https://doi.org/10.1016/j.imlet.2007.04.011>
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498. <https://doi.org/10.1038/35078107>

- Enache, O.M., Rendo, V., Abdusamad, M., Lam, D., Davison, D., Pal, S., Currimjee, N., Hess, J., Pantel, S., Nag, A., Thorner, A.R., Doench, J.G., Vazquez, F., Beroukhir, R., Golub, T.R., Ben-David, U., 2020. Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nature Genetics* 52, 662–668. <https://doi.org/10.1038/s41588-020-0623-4>
- Eriksson, S., Chambers, B.J., Rhen, M., 2003. Nitric Oxide Produced by Murine Dendritic Cells is Cytotoxic for Intracellular *Salmonella enterica* sv. Typhimurium. *Scandinavian Journal of Immunology* 58, 493–502. <https://doi.org/10.1046/j.1365-3083.2003.01330.x>
- Eske, K., Breitbach, K., Köhler, J., Wongprompitak, P., Steinmetz, I., 2009. Generation of murine bone marrow derived macrophages in a standardised serum-free cell culture system. *Journal of Immunological Methods* 342, 13–19. <https://doi.org/10.1016/j.jim.2008.11.011>
- Esteve-Codina, A., Arpi, O., Martínez-García, M., Pineda, E., Mallo, M., Gut, M., Carrato, C., Rovira, A., Lopez, R., Tortosa, A., Dabad, M., Barco, S.D., Heath, S., Bagué, S., Ribalta, T., Alameda, F., Iglesia, N. de la, Balaña, C., Group, on behalf of the G., 2017. A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples. *PLOS ONE* 12, e0170632. <https://doi.org/10.1371/journal.pone.0170632>
- Evans, R.J., Sundaramurthy, V., Frickel, E.-M., 2018. The Interplay of Host Autophagy and Eukaryotic Pathogens. *Front. Cell Dev. Biol.* 6. <https://doi.org/10.3389/fcell.2018.00118>
- Everman, J.L., Sajuthi, S., Saef, B., Rios, C., Stoner, A.M., Numata, M., Hu, D., Eng, C., Oh, S., Rodriguez-Santana, J., Vldar, E.K., Voelker, D.R., Burchard, E.G., Seibold, M.A., 2019. Functional genomics of CDHR3 confirms its role in HRV-C infection and childhood asthma exacerbations. *Journal of Allergy and Clinical Immunology* 144, 962–971. <https://doi.org/10.1016/j.jaci.2019.01.052>
- Ewels, P., Magnusson, M., Lundin, S., Källér, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Fabbro, C.D., Scalabrin, S., Morgante, M., Giorgi, F.M., 2013. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLOS ONE* 8, e85024. <https://doi.org/10.1371/journal.pone.0085024>
- Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., Sonenberg, N., 2011. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nature Structural & Molecular Biology* 18, 1211–1217. <https://doi.org/10.1038/nsmb.2149>
- Fang, J., Ying, H., Mao, T., Fang, Y., Lu, Y., Wang, H., Zang, I., Wang, Zhaofu, Lin, Y., Zhao, M., Luo, X., Wang, Zongyao, Zhang, Y., Zhang, C., Xiao, W., Wang, Y., Tan, W., Chen, Z., Lu, C., Atadja, P., Li, E., Zhao, K., Liu, J., Gu, J., 2017. Upregulation of CD11b and CD86 through LSD1 inhibition promotes myeloid differentiation and suppresses cell proliferation in human monocytic leukemia cells. *Oncotarget* 8, 85085–85101. <https://doi.org/10.18632/oncotarget.18564>
- Feeley, E.M., Pilla-Moffett, D.M., Zwack, E.E., Piro, A.S., Finethy, R., Kolb, J.P., Martinez, J., Brodsky, I.E., Coers, J., 2017. Galectin-3 directs antimicrobial guanylate binding proteins to vacuoles furnished with bacterial secretion systems. *PNAS* 114, E1698–E1706. <https://doi.org/10.1073/pnas.1615771114>

- Fernandes-Alnemri, T., Yu, J.-W., Datta, P., Wu, J., Alnemri, E.S., 2009. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* 458, 509–513. <https://doi.org/10.1038/nature07710>
- Fernandes-Alnemri, T., Yu, J.-W., Juliana, C., Solorzano, L., Kang, S., Wu, J., Datta, P., McCormick, M., Huang, L., McDermott, E., Eisenlohr, L., Landel, C.P., Alnemri, E.S., 2010. The AIM2 inflammasome is critical for innate immunity to *Francisella tularensis*. *Nat Immunol* 11, 385–393. <https://doi.org/10.1038/ni.1859>
- Fettucciari, K., Ponsini, P., Palumbo, C., Rosati, E., Mannucci, R., Bianchini, R., Modesti, A., Marconi, P., 2015. Macrophage induced gelsolin in response to Group B *Streptococcus* (GBS) infection. *Cellular Microbiology* 17, 79–104. <https://doi.org/10.1111/cmi.12338>
- Fields, K.A., Heinzen, R.A., Carabeo, R., 2011. The Obligate Intracellular Lifestyle. *Front Microbiol* 2. <https://doi.org/10.3389/fmicb.2011.00099>
- Figueiredo, P. de, Ficht, T.A., Rice-Ficht, A., Rossetti, C.A., Adams, L.G., 2015. Pathogenesis and Immunobiology of Brucellosis: Review of Brucella–Host Interactions. *The American Journal of Pathology* 185, 1505–1517. <https://doi.org/10.1016/j.ajpath.2015.03.003>
- Fisch, D., Bando, H., Clough, B., Hornung, V., Yamamoto, M., Shenoy, A.R., Frickel, E., 2019. Human GBP1 is a microbe-specific gatekeeper of macrophage apoptosis and pyroptosis. *EMBO J* 38. <https://doi.org/10.15252/embj.2018100926>
- Fischer, J., Gutiérrez, S., Ganesan, R., Calabrese, C., Ranjan, R., Cildir, G., Hos, N.J., Rybniker, J., Wolke, M., Fries, J.W.U., Tergaonkar, V., Plum, G., Antebi, A., Robinson, N., 2019. Leptin signaling impairs macrophage defenses against *Salmonella Typhimurium*. *PNAS* 116, 16551–16560. <https://doi.org/10.1073/pnas.1904885116>
- Fleischer, T.C., Yun, U.J., Ayer, D.E., 2003. Identification and Characterization of Three New Components of the mSin3A Corepressor Complex. *Mol. Cell. Biol.* 23, 3456–3467. <https://doi.org/10.1128/MCB.23.10.3456-3467.2003>
- Fletcher, K., Ulferts, R., Jacquin, E., Veith, T., Gammoh, N., Arasteh, J.M., Mayer, U., Carding, S.R., Wileman, T., Beale, R., Florey, O., 2018. The WD40 domain of ATG16L1 is required for its non-canonical role in lipidation of LC3 at single membranes. *The EMBO Journal* 37, e97840. <https://doi.org/10.15252/embj.201797840>
- Foley, J.E., Nieto, N.C., 2010. Tularemia. *Veterinary Microbiology, Zoonoses: Advances and Perspectives* 140, 332–338. <https://doi.org/10.1016/j.vetmic.2009.07.017>
- Fookes, M., Schroeder, G.N., Langridge, G.C., Blondel, C.J., Mammina, C., Connor, T.R., Seth-Smith, H., Vernikos, G.S., Robinson, K.S., Sanders, M., Petty, N.K., Kingsley, R.A., Bäuml, A.J., Nuccio, S.-P., Contreras, I., Santiviago, C.A., Maskell, D., Barrow, P., Humphrey, T., Nastasi, A., Roberts, M., Frankel, G., Parkhill, J., Dougan, G., Thomson, N.R., 2011. *Salmonella bongori* Provides Insights into the Evolution of the Salmonellae. *PLOS Pathogens* 7, e1002191. <https://doi.org/10.1371/journal.ppat.1002191>
- Foote, J.R., Patel, A.A., Yona, S., Segal, A.W., 2019. Variations in the Phagosomal Environment of Human Neutrophils and Mononuclear Phagocyte Subsets. *Front. Immunol.* 10, 188. <https://doi.org/10.3389/fimmu.2019.00188>
- Forestal, C.A., Malik, M., Catlett, S.V., Savitt, A.G., Benach, J.L., Sellati, T.J., Furie, M.B., 2007. *Francisella tularensis* Has a Significant Extracellular Phase in Infected Mice. *J Infect Dis* 196, 134–137. <https://doi.org/10.1086/518611>
- Fotis, C., Antoranz, A., Hatzivramidis, D., Sakellaropoulos, T., Alexopoulos, L.G., 2018. Network-based technologies for early drug discovery. *Drug Discovery Today* 23, 626–635. <https://doi.org/10.1016/j.drudis.2017.12.001>

- Franceschini, A., Meier, R., Casanova, A., Kreibich, S., Daga, N., Andritschke, D., Dilling, S., Rämö, P., Emmenlauer, M., Kaufmann, A., Conde-Álvarez, R., Low, S.H., Pelkmans, L., Helenius, A., Hardt, W.-D., Dehio, C., Mering, C. von, 2014. Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *PNAS* 111, 4548–4553.  
<https://doi.org/10.1073/pnas.1402353111>
- Franchi, L., Kanneganti, T.-D., Dubyak, G.R., Núñez, G., 2007. Differential Requirement of P2X7 Receptor and Intracellular K<sup>+</sup> for Caspase-1 Activation Induced by Intracellular and Extracellular Bacteria. *J. Biol. Chem.* 282, 18810–18818.  
<https://doi.org/10.1074/jbc.M610762200>
- Franklin-Murray, A.L., Mallya, S., Jankeel, A., Sureshchandra, S., Messaoudi, I., Lodoen, M.B., 2020. *Toxoplasma gondii* Dysregulates Barrier Function and Mechanotransduction Signaling in Human Endothelial Cells. *mSphere* 5.  
<https://doi.org/10.1128/mSphere.00550-19>
- French, P.J., Bijman, J., Edixhoven, M., Vaandrager, A.B., Scholte, B.J., Lohmann, S.M., Nairn, A.C., Jonge, H.R. de, 1995. Isotype-specific Activation of Cystic Fibrosis Transmembrane Conductance Regulator-Chloride Channels by cGMP-dependent Protein Kinase II. *J. Biol. Chem.* 270, 26626–26631.  
<https://doi.org/10.1074/jbc.270.44.26626>
- Freund, E.C., Lock, J.Y., Oh, J., Maculins, T., Delamarre, L., Bohlen, C.J., Haley, B., Murthy, A., 2020. Efficient gene knockout in primary human and murine myeloid cells by non-viral delivery of CRISPR-Cas9. *J Exp Med* 217.  
<https://doi.org/10.1084/jem.20191692>
- Friebel, A., Ilchmann, H., Aepfelbacher, M., Ehrbar, K., Machleidt, W., Hardt, W.-D., 2001. SopE and SopE2 from *Salmonella typhimurium* Activate Different Sets of RhoGTPases of the Host Cell. *J. Biol. Chem.* 276, 34035–34040.  
<https://doi.org/10.1074/jbc.M100609200>
- Fritah, S., Lhocine, N., Golebiowski, F., Mounier, J., Andrieux, A., Jouvion, G., Hay, R.T., Sansonetti, P., Dejean, A., 2014. Sumoylation controls host anti-bacterial response to the gut invasive pathogen *Shigella flexneri*. *EMBO reports* 15, 965–972.  
<https://doi.org/10.15252/embr.201338386>
- Fu, H., Doelling, J.H., Arendt, C.S., Hochstrasser, M., Vierstra, R.D., 1998. Molecular organization of the 20S proteasome gene family from *Arabidopsis thaliana*. *Genetics* 149, 677–692.
- Fu, J., Qi, L., Hu, M., Liu, Y., Yu, K., Liu, Q., Liu, X., 2017. *Salmonella* proteomics under oxidative stress reveals coordinated regulation of antioxidant defense with iron metabolism and bacterial virulence. *Journal of Proteomics* 157, 52–58.  
<https://doi.org/10.1016/j.jprot.2017.02.004>
- Fujita, N., Morita, E., Itoh, T., Tanaka, A., Nakaoka, M., Osada, Y., Umemoto, T., Saitoh, T., Nakatogawa, H., Kobayashi, S., Haraguchi, T., Guan, J.-L., Iwai, K., Tokunaga, F., Saito, K., Ishibashi, K., Akira, S., Fukuda, M., Noda, T., Yoshimori, T., 2013. Recruitment of the autophagic machinery to endosomes during infection is mediated by ubiquitin. *J Cell Biol* 203, 115–128.  
<https://doi.org/10.1083/jcb.201304188>
- Fujita, N., Saitoh, T., Kageyama, S., Akira, S., Noda, T., Yoshimori, T., 2009. Differential Involvement of Atg16L1 in Crohn Disease and Canonical Autophagy ANALYSIS OF THE ORGANIZATION OF THE Atg16L1 COMPLEX IN FIBROBLASTS. *J. Biol. Chem.* 284, 32602–32609. <https://doi.org/10.1074/jbc.M109.037671>

- Furne, J., Springfield, J., Koenig, T., DeMaster, E., Levitt, M.D., 2001. Oxidation of hydrogen sulfide and methanethiol to thiosulfate by rat tissues: a specialized function of the colonic mucosa. *Biochemical Pharmacology* 62, 255–259. [https://doi.org/10.1016/S0006-2952\(01\)00657-8](https://doi.org/10.1016/S0006-2952(01)00657-8)
- Gallego Romero, I., Pai, A.A., Tung, J., Gilad, Y., 2014. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology* 12, 42. <https://doi.org/10.1186/1741-7007-12-42>
- Garaude, J., Acín-Pérez, R., Martínez-Cano, S., Enamorado, M., Ugolini, M., Nistal-Villán, E., Hervás-Stubbs, S., Pelegrín, P., Sander, L.E., Enríquez, J.A., Sancho, D., 2016. Mitochondrial respiratory-chain adaptations in macrophages contribute to antibacterial host defense. *Nature Immunology* 17, 1037–1045. <https://doi.org/10.1038/ni.3509>
- García-Cattaneo, A., Gobert, F.-X., Müller, M., Toscano, F., Flores, M., Lescure, A., Nery, E.D., Benaroch, P., 2012. Cleavage of Toll-like receptor 3 by cathepsins B and H is essential for signaling. *PNAS* 109, 9053–9058. <https://doi.org/10.1073/pnas.1115091109>
- Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., Mazloom, A.R., Ma'ayan, A., Chua, W.-J., Hansen, T.H., Turley, S.J., Merad, M., Randolph, G.J., Consortium, the I.G., Gautier, E.L., Jakubzick, C., Randolph, G.J., Best, A.J., Knell, J., Goldrath, A., Miller, J., Brown, B., Merad, M., Jojic, V., Koller, D., Cohen, N., Brennan, P., Brenner, M., Shay, T., Regev, A., Fletcher, A., Elpek, K., Bellemare-Pelletier, A., Malhotra, D., Turley, S., Jianu, R., Laidlaw, D., Collins, J., Narayan, K., Sylvia, K., Kang, J., Gazit, R., Garrison, B.S., Rossi, D.J., Kim, F., Rao, T.N., Wagers, A., Shinton, S.A., Hardy, R.R., Monach, P., Bezman, N.A., Sun, J.C., Kim, C.C., Lanier, L.L., Heng, T., Kreslavsky, T., Painter, M., Ericson, J., Davis, S., Mathis, D., Benoist, C., 2012. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature Immunology* 13, 1118–1128. <https://doi.org/10.1038/ni.2419>
- Gaviria-Agudelo, C., Carter, K., Tareen, N., Pascual, V., Copley, L.A., 2014. Gene Expression Analysis of Children with Acute Hematogenous Osteomyelitis Caused by Methicillin-Resistant *Staphylococcus aureus*: Correlation with Clinical Severity of Illness. *PLOS ONE* 9, e103523. <https://doi.org/10.1371/journal.pone.0103523>
- Ge, Y., Jiang, W., Gan, L., Wang, L., Sun, C., Ni, P., Liu, Y., Wu, S., Gu, L., Zheng, W., Lund, F.E., Xin, H.-B., 2010. Mouse embryonic fibroblasts from CD38 knockout mice are resistant to oxidative stresses through inhibition of reactive oxygen species production and Ca<sup>2+</sup> overload. *Biochemical and Biophysical Research Communications* 399, 167–172. <https://doi.org/10.1016/j.bbrc.2010.07.040>
- Geier, H., Celli, J., 2011. Phagocytic Receptors Dictate Phagosomal Escape and Intracellular Proliferation of *Francisella tularensis*. *Infection and Immunity* 79, 2204–2214. <https://doi.org/10.1128/IAI.01382-10>
- Gershlick, D.C., Schindler, C., Chen, Y., Bonifacino, J.S., 2016. TSSC1 is novel component of the endosomal retrieval machinery. *Mol Biol Cell* 27, 2867–2878. <https://doi.org/10.1091/mbc.E16-04-0209>
- Gharib, S.A., McMahan, R.S., Eddy, W.E., Long, M.E., Parks, W.C., Aitken, M.L., Manicone, A.M., 2019. Transcriptional and functional diversity of human macrophage repolarization. *Journal of Allergy and Clinical Immunology* 143, 1536–1548. <https://doi.org/10.1016/j.jaci.2018.10.046>

- Ghasemi, A., Zahediasl, S., 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab* 10, 486–489.  
<https://doi.org/10.5812/ijem.3505>
- Gibbs, K.D., Washington, E.J., Jaslow, S.L., Bourgeois, J.S., Foster, M.W., Guo, R., Brennan, R.G., Ko, D.C., 2020. The Salmonella Secreted Effector SarA/SteE Mimics Cytokine Receptor Signaling to Activate STAT3. *Cell Host & Microbe* 27, 129-139.e4.  
<https://doi.org/10.1016/j.chom.2019.11.012>
- Gillies, P.J., Bhatia, S.K., Belcher, L.A., Hannon, D.B., Thompson, J.T., Heuvel, J.P.V., 2012. Regulation of inflammatory and lipid metabolism genes by eicosapentaenoic acid-rich oil. *J. Lipid Res.* 53, 1679–1689. <https://doi.org/10.1194/jlr.M022657>
- Girardin, S.E., Boneca, I.G., Carneiro, L.A.M., Antignac, A., Jéhanno, M., Viala, J., Tedin, K., Taha, M.-K., Labigne, A., Zähringer, U., Coyle, A.J., DiStefano, P.S., Bertin, J., Sansonetti, P.J., Philpott, D.J., 2003a. Nod1 Detects a Unique Muropeptide from Gram-Negative Bacterial Peptidoglycan. *Science* 300, 1584–1587.  
<https://doi.org/10.1126/science.1084677>
- Girardin, S.E., Boneca, I.G., Viala, J., Chamaillard, M., Labigne, A., Thomas, G., Philpott, D.J., Sansonetti, P.J., 2003b. Nod2 Is a General Sensor of Peptidoglycan through Muramyl Dipeptide (MDP) Detection. *J. Biol. Chem.* 278, 8869–8872.  
<https://doi.org/10.1074/jbc.C200651200>
- Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., Ruepp, A., 2019. CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res* 47, D559–D563.  
<https://doi.org/10.1093/nar/gky973>
- Gluschko, A., Herb, M., Wiegmann, K., Krut, O., Neiss, W.F., Utermöhlen, O., Krönke, M., Schramm, M., 2018. The  $\beta$ 2 Integrin Mac-1 Induces Protective LC3-Associated Phagocytosis of *Listeria monocytogenes*. *Cell Host & Microbe* 23, 324-337.e5.  
<https://doi.org/10.1016/j.chom.2018.01.018>
- Gogoi, M., Ravikumar, V., Dixit, N.M., Chakravorty, D., 2018. Salmonella escapes antigen presentation through K63 ubiquitination mediated endosomal proteolysis of MHC II via modulation of endosomal acidification in dendritic cells. *Pathog Dis* 76.  
<https://doi.org/10.1093/femspd/ftx125>
- Gong, L., Yeh, E.T.H., 2006. Characterization of a Family of Nucleolar SUMO-specific Proteases with Preference for SUMO-2 or SUMO-3. *J. Biol. Chem.* 281, 15869–15877. <https://doi.org/10.1074/jbc.M511658200>
- González-Pérez, C.J., Tanori-Cordova, J., Aispuro-Hernández, E., Vargas-Arispuro, I., Martínez-Téllez, M.A., 2019. Morphometric parameters of foodborne related-pathogens estimated by transmission electron microscopy and their relation to optical density and colony forming units. *Journal of Microbiological Methods* 165, 105691. <https://doi.org/10.1016/j.mimet.2019.105691>
- Gordon, M.A., 2008. Salmonella infections in immunocompromised adults. *Journal of Infection* 56, 413–422. <https://doi.org/10.1016/j.jinf.2008.03.012>
- Gordon, S., 2016. Phagocytosis: An Immunobiologic Process. *Immunity* 44, 463–475.  
<https://doi.org/10.1016/j.immuni.2016.02.026>
- Goudot, C., Coillard, A., Villani, A.-C., Gueguen, P., Cros, A., Sarkizova, S., Tang-Huau, T.-L., Bohec, M., Baulande, S., Hacohen, N., Amigorena, S., Segura, E., 2017. Aryl Hydrocarbon Receptor Controls Monocyte Differentiation into Dendritic Cells versus Macrophages. *Immunity* 47, 582-596.e6.  
<https://doi.org/10.1016/j.immuni.2017.08.016>

- Goulet, D.R., Atkins, W.M., 2020. Considerations for the Design of Antibody-Based Therapeutics. *JPharmSci* 109, 74–103. <https://doi.org/10.1016/j.xphs.2019.05.031>
- Gow, M., Mirembe, D., Longwe, Z., Pickard, B.S., 2013. A gene trap mutagenesis screen for genes underlying cellular response to the mood stabilizer lithium. *Journal of Cellular and Molecular Medicine* 17, 657–663. <https://doi.org/10.1111/jcmm.12048>
- Grandjean, T., Boucher, A., Thepaut, M., Monlezun, L., Guery, B., Faudry, E., Kipnis, E., Dessein, R., 2017. The human NAIP-NLRC4-inflammasome senses the *Pseudomonas aeruginosa* T3SS inner-rod protein. *Int Immunol* 29, 377–384. <https://doi.org/10.1093/intimm/dxx047>
- Grandvaux, N., Servant, M.J., tenOever, B., Sen, G.C., Balachandran, S., Barber, G.N., Lin, R., Hiscott, J., 2002. Transcriptional Profiling of Interferon Regulatory Factor 3 Target Genes: Direct Involvement in the Regulation of Interferon-Stimulated Genes. *Journal of Virology* 76, 5532–5539. <https://doi.org/10.1128/JVI.76.11.5532-5539.2002>
- Griffin, A.J., Crane, D.D., Wehrly, T.D., Scott, D.P., Bosio, C.M., 2013. Alternative Activation of Macrophages and Induction of Arginase Are Not Components of Pathogenesis Mediated by *Francisella* Species. *PLOS ONE* 8, e82096. <https://doi.org/10.1371/journal.pone.0082096>
- Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J.V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway, M.B., Eldred, J.M., Miller, C.A., Subramanian, J., Govindan, R., Kumar, R.D., Bose, R., Ding, L., Walker, J.R., Larson, D.E., Dooling, D.J., Smith, S.M., Ley, T.J., Mardis, E.R., Wilson, R.K., 2013. DGIdb: mining the druggable genome. *Nature Methods* 10, 1209–1210. <https://doi.org/10.1038/nmeth.2689>
- Grobarczyk, B., Franco, B., Hanon, K., Malgrange, B., 2015. Generation of Isogenic Human iPS Cell Line Precisely Corrected by Genome Editing Using the CRISPR/Cas9 System. *Stem Cell Rev and Rep* 11, 774–787. <https://doi.org/10.1007/s12015-015-9600-1>
- Groß, C.J., Mishra, R., Schneider, K.S., Médard, G., Wettmarshausen, J., Dittlein, D.C., Shi, H., Gorka, O., Koenig, P.-A., Fromm, S., Magnani, G., Ćiković, T., Hartjes, L., Smollich, J., Robertson, A.A.B., Cooper, M.A., Schmidt-Supprian, M., Schuster, M., Schroder, K., Broz, P., Traidl-Hoffmann, C., Beutler, B., Kuster, B., Ruland, J., Schneider, S., Perocchi, F., Groß, O., 2016. K<sup>+</sup> Efflux-Independent NLRP3 Inflammasome Activation by Small Molecules Targeting Mitochondria. *Immunity* 45, 761–773. <https://doi.org/10.1016/j.immuni.2016.08.010>
- Guarás, A., Perales-Clemente, E., Calvo, E., Acín-Pérez, R., Loureiro-Lopez, M., Pujol, C., Martínez-Carrascoso, I., Nuñez, E., García-Marqués, F., Rodríguez-Hernández, M.A., Cortés, A., Diaz, F., Pérez-Martos, A., Moraes, C.T., Fernández-Silva, P., Trifunovic, A., Navas, P., Vazquez, J., Enríquez, J.A., 2016. The CoQH<sub>2</sub>/CoQ Ratio Serves as a Sensor of Respiratory Chain Efficiency. *Cell Reports* 15, 197–209. <https://doi.org/10.1016/j.celrep.2016.03.009>
- Gudi, T., Lohmann, S.M., Pilz, R.B., 1997. Regulation of gene expression by cyclic GMP-dependent protein kinase requires nuclear translocation of the kinase: identification of a nuclear localization signal. *Mol. Cell. Biol.* 17, 5244–5254. <https://doi.org/10.1128/MCB.17.9.5244>
- Gudipaty, S.A., Rosenblatt, J., 2017. Epithelial cell extrusion: Pathways and pathologies. *Seminars in Cell & Developmental Biology, Extracellular Vesicles* 67, 132–140. <https://doi.org/10.1016/j.semcd.2016.05.010>
- Guest, C.B., Deszo, E.L., Hartman, M.E., York, J.M., Kelley, K.W., Freund, G.G., 2008. Ca<sup>2+</sup>/Calmodulin-Dependent Kinase Kinase  $\alpha$  Is Expressed by Monocytic Cells and

- Regulates the Activation Profile. *PLOS ONE* 3, e1606.  
<https://doi.org/10.1371/journal.pone.0001606>
- Guilliams, M., van de Laar, L., 2015. A Hitchhiker's Guide to Myeloid Cell Subsets: Practical Implementation of a Novel Mononuclear Phagocyte Classification System. *Front. Immunol.* 6. <https://doi.org/10.3389/fimmu.2015.00406>
- Guinn, Z.P., Petro, T.M., 2019. Interferon regulatory factor 3 plays a role in macrophage responses to interferon- $\gamma$ . *Immunobiology* 224, 565–574.  
<https://doi.org/10.1016/j.imbio.2019.04.004>
- Guo, X., Wang, H., Li, Y., Leng, X., Huang, W., Ma, Y., Xu, T., Qi, X., 2019. Transfection reagent Lipofectamine triggers type I interferon signaling activation in macrophages. *Immunology & Cell Biology* 97, 92–96.  
<https://doi.org/10.1111/imcb.12194>
- Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., Taipale, J., 2018. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine* 24, 927–930. <https://doi.org/10.1038/s41591-018-0049-z>
- Hall, J.D., Craven, R.R., Fuller, J.R., Pickles, R.J., Kawula, T.H., 2007. *Francisella tularensis* Replicates within Alveolar Type II Epithelial Cells In Vitro and In Vivo following Inhalation. *Infection and Immunity* 75, 1034–1039.  
<https://doi.org/10.1128/IAI.01254-06>
- Hall, J.D., Woolard, M.D., Gunn, B.M., Craven, R.R., Taft-Benz, S., Frelinger, J.A., Kawula, T.H., 2008. Infected-Host-Cell Repertoire and Cellular Response in the Lung following Inhalation of *Francisella tularensis* Schu S4, LVS, or U112. *Infection and Immunity* 76, 5843–5852. <https://doi.org/10.1128/IAI.01176-08>
- Han, J., Pluhackova, K., Böckmann, R.A., 2017. The Multifaceted Role of SNARE Proteins in Membrane Fusion. *Front. Physiol.* 8. <https://doi.org/10.3389/fphys.2017.00005>
- Handa, Y., Suzuki, M., Ohya, K., Iwai, H., Ishijima, N., Koleske, A.J., Fukui, Y., Sasakawa, C., 2007. *Shigella* IpgB1 promotes bacterial entry through the ELMO–Dock180 machinery. *Nature Cell Biology* 9, 121–128. <https://doi.org/10.1038/ncb1526>
- Haney, M.S., Bohlen, C.J., Morgens, D.W., Ousey, J.A., Barkal, A.A., Tsui, C.K., Ego, B., Levin, R., Kamber, R., Collins, H., Tucker, A., Li, A., Vorselen, D., Labitigan, L., Crane, E., Boyle, E., Jiang, L., Chan, J., Rincón, E., Greenleaf, W.J., Li, B., Snyder, M.P., Weissman, I.L., Theriot, J.A., Collins, S.R., Barres, B.A., Bassik, M.C., 2018. Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat Genet* 50, 1716–1727. <https://doi.org/10.1038/s41588-018-0254-1>
- Hass, R., Giese, G., Meyer, G., Hartmann, A., Dörk, T., Köhler, L., Resch, K., Traub, P., Goppelt-Strübe, M., 1990. Differentiation and retrodifferentiation of U937 cells: reversible induction and suppression of intermediate filament protein synthesis. *Eur. J. Cell Biol.* 51, 265–271.
- Hattula, K., Furuholm, J., Tikkanen, J., Tanhuanpää, K., Laakkonen, P., Peränen, J., 2006. Characterization of the Rab8-specific membrane traffic route linked to protrusion formation. *Journal of Cell Science* 119, 4866–4877.  
<https://doi.org/10.1242/jcs.03275>
- Hawkey, J., Monk, J.M., Billman-Jacobe, H., Palsson, B., Holt, K.E., 2020. Impact of insertion sequences on convergent evolution of *Shigella* species. *PLOS Genetics* 16, e1008931. <https://doi.org/10.1371/journal.pgen.1008931>
- Hawkins, J.L., Uknalis, J., Oscar, T.P., Schwarz, J.G., Vimini, B., Parveen, S., 2019. The Effect of Previous Life Cycle Phase on the Growth Kinetics, Morphology, and Antibiotic Resistance of *Salmonella Typhimurium* DT104 in Brain Heart Infusion and Ground Chicken Extract. *Front. Microbiol.* 10. <https://doi.org/10.3389/fmicb.2019.01043>

- Hedl, M., Yan, J., Witt, H., Abraham, C., 2019. IRF5 Is Required for Bacterial Clearance in Human M1-Polarized Macrophages, and IRF5 Immune-Mediated Disease Risk Variants Modulate This Outcome. *The Journal of Immunology* 202, 920–930. <https://doi.org/10.4049/jimmunol.1800226>
- Hefele, M., Stolzer, I., Ruder, B., He, G.-W., Mahapatro, M., Wirtz, S., Neurath, M.F., Günther, C., 2018. Intestinal epithelial Caspase-8 signaling is essential to prevent necroptosis during *Salmonella Typhimurium* induced enteritis. *Mucosal Immunology* 11, 1191–1202. <https://doi.org/10.1038/s41385-018-0011-x>
- Heijden, J. van der, Bosman, E.S., Reynolds, L.A., Finlay, B.B., 2015. Direct measurement of oxidative and nitrosative stress dynamics in *Salmonella* inside macrophages. *PNAS* 112, 560–565. <https://doi.org/10.1073/pnas.1414569112>
- Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A.A., Mann, M., 2015. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* 163, 712–723. <https://doi.org/10.1016/j.cell.2015.09.053>
- Helaine, S., Cheverton, A.M., Watson, K.G., Faure, L.M., Matthews, S.A., Holden, D.W., 2014. Internalization of *Salmonella* by Macrophages Induces Formation of Nonreplicating Persisters. *Science* 343, 204. <https://doi.org/10.1126/science.1244705>
- Helaine, S., Thompson, J.A., Watson, K.G., Liu, M., Boyle, C., Holden, D.W., 2010. Dynamics of intracellular bacterial replication at the single cell level. *PNAS* 107, 3746–3751. <https://doi.org/10.1073/pnas.1000041107>
- Hermansson, A.-K., Paciello, I., Bernardini, M.L., 2016. The Orchestra and Its Maestro: *Shigella's* Fine-Tuning of the Inflammasome Platforms, in: Backert, S. (Ed.), *Inflammasome Signaling and Bacterial Infections*, Current Topics in Microbiology and Immunology. Springer International Publishing, Cham, pp. 91–115. [https://doi.org/10.1007/978-3-319-41171-2\\_5](https://doi.org/10.1007/978-3-319-41171-2_5)
- Hersh, D., Monack, D.M., Smith, M.R., Ghori, N., Falkow, S., Zychlinsky, A., 1999. The *Salmonella* invasin SipB induces macrophage apoptosis by binding to caspase-1. *PNAS* 96, 2396–2401. <https://doi.org/10.1073/pnas.96.5.2396>
- Herwig, R., Hardt, C., Lienhard, M., Kamburov, A., 2016. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature Protocols* 11, 1889–1907. <https://doi.org/10.1038/nprot.2016.117>
- Hida, T., Ikeda, H., Kametaka, S., Akazawa, C., Kohsaka, S., Ebisu, S., Uchiyama, Y., Waguri, S., 2007. Specific depletion of GGA2 causes cathepsin D missorting in HeLa cells. *Archives of Histology and Cytology* 70, 303–312. <https://doi.org/10.1679/aohc.70.303>
- Hoffman, G.E., Roussos, P., 2020. dream: Powerful differential expression analysis for repeated measures designs. *bioRxiv* 432567. <https://doi.org/10.1101/432567>
- Holt, O., Kanno, E., Bossi, G., Booth, S., Daniele, T., Santoro, A., Arico, M., Saegusa, C., Fukuda, M., Griffiths, G.M., 2008. Slp1 and Slp2-a Localize to the Plasma Membrane of CTL and Contribute to Secretion from the Immunological Synapse. *Traffic* 9, 446–457. <https://doi.org/10.1111/j.1600-0854.2008.00714.x>
- Hölzer, M., Marz, M., 2019. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *Gigascience* 8. <https://doi.org/10.1093/gigascience/giz039>
- Homer, C.R., Richmond, A.L., Rebert, N.A., Achkar, J.-P., McDonald, C., 2010. ATG16L1 and NOD2 Interact in an Autophagy-Dependent Antibacterial Pathway Implicated in

- Crohn's Disease Pathogenesis. *Gastroenterology* 139, 1630-1641.e2.  
<https://doi.org/10.1053/j.gastro.2010.07.006>
- Hong, E.P., Go, M.J., Kim, H.-L., Park, J.W., 2017. Risk prediction of pulmonary tuberculosis using genetic and conventional risk factors in adult Korean population. *PLoS ONE* 12, e0174642. <https://doi.org/10.1371/journal.pone.0174642>
- Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., Fitzgerald, K.A., 2009. AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* 458, 514–518.  
<https://doi.org/10.1038/nature07725>
- Hornung, V., Bauernfeind, F., Halle, A., Samstad, E.O., Kono, H., Rock, K.L., Fitzgerald, K.A., Latz, E., 2008. Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunology* 9, 847–856.  
<https://doi.org/10.1038/ni.1631>
- Huang, B., Zhang, Q., Yuan, Y., Xin, N., He, K., Huang, Y., Tang, H., Gong, P., 2018. Sema3a inhibits the differentiation of Raw264.7 cells to osteoclasts under 2Gy radiation by reducing inflammation. *PLOS ONE* 13, e0200000.  
<https://doi.org/10.1371/journal.pone.0200000>
- Huang, J., Canadien, V., Lam, G.Y., Steinberg, B.E., Dinauer, M.C., Magalhaes, M.A.O., Glogauer, M., Grinstein, S., Brumell, J.H., 2009. Activation of antibacterial autophagy by NADPH oxidases. *PNAS* 106, 6226–6231.  
<https://doi.org/10.1073/pnas.0811045106>
- Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., Ideker, T., 2018. Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *cells* 6, 484-495.e5. <https://doi.org/10.1016/j.cels.2018.03.001>
- Huang, K.-C., Yang, K.-C., Lin, H., Tsao Tsun-Hui, T., Lee, W.-K., Lee, S.-A., Kao, C.-Y., 2013. Analysis of schizophrenia and hepatocellular carcinoma genetic network with corresponding modularity and pathways: novel insights to the immune system. *BMC Genomics* 14, S10. <https://doi.org/10.1186/1471-2164-14-S5-S10>
- Huang, T., Huang, X., Shi, B., Wang, F., Feng, W., Yao, M., 2018. Regulators of Salmonella-host interaction identified by peripheral blood transcriptome profiling: roles of TGFβ1 and TRP53 in intracellular Salmonella replication in pigs. *Veterinary Research* 49, 121. <https://doi.org/10.1186/s13567-018-0616-9>
- Hubber, A., Kubori, T., Coban, C., Matsuzawa, T., Ogawa, M., Kawabata, T., Yoshimori, T., Nagai, H., 2017. Bacterial secretion system skews the fate of Legionella -containing vacuoles towards LC3-associated phagocytosis. *Scientific Reports* 7, 44795.  
<https://doi.org/10.1038/srep44795>
- Huebener, P., Pradere, J.-P., Hernandez, C., Gwak, G.-Y., Caviglia, J.M., Mu, X., Loike, J.D., Jenkins, R.E., Antoine, D.J., Schwabe, R.F., 2015. The HMGB1/RAGE axis triggers neutrophil-mediated injury amplification following necrosis. *J Clin Invest* 125, 539–550. <https://doi.org/10.1172/JCI76887>
- Hultquist, J.F., Schumann, K., Woo, J.M., Manganaro, L., McGregor, M.J., Doudna, J., Simon, V., Krogan, N.J., Marson, A., 2016. A Cas9 Ribonucleoprotein Platform for Functional Genetic Studies of HIV-Host Interactions in Primary Human T Cells. *Cell Rep* 17, 1438–1452. <https://doi.org/10.1016/j.celrep.2016.09.080>
- Hume, P.J., Singh, V., Davidson, A.C., Koronakis, V., 2017. Swiss Army Pathogen: The Salmonella Entry Toolkit. *Front Cell Infect Microbiol* 7.  
<https://doi.org/10.3389/fcimb.2017.00348>

- Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P., Cunningham, F., 2018. Ensembl variation resources. Database (Oxford) 2018. <https://doi.org/10.1093/database/bay119>
- Hutchings, M.I., Truman, A.W., Wilkinson, B., 2019. Antibiotics: past, present and future. *Current Opinion in Microbiology, Antimicrobials* 51, 72–80. <https://doi.org/10.1016/j.mib.2019.10.008>
- Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., Szpyt, J., Tam, S., Zarraga, G., Pontano-Vaites, L., Swarup, S., White, A.E., Schweppe, D.K., Rad, R., Erickson, B.K., Obar, R.A., Guruharsha, K.G., Li, K., Artavanis-Tsakonas, S., Gygi, S.P., Harper, J.W., 2017. Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509. <https://doi.org/10.1038/nature22366>
- Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G.R., Dolmetsch, R., Kaykas, A., 2018. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature Medicine* 24, 939–946. <https://doi.org/10.1038/s41591-018-0050-6>
- Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., Fukase, K., Inamura, S., Kusumoto, S., Hashimoto, M., Foster, S.J., Moran, A.P., Fernandez-Luna, J.L., Nuñez, G., 2003. Host Recognition of Bacterial Muramyl Dipeptide Mediated through NOD2 IMPLICATIONS FOR CROHN'S DISEASE. *J. Biol. Chem.* 278, 5509–5512. <https://doi.org/10.1074/jbc.C200673200>
- Irving, A.T., Mimuro, H., Kufer, T.A., Lo, C., Wheeler, R., Turner, L.J., Thomas, B.J., Malosse, C., Gantier, M.P., Casillas, L.N., Votta, B.J., Bertin, J., Boneca, I.G., Sasakawa, C., Philpott, D.J., Ferrero, R.L., Kaparakis-Liaskos, M., 2014. The Immune Receptor NOD1 and Kinase RIP2 Interact with Bacterial Peptidoglycan on Early Endosomes to Promote Autophagy and Inflammatory Signaling. *Cell Host & Microbe* 15, 623–635. <https://doi.org/10.1016/j.chom.2014.04.001>
- Isakov, E., Weisman-Shomer, P., Benhar, M., 2014. Suppression of the pro-inflammatory NLRP3/interleukin-1 $\beta$  pathway in macrophages by the thioredoxin reductase inhibitor auranofin. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1840, 3153–3161. <https://doi.org/10.1016/j.bbagen.2014.07.012>
- Jangra, R.K., Herbert, A.S., Li, R., Jae, L.T., Kleinfelter, L.M., Slough, M.M., Barker, S.L., Guardado-Calvo, P., Román-Sosa, G., Dieterle, M.E., Kuehne, A.I., Muenz, N.A., Wirchnianski, A.S., Nyakatura, E.K., Fels, J.M., Ng, M., Mittler, E., Pan, J., Bharrhan, S., Wec, A.Z., Lai, J.R., Sidhu, S.S., Tischler, N.D., Rey, F.A., Moffat, J., Brummelkamp, T.R., Wang, Z., Dye, J.M., Chandran, K., 2018. Protocadherin-1 is essential for cell entry by New World hantaviruses. *Nature* 563, 559–563. <https://doi.org/10.1038/s41586-018-0702-1>
- Jeng, E.E., Bhadkamkar, V., Ibe, N.U., Gause, H., Jiang, L., Chan, J., Jian, R., Jimenez-Morales, D., Stevenson, E., Krogan, N.J., Swaney, D.L., Snyder, M.P., Mukherjee, S., Bassik, M.C., 2019. Systematic Identification of Host Cell Regulators of Legionella pneumophila Pathogenesis Using a Genome-wide CRISPR Screen. *Cell Host & Microbe* 26, 551-563.e6. <https://doi.org/10.1016/j.chom.2019.08.017>
- Jennings, E., Thurston, T.L.M., Holden, D.W., 2017. Salmonella SPI-2 Type III Secretion System Effectors: Molecular Mechanisms And Physiological Consequences. *Cell Host & Microbe* 22, 217–231. <https://doi.org/10.1016/j.chom.2017.07.009>

- Jensen, T.J., Loo, M.A., Pind, S., Williams, D.B., Goldberg, A.L., Riordan, J.R., 1995. Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* 83, 129–135. [https://doi.org/10.1016/0092-8674\(95\)90241-4](https://doi.org/10.1016/0092-8674(95)90241-4)
- Jha, A.K., Huang, S.C.-C., Sergushichev, A., Lampropoulou, V., Ivanova, Y., Loginicheva, E., Chmielewski, K., Stewart, K.M., Ashall, J., Everts, B., Pearce, E.J., Driggers, E.M., Artyomov, M.N., 2015. Network Integration of Parallel Metabolic and Transcriptional Data Reveals Metabolic Modules that Regulate Macrophage Polarization. *Immunity* 42, 419–430. <https://doi.org/10.1016/j.immuni.2015.02.005>
- Jia, Q., Horwitz, M.A., 2018. Live Attenuated Tularemia Vaccines for Protection Against Respiratory Challenge With Virulent *F. tularensis* subsp. *tularensis*. *Front. Cell. Infect. Microbiol.* 8. <https://doi.org/10.3389/fcimb.2018.00154>
- Jin, M.S., Kim, S.E., Heo, J.Y., Lee, M.E., Kim, H.M., Paik, S.-G., Lee, H., Lee, J.-O., 2007. Crystal Structure of the TLR1-TLR2 Heterodimer Induced by Binding of a Tri-Acylated Lipopeptide. *Cell* 130, 1071–1082. <https://doi.org/10.1016/j.cell.2007.09.008>
- Jin, X., Kruth, H.S., 2016. Culture of Macrophage Colony-stimulating Factor Differentiated Human Monocyte-derived Macrophages. *J Vis Exp.* <https://doi.org/10.3791/54244>
- Jin, Z., Wei, W., Yang, M., Du, Y., Wan, Y., 2014. Mitochondrial Complex I Activity Suppresses Inflammation and Enhances Bone Resorption by Shifting Macrophage-Osteoclast Polarization. *Cell Metabolism* 20, 483–498. <https://doi.org/10.1016/j.cmet.2014.07.011>
- Johnson, J.L., Ellis, B.A., Noack, D., Seabra, M.C., Catz, S.D., 2005. The Rab27a-binding protein, JFC1, regulates androgen-dependent secretion of prostate-specific antigen and prostatic-specific acid phosphatase1. *Biochem J* 391, 699–710. <https://doi.org/10.1042/BJ20050380>
- Jones, B.D., Ghorri, N., Falkow, S., 1994. Salmonella typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *J Exp Med* 180, 15–23. <https://doi.org/10.1084/jem.180.1.15>
- Jorgensen, I., Rayamajhi, M., Miao, E.A., 2017. Programmed cell death as a defence against infection. *Nature Reviews Immunology* 17, 151–164. <https://doi.org/10.1038/nri.2016.147>
- Jorgensen, I., Zhang, Y., Krantz, B.A., Miao, E.A., 2016. Pyroptosis triggers pore-induced intracellular traps (PITs) that capture bacteria and lead to their clearance by efferocytosis. *J Exp Med* 213, 2113–2128. <https://doi.org/10.1084/jem.20151613>
- Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., Cristea, I.M., 2013. The functional interactome landscape of the human histone deacetylase family. *Molecular Systems Biology* 9, 672. <https://doi.org/10.1038/msb.2013.26>
- Kagan, J.C., Su, T., Horng, T., Chow, A., Akira, S., Medzhitov, R., 2008. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon- $\beta$ . *Nature Immunology* 9, 361–368. <https://doi.org/10.1038/ni1569>
- Kageyama, S., Omori, H., Saitoh, T., Sone, T., Guan, J.-L., Akira, S., Imamoto, F., Noda, T., Yoshimori, T., 2011. The LC3 recruitment mechanism is separate from Atg9L1-dependent membrane formation in the autophagic response against Salmonella. *Mol. Biol. Cell* 22, 2290–2300. <https://doi.org/10.1091/mbc.E10-11-0893>
- Kajaste-Rudnitski, A., Naldini, L., 2015. Cellular Innate Immunity and Restriction of Viral Infection: Implications for Lentiviral Gene Therapy in Human Hematopoietic Cells. *Human Gene Therapy* 26, 201–209. <https://doi.org/10.1089/hum.2015.036>

- Kamburov, A., Stelzl, U., Lehrach, H., Herwig, R., 2013. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 41, D793-800.  
<https://doi.org/10.1093/nar/gks1055>
- Kampmann, M., Horlbeck, M.A., Chen, Y., Tsai, J.C., Bassik, M.C., Gilbert, L.A., Villalta, J.E., Kwon, S.C., Chang, H., Kim, V.N., Weissman, J.S., 2015. Next-generation libraries for robust RNA interference-based genome-wide screens. *PNAS* 112, E3384–E3391.  
<https://doi.org/10.1073/pnas.1508821112>
- Kamynina, E., Stover, P.J., 2017. The Roles of SUMO in Metabolic Regulation. *Adv. Exp. Med. Biol.* 963, 143–168. [https://doi.org/10.1007/978-3-319-50044-7\\_9](https://doi.org/10.1007/978-3-319-50044-7_9)
- Kang, H.S., Park, E.K., Kim, K.H., Park, J.-Y., Choi, J.-Y., Shin, H.-I., Jun, C.-D., Kang, S.-S., Kim, S.-Y., 2004. Receptor Activator of Nuclear Factor- $\kappa$ B Is Induced by a Rottlerin-sensitive and p38 MAP Kinase-dependent Pathway during Monocyte Differentiation. *Molecules and Cells* 17, 438–445.
- Kang, J.Y., Nan, X., Jin, M.S., Youn, S.-J., Ryu, Y.H., Mah, S., Han, S.H., Lee, H., Paik, S.-G., Lee, J.-O., 2009. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. *Immunity* 31, 873–884.  
<https://doi.org/10.1016/j.immuni.2009.09.018>
- Karlinsey, J.E., Stepien, T.A., Mayho, M., Singletary, L.A., Bingham-Ramos, L.K., Brehm, M.A., Greiner, D.L., Shultz, L.D., Gallagher, L.A., Bawn, M., Kingsley, R.A., Libby, S.J., Fang, F.C., 2019. Genome-wide Analysis of *Salmonella enterica* serovar Typhi in Humanized Mice Reveals Key Virulence Features. *Cell Host & Microbe* 26, 426-434.e6. <https://doi.org/10.1016/j.chom.2019.08.001>
- Karlsson, E.K., Harris, J.B., Tabrizi, S., Rahman, A., Shlyakhter, I., Patterson, N., O’Dushlaine, C., Schaffner, S.F., Gupta, S., Chowdhury, F., Sheikh, A., Shin, O.S., Ellis, C., Becker, C.E., Stuart, L.M., Calderwood, S.B., Ryan, E.T., Qadri, F., Sabeti, P.C., LaRocque, R.C., 2013. Natural Selection in a Bangladeshi Population from the Cholera-Endemic Ganges River Delta. *Science Translational Medicine* 5, 192ra86-192ra86.  
<https://doi.org/10.1126/scitranslmed.3006338>
- Kasper, C.A., Sorg, I., Schmutz, C., Tschon, T., Wischnewski, H., Kim, M.L., Arrieumerlou, C., 2010. Cell-Cell Propagation of NF- $\kappa$ B Transcription Factor and MAP Kinase Activation Amplifies Innate Immunity against Bacterial Infection. *Immunity* 33, 804–816. <https://doi.org/10.1016/j.immuni.2010.10.015>
- Kaufmann, S.H.E., Dorhoi, A., Hotchkiss, R.S., Bartenschlager, R., 2017. Host-directed therapies for bacterial and viral infections. *Nat Rev Drug Discov* advance online publication. <https://doi.org/10.1038/nrd.2017.162>
- Kaushik, S., Cuervo, A.M., 2015. Proteostasis and aging. *Nat Med* 21, 1406–1415.  
<https://doi.org/10.1038/nm.4001>
- Kayagaki, N., Stowe, I.B., Lee, B.L., O’Rourke, K., Anderson, K., Warming, S., Cuellar, T., Haley, B., Roose-Girma, M., Phung, Q.T., Liu, P.S., Lill, J.R., Li, H., Wu, J., Kummerfeld, S., Zhang, J., Lee, W.P., Snipas, S.J., Salvesen, G.S., Morris, L.X., Fitzgerald, L., Zhang, Y., Bertram, E.M., Goodnow, C.C., Dixit, V.M., 2015. Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* 526, 666–671. <https://doi.org/10.1038/nature15541>
- Ke, Z.-B., Cai, H., Wu, Y.-P., Lin, Y.-Z., Li, X.-D., Huang, J.-B., Sun, X.-L., Zheng, Q.-S., Xue, X.-Y., Wei, Y., Xu, N., 2019. Identification of key genes and pathways in benign prostatic hyperplasia. *Journal of Cellular Physiology* 234, 19942–19950.  
<https://doi.org/10.1002/jcp.28592>

- Kelly, R.D.W., Cowley, S.M., 2013. The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochemical Society Transactions* 41, 741–749. <https://doi.org/10.1042/BST20130010>
- Kent, W.J., 2002. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* 12, 656–664. <https://doi.org/10.1101/gr.229202>
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, and D., 2002. The Human Genome Browser at UCSC. *Genome Res.* 12, 996–1006. <https://doi.org/10.1101/gr.229102>
- Khan, R., Sancho-Shimizu, V., Prendergast, C., Roy, M.-F., Loredano-Osti, J.-C., Malo, D., 2012. Refinement of the genetics of the host response to Salmonella infection in MOLF/Ei: regulation of type 1 IFN and TRP3 pathways by Ity2. *Genes & Immunity* 13, 175–183. <https://doi.org/10.1038/gene.2011.69>
- Kiama, S.G., Dreher, D., Cochand, L., Kok, M., Obregon, C., Nicod, L., Gehr, P., 2006. Host cell responses of Salmonella typhimurium infected human dendritic cells. *Immunology & Cell Biology* 84, 475–481. <https://doi.org/10.1111/j.1440-1711.2006.01461.x>
- Kim, B.-H., Shenoy, A.R., Kumar, P., Das, R., Tiwari, S., MacMicking, J.D., 2011. A Family of IFN- $\gamma$ -Inducible 65-kD GTPases Protects Against Bacterial Infection. *Science* 332, 717–721. <https://doi.org/10.1126/science.1201711>
- Kim, C.C., Monack, D., Falkow, S., 2003. Modulation of Virulence by Two Acidified Nitrite-Responsive Loci of Salmonella enterica Serovar Typhimurium. *Infection and Immunity* 71, 3196–3205. <https://doi.org/10.1128/IAI.71.6.3196-3205.2003>
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kim, H.Y., Kim, K.S., Kim, M.J., Kim, H.-S., Lee, K.-Y., Kang, K.W., 2019. Auranofin Inhibits RANKL-Induced Osteoclastogenesis by Suppressing Inhibitors of  $\kappa$ B Kinase and Inflammasome-Mediated Interleukin-1 $\beta$  Secretion [WWW Document]. *Oxidative Medicine and Cellular Longevity*. <https://doi.org/10.1155/2019/3503912>
- Kim, S.-Y., Yasuda, S., Tanaka, H., Yamagata, K., Kim, H., 2011. Non-clustered protocadherin. *Cell Adhesion & Migration* 5, 97–105. <https://doi.org/10.4161/cam.5.2.14374>
- Kirschke, H., 2013. Chapter 408 - Cathepsin H, in: Rawlings, N.D., Salvesen, G. (Eds.), *Handbook of Proteolytic Enzymes (Third Edition)*. Academic Press, pp. 1795–1800. <https://doi.org/10.1016/B978-0-12-382219-2.00408-7>
- Kissing, S., Saftig, P., Haas, A., 2018. Vacuolar ATPase in phago(lyso)some biology. *International Journal of Medical Microbiology, Intracellular Compartments as Places of Pathogen-Host Interaction* 308, 58–67. <https://doi.org/10.1016/j.ijmm.2017.08.007>
- Klein, E.Y., Boeckel, T.P.V., Martinez, E.M., Pant, S., Gandra, S., Levin, S.A., Goossens, H., Laxminarayan, R., 2018. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *PNAS* 115, E3463–E3470. <https://doi.org/10.1073/pnas.1717295115>
- Knodler, L.A., 2015. Salmonella enterica: living a double life in epithelial cells. *Current Opinion in Microbiology, Host-microbe interactions: bacteria • Genomics* 23, 23–31. <https://doi.org/10.1016/j.mib.2014.10.010>
- Knodler, L.A., Crowley, S.M., Sham, H.P., Yang, H., Wrands, M., Ma, C., Ernst, R.K., Steele-Mortimer, O., Celli, J., Vallance, B.A., 2014. Noncanonical Inflammasome Activation

- of Caspase-4/Caspase-11 Mediates Epithelial Defenses against Enteric Bacterial Pathogens. *Cell Host & Microbe* 16, 249–256.  
<https://doi.org/10.1016/j.chom.2014.07.002>
- Knodler, L.A., Vallance, B.A., Celli, J., Winfree, S., Hansen, B., Montero, M., Steele-Mortimer, O., 2010. Dissemination of invasive *Salmonella* via bacterial-induced extrusion of mucosal epithelia. *PNAS* 107, 17733–17738.  
<https://doi.org/10.1073/pnas.1006098107>
- Ko, D.C., Shukla, K.P., Fong, C., Wasnick, M., Brittnacher, M.J., Wurfel, M.M., Holden, T.D., O’Keefe, G.E., Yserloo, B.V., Akey, J.M., Miller, S.I., 2009. A Genome-wide In Vitro Bacterial-Infection Screen Reveals Human Variation in the Host Response Associated with Inflammatory Disease. *The American Journal of Human Genetics* 85, 214–227. <https://doi.org/10.1016/j.ajhg.2009.07.012>
- Kobayashi, T., Ogawa, M., Sanada, T., Mimuro, H., Kim, M., Ashida, H., Akakura, R., Yoshida, M., Kawalec, M., Reichhart, J.-M., Mizushima, T., Sasakawa, C., 2013. The *Shigella* OspC3 effector inhibits caspase-4, antagonizes inflammatory cell death, and promotes epithelial infection. *Cell Host Microbe* 13, 570–583.  
<https://doi.org/10.1016/j.chom.2013.04.012>
- Koch-Edelmann, S., Banhart, S., Saied, E.M., Rose, L., Aeberhard, L., Laue, M., Doellinger, J., Arenz, C., Heuer, D., 2017. The cellular ceramide transport protein CERT promotes *Chlamydia psittaci* infection and controls bacterial sphingolipid uptake. *Cellular Microbiology* 19, e12752. <https://doi.org/10.1111/cmi.12752>
- Kochi, L.T., Fernandes, L.G.V., Souza, G.O., Vasconcellos, S.A., Heinemann, M.B., Romero, E.C., Kirchgatter, K., Nascimento, A.L.T.O., 2019. The interaction of two novel putative proteins of *Leptospira interrogans* with E-cadherin, plasminogen and complement components with potential role in bacterial infection. *Virulence* 10, 734–753. <https://doi.org/10.1080/21505594.2019.1650613>
- Kofoed, E.M., Vance, R.E., 2011. Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* 477, 592–595.  
<https://doi.org/10.1038/nature10394>
- Kopp, F., Mendell, J.T., 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393–407. <https://doi.org/10.1016/j.cell.2018.01.011>
- Kortmann, J., Brubaker, S.W., Monack, D.M., 2015. Cutting Edge: Inflammasome Activation in Primary Human Macrophages Is Dependent on Flagellin. *The Journal of Immunology* 195, 815–819. <https://doi.org/10.4049/jimmunol.1403100>
- Köster, S., Upadhyay, S., Chandra, P., Papavinasasundaram, K., Yang, G., Hassan, A., Grigsby, S.J., Mittal, E., Park, H.S., Jones, V., Hsu, F.-F., Jackson, M., Sasseti, C.M., Philips, J.A., 2017. *Mycobacterium tuberculosis* is protected from NADPH oxidase and LC3-associated phagocytosis by the LCP protein CpsA. *PNAS* 114, E8711–E8720.  
<https://doi.org/10.1073/pnas.1707792114>
- Koutsoumanis, K.P., Lianou, A., 2013. Stochasticity in Colonial Growth Dynamics of Individual Bacterial Cells. *Appl. Environ. Microbiol.* 79, 2294–2301.  
<https://doi.org/10.1128/AEM.03629-12>
- Kravets, E., Degrandi, D., Ma, Q., Peulen, T.-O., Klümpers, V., Felekyan, S., Kühnemuth, R., Weidtkamp-Peters, S., Seidel, C.A., Pfeffer, K., 2016. Guanylate binding proteins directly attack *Toxoplasma gondii* via supramolecular complexes. *eLife* 5, e11479.  
<https://doi.org/10.7554/eLife.11479>
- Kumar, D.K.V., Choi, S.H., Washicosky, K.J., Eimer, W.A., Tucker, S., Ghofrani, J., Lefkowitz, A., McColl, G., Goldstein, L.E., Tanzi, R.E., Moir, R.D., 2016. Amyloid- $\beta$  peptide protects against microbial infection in mouse and worm models of Alzheimer’s

- disease. *Science Translational Medicine* 8, 340ra72-340ra72.  
<https://doi.org/10.1126/scitranslmed.aaf1059>
- Kumar, M., Sulfikar, Chaminda, T., Patel, A.K., Sewwandi, H., Mazumder, P., Joshi, M., Honda, R., 2020. Prevalence of antibiotic resistance in the tropical rivers of Sri Lanka and India. *Environmental Research* 188, 109765.  
<https://doi.org/10.1016/j.envres.2020.109765>
- Kurz, A.R.M., Pruenster, M., Rohwedder, I., Ramadass, M., Schäfer, K., Harrison, U., Gouveia, G., Nussbaum, C., Immler, R., Wiessner, J.R., Margraf, A., Lim, D.-S., Walzog, B., Dietzel, S., Moser, M., Klein, C., Vestweber, D., Haas, R., Catz, S.D., Sperandio, M., 2016. MST1-dependent vesicle trafficking regulates neutrophil transmigration through the vascular basement membrane. *J Clin Invest* 126, 4125–4139. <https://doi.org/10.1172/JCI87043>
- Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., Valen, E., 2016. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* 44, W272–W276. <https://doi.org/10.1093/nar/gkw398>
- Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., Valen, E., 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* 47, W171–W174. <https://doi.org/10.1093/nar/gkz365>
- LaFleur, M.W., Nguyen, T.H., Coxe, M.A., Yates, K.B., Trombley, J.D., Weiss, S.A., Brown, F.D., Gillis, J.E., Coxe, D.J., Doench, J.G., Haining, W.N., Sharpe, A.H., 2019. A CRISPR-Cas9 delivery system for in vivo screening of genes in the immune system. *Nature Communications* 10, 1–10. <https://doi.org/10.1038/s41467-019-09656-2>
- Lahiri, A., Das, P., Chakravorty, D., 2008. Arginase modulates Salmonella induced nitric oxide production in RAW264.7 macrophages and is required for Salmonella pathogenesis in mice model of infection. *Microbes and Infection* 10, 1166–1174. <https://doi.org/10.1016/j.micinf.2008.06.008>
- Lai, A.C., Crews, C.M., 2017. Induced protein degradation: an emerging drug discovery paradigm. *Nature Reviews Drug Discovery* 16, 101–114. <https://doi.org/10.1038/nrd.2016.211>
- Lam, G.Y., Cemma, M., Muise, A.M., Higgins, D.E., Brumell, J.H., 2013. Host and bacterial factors that regulate LC3 recruitment to *Listeria monocytogenes* during the early stages of macrophage infection. *Autophagy* 9, 985–995. <https://doi.org/10.4161/auto.24406>
- Lanier, L.L., Warner, N.L., 1981. Paraformaldehyde fixation of hematopoietic cells for quantitative flow cytometry (FACS) analysis. *Journal of Immunological Methods* 47, 25–30. [https://doi.org/10.1016/0022-1759\(81\)90253-2](https://doi.org/10.1016/0022-1759(81)90253-2)
- Lapaquette, P., Glasser, A.-L., Huett, A., Xavier, R.J., Darfeuille-Michaud, A., 2009. Crohn's disease-associated adherent-invasive *E. coli* are selectively favoured by impaired autophagy to replicate intracellularly. *Cellular Microbiology* 12, 99–113. <https://doi.org/10.1111/j.1462-5822.2009.01381.x>
- Lapuente-Brun, E., Moreno-Loshuertos, R., Acín-Pérez, R., Latorre-Pellicer, A., Colás, C., Balsa, E., Perales-Clemente, E., Quirós, P.M., Calvo, E., Rodríguez-Hernández, M.A., Navas, P., Cruz, R., Carracedo, Á., López-Otín, C., Pérez-Martos, A., Fernández-Silva, P., Fernández-Vizarra, E., Enríquez, J.A., 2013. Supercomplex Assembly Determines Electron Flux in the Mitochondrial Electron Transport Chain. *Science* 340, 1567–1570. <https://doi.org/10.1126/science.1230381>
- LaRock, D.L., Chaudhary, A., Miller, S.I., 2015. Salmonellae interactions with host processes. *Nature Reviews Microbiology* 13, 191–205. <https://doi.org/10.1038/nrmicro3420>

- Lathrop, S.K., Binder, K.A., Starr, T., Cooper, K.G., Chong, A., Carmody, A.B., Steele-Mortimer, O., 2015. Replication of *Salmonella enterica* Serovar Typhimurium in Human Monocyte-Derived Macrophages. *Infection and Immunity* 83, 2661–2671. <https://doi.org/10.1128/IAI.00033-15>
- Lathrop, S.K., Cooper, K.G., Binder, K.A., Starr, T., Mampilli, V., Detweiler, C.S., Steele-Mortimer, O., 2018. *Salmonella* Typhimurium Infection of Human Monocyte-Derived Macrophages. *Curr Protoc Microbiol* 50, e56. <https://doi.org/10.1002/cpmc.56>
- Laughlin, R.C., Knodler, L.A., Barhoumi, R., Payne, H.R., Wu, J., Gomez, G., Pugh, R., Lawhon, S.D., Bäumlner, A.J., Steele-Mortimer, O., Adams, L.G., 2014. Spatial Segregation of Virulence Gene Expression during Acute Enteric Infection with *Salmonella enterica* serovar Typhimurium. *mBio* 5. <https://doi.org/10.1128/mBio.00946-13>
- Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* 15, R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
- Law, H.T., Lin, A.E.-J., Kim, Y., Quach, B., Nano, F.E., Guttman, J.A., 2011. *Francisella tularensis* Uses Cholesterol and Clathrin-Based Endocytic Mechanisms to Invade Hepatocytes. *Scientific Reports* 1, 1–11. <https://doi.org/10.1038/srep00192>
- Leber, J.H., Crimmins, G.T., Raghavan, S., Meyer-Morse, N.P., Cox, J.S., Portnoy, D.A., 2008. Distinct TLR- and NLR-Mediated Transcriptional Responses to an Intracellular Pathogen. *PLOS Pathogens* 4, e6. <https://doi.org/10.1371/journal.ppat.0040006>
- Lee, H.C., 2006. Structure and Enzymatic Functions of Human CD38. *Mol Med* 12, 317–323. <https://doi.org/10.2119/2006-00086.Lee>
- Lee, H.-J., Woo, Y., Hahn, T.-W., Jung, Y.M., Jung, Y.-J., 2020. Formation and Maturation of the Phagosome: A Key Mechanism in Innate Immunity against Intracellular Bacterial Infection. *Microorganisms* 8, 1298. <https://doi.org/10.3390/microorganisms8091298>
- Lee, S.-M., Suk, K., Lee, W.-H., 2015. Myristoylated alanine-rich C kinase substrate (MARCKS) regulates the expression of proinflammatory cytokines in macrophages through activation of p38/JNK MAPK and NF- $\kappa$ B. *Cellular Immunology* 296, 115–121. <https://doi.org/10.1016/j.cellimm.2015.04.004>
- Lehtinen, J., Nuutila, J., Lilius, E.-M., 2004. Green fluorescent protein–propidium iodide (GFP-PI) based assay for flow cytometric measurement of bacterial viability. *Cytometry Part A* 60A, 165–172. <https://doi.org/10.1002/cyto.a.20026>
- Levenson, E.A., Martens, C., Kanakabandi, K., Turner, C., Virtaneva, K., Paneru, M., Ricklefs, S., Sosnovtsev, S.V., Johnson, J.A., Porcella, S.F., Green, K.Y., 2018. Comparative transcriptomic response of primary and immortalized macrophage to murine norovirus infection. *J Immunol* 200, 4157–4169. <https://doi.org/10.4049/jimmunol.1700384>
- Levitt, M.D., Furne, J., Springfield, J., Suarez, F., DeMaster, E., 1999. Detoxification of hydrogen sulfide and methanethiol in the cecal mucosa. *J Clin Invest* 104, 1107–1114. <https://doi.org/10.1172/JCI7712>
- Lewerenz, J., Hewett, S.J., Huang, Y., Lambros, M., Gout, P.W., Kalivas, P.W., Massie, A., Smolders, I., Methner, A., Pergande, M., Smith, S.B., Ganapathy, V., Maher, P., 2013. The Cystine/Glutamate Antiporter System xc<sup>-</sup> in Health and Disease: From Molecular Mechanisms to Novel Therapeutic Opportunities. *Antioxid Redox Signal* 18, 522–555. <https://doi.org/10.1089/ars.2011.4391>
- Lewis, K., 2020. The Science of Antibiotic Discovery. *Cell* 181, 29–45. <https://doi.org/10.1016/j.cell.2020.02.056>

- Li, B., Clohisey, S.M., Chia, B.S., Wang, B., Cui, A., Eisenhaure, T., Schweitzer, L.D., Hoover, P., Parkinson, N.J., Nachshon, A., Smith, N., Regan, T., Farr, D., Gutmann, M.U., Bukhari, S.I., Law, A., Sangesland, M., Gat-Viks, I., Digard, P., Vasudevan, S., Lingwood, D., Dockrell, D.H., Doench, J.G., Baillie, J.K., Hacohen, N., 2020. Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nat Commun* 11. <https://doi.org/10.1038/s41467-019-13965-x>
- Li, C., Lee, J., Ding, J., Sun, S., 2018. Integrative analysis of gene expression and methylation data for breast cancer cell lines. *BioData Mining* 11, 13. <https://doi.org/10.1186/s13040-018-0174-8>
- Li, G.H., Arora, P.D., Chen, Y., McCulloch, C.A., Liu, P., 2012. Multifunctional roles of gelsolin in health and diseases. *Medicinal Research Reviews* 32, 999–1025. <https://doi.org/10.1002/med.20231>
- Li, H., Liu, C.-C., Zheng, H., Huang, T.Y., 2018. Amyloid, tau, pathogen infection and antimicrobial protection in Alzheimer's disease -conformist, nonconformist, and realistic prospects for AD pathogenesis. *Transl Neurodegener* 7, 34. <https://doi.org/10.1186/s40035-018-0139-3>
- Li, H., Marshall, Z.M., Whorton, A.R., 1999. Stimulation of cystine uptake by nitric oxide: regulation of endothelial cell glutathione levels. *American Journal of Physiology-Cell Physiology* 276, C803–C811. <https://doi.org/10.1152/ajpcell.1999.276.4.C803>
- Li, H., Zhou, H., Wang, D., Qiu, J., Zhou, Y., Li, X., Rosenfeld, M.G., Ding, S., Fu, X.-D., 2012. Versatile pathway-centric approach based on high-throughput sequencing to anticancer drug discovery. *PNAS* 109, 4609–4614. <https://doi.org/10.1073/pnas.1200305109>
- Li, H.-Z., Zhang, L., Chen, J.-X., Zheng, Y., Zhu, X.-N., 2017. Silver-containing dressing for surgical site infection in clean and clean-contaminated operations: a systematic review and meta-analysis of randomized controlled trials. *Journal of Surgical Research* 215, 98–107. <https://doi.org/10.1016/j.jss.2017.03.040>
- Li, P., Jiang, W., Yu, Q., Liu, W., Zhou, P., Li, J., Xu, J., Xu, B., Wang, F., Shao, F., 2017. Ubiquitination and degradation of GBPs by a Shigella effector to suppress host defence. *Nature* 551, 378–383. <https://doi.org/10.1038/nature24467>
- Li, P., Li, M., Lindberg, M.R., Kennett, M.J., Xiong, N., Wang, Y., 2010. PAD4 is essential for antibacterial innate immunity mediated by neutrophil extracellular traps/PAD4 in NET-mediated bacterial killing. *J Exp Med* 207, 1853–1862. <https://doi.org/10.1084/jem.20100239>
- Li, Q., Sodroski, C., Lowey, B., Schweitzer, C.J., Cha, H., Zhang, F., Liang, T.J., 2016. Hepatitis C virus depends on E-cadherin as an entry factor and regulates its expression in epithelial-to-mesenchymal transition. *Proc Natl Acad Sci USA* 113, 7620–7625. <https://doi.org/10.1073/pnas.1602701113>
- Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., Lopez, R., 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* 43, W580–W584. <https://doi.org/10.1093/nar/gkv279>
- Li, X., Massa, P.E., Hanidu, A., Peet, G.W., Aro, P., Savitt, A., Mische, S., Li, J., Marcu, K.B., 2002. IKK $\alpha$ , IKK $\beta$ , and NEMO/IKK $\gamma$  Are Each Required for the NF- $\kappa$ B-mediated Inflammatory Response Program. *J. Biol. Chem.* 277, 45129–45140. <https://doi.org/10.1074/jbc.M205165200>
- Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., Chesnut, J.D., 2015. Rapid and highly

- efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology* 208, 44–53. <https://doi.org/10.1016/j.jbiotec.2015.04.024>
- Liao, G., Driel, B. van, Magelky, E., O’Keeffe, M.S., Malefyt, R. de W., Engel, P., Herzog, R.W., Mizoguchi, E., Bhan, A.K., Terhorst, C., 2014. Glucocorticoid-induced TNF receptor family-related protein ligand regulates the migration of monocytes to the inflamed intestine. *The FASEB Journal* 28, 474–484. <https://doi.org/10.1096/fj.13-236505>
- Liao, Y., Shi, W., 2019. Read trimming is not required for mapping and quantification of RNA-seq reads. *bioRxiv* 833962. <https://doi.org/10.1101/833962>
- Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Roy, N.V., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Peer, Y.V. de, Tavernier, J., Callewaert, N., 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications* 5, ncomms5767. <https://doi.org/10.1038/ncomms5767>
- Lindgren, H., Golovliov, I., Baranov, V., Ernst, R.K., Telepnev, M., Sjöstedt, A., 2004. Factors affecting the escape of *Francisella tularensis* from the phagolysosome. *Journal of Medical Microbiology*, 53, 953–958. <https://doi.org/10.1099/jmm.0.45685-0>
- Lindholm, D., Korhonen, L., Eriksson, O., Köks, S., 2017. Recent Insights into the Role of Unfolded Protein Response in ER Stress in Health and Disease. *Front. Cell Dev. Biol.* 5. <https://doi.org/10.3389/fcell.2017.00048>
- Liss, V., Hensel, M., 2015. Take the tube: remodelling of the endosomal system by intracellular *Salmonella enterica*. *Cellular Microbiology* 17, 639–647. <https://doi.org/10.1111/cmi.12441>
- Liu, J., Akahoshi, T., Namai, R., Matsui, T., Kondo, H., 2000. Effect of auranofin, an antirheumatic drug, on neutrophil apoptosis. *Inflamm res.* 49, 445–451. <https://doi.org/10.1007/s000110050615>
- Liu, R., Holik, A.Z., Su, S., Jansz, N., Chen, K., Leong, H.S., Blewitt, M.E., Asselin-Labat, M.-L., Smyth, G.K., Ritchie, M.E., 2015. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res* 43, e97–e97. <https://doi.org/10.1093/nar/gkv412>
- Liu, Y.-Y., Wang, Y., Walsh, T.R., Yi, L.-X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, L.-F., Gu, D., Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, J.-H., Shen, J., 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 16, 161–168. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7)
- Llewellyn, A.C., Jones, C.L., Napier, B.A., Bina, J.E., Weiss, D.S., 2011. Macrophage Replication Screen Identifies a Novel *Francisella* Hydroperoxide Resistance Protein Involved in Virulence. *PLOS ONE* 6, e24201. <https://doi.org/10.1371/journal.pone.0024201>
- Llufrio, E.M., Wang, L., Naser, F.J., Patti, G.J., 2018. Sorting cells alters their redox state and cellular metabolome. *Redox Biology* 16, 381–387. <https://doi.org/10.1016/j.redox.2018.03.004>
- Long, M., Zhao, J., Li, T., Tafalla, C., Zhang, Q., Wang, X., Gong, X., Shen, Z., Li, A., 2015. Transcriptomic and proteomic analyses of splenic immune mechanisms of rainbow trout (*Oncorhynchus mykiss*) infected by *Aeromonas salmonicida* subsp. *salmonicida*. *Journal of Proteomics* 122, 41–54. <https://doi.org/10.1016/j.jprot.2015.03.031>
- Lopez-Fabuel, I., Douce, J.L., Logan, A., James, A.M., Bonvento, G., Murphy, M.P., Almeida, A., Bolaños, J.P., 2016. Complex I assembly into supercomplexes determines

- differential mitochondrial ROS production in neurons and astrocytes. *PNAS* 113, 13063–13068. <https://doi.org/10.1073/pnas.1613701113>
- López-Montero, N., Ramos-Marquès, E., Risco, C., Portillo, F.G., 2016. Intracellular Salmonella induces autophagy of host endomembranes in persistent infections. *Autophagy* 12, 1886–1901. <https://doi.org/10.1080/15548627.2016.1208888>
- Loprasert, S., Whangsuk, W., Sallabhan, R., Mongkolsuk, S., 2003. Regulation of the katG-dpsA operon and the importance of KatG in survival of Burkholderia pseudomallei exposed to oxidative stress. *FEBS Letters* 542, 17–21. [https://doi.org/10.1016/S0014-5793\(03\)00328-4](https://doi.org/10.1016/S0014-5793(03)00328-4)
- Low, E.N.D., Mokhtar, N.M., Wong, Z., Raja Ali, R.A., 2019. Colonic Mucosal Transcriptomic Changes in Patients with Long-Duration Ulcerative Colitis Revealed Colitis-Associated Cancer Pathways. *J Crohns Colitis* 13, 755–763. <https://doi.org/10.1093/ecco-jcc/jjz002>
- Lu, J., Holmgren, A., 2014. The thioredoxin antioxidant system. *Free Radical Biology and Medicine, Antioxidants* 66, 75–87. <https://doi.org/10.1016/j.freeradbiomed.2013.07.036>
- Lu, R., Pitha, P.M., 2001. Monocyte Differentiation to Macrophage Requires Interferon Regulatory Factor 7. *J. Biol. Chem.* 276, 45491–45496. <https://doi.org/10.1074/jbc.C100421200>
- Lucarelli, M., 2017. New era of cystic fibrosis: Full mutational analysis and personalized therapy. *World Journal of Medical Genetics* 7, 1–9. <https://doi.org/10.5496/wjmg.v7.i1.1>
- Luciani, A., Vilella, V.R., Vasaturo, A., Giardino, I., Raia, V., Pettoello-Mantovani, M., D'Apolito, M., Guido, S., Leal, T., Quarantino, S., Maiuri, L., 2009. SUMOylation of Tissue Transglutaminase as Link between Oxidative Stress and Inflammation. *The Journal of Immunology* 183, 2775–2784. <https://doi.org/10.4049/jimmunol.0900993>
- Lv, J., He, X., Wang, H., Wang, Z., Kelly, G.T., Wang, X., Chen, Y., Wang, T., Qian, Z., 2017. TLR4-NOX2 axis regulates the phagocytosis and killing of Mycobacterium tuberculosis by macrophages. *BMC Pulmonary Medicine* 17, 194. <https://doi.org/10.1186/s12890-017-0517-0>
- Ma, J., Chen, C., Barth, A.S., Cheadle, C., Guan, X., Gao, L., 2015. Lysosome and Cytoskeleton Pathways Are Robustly Enriched in the Blood of Septic Patients: A Meta-Analysis of Transcriptomic Data. *Mediators Inflamm* 2015. <https://doi.org/10.1155/2015/984825>
- Ma, Zhuo, King, K., Alqahtani, M., Worden, M., Muthuraman, P., Cioffi, C.L., Bakshi, C.S., Malik, M., 2019. Stringent response governs the oxidative stress resistance and virulence of Francisella tularensis. *PLOS ONE* 14, e0224094. <https://doi.org/10.1371/journal.pone.0224094>
- Ma, Z., Russo, V.C., Rabadi, S.M., Jen, Y., Catlett, S.V., Bakshi, C.S., Malik, M., 2016. Elucidation of a Mechanism of Oxidative Stress Regulation in Francisella tularensis Live Vaccine Strain. *Mol Microbiol* 101, 856–878. <https://doi.org/10.1111/mmi.13426>
- Ma, Zhipeng, Zhu, P., Shi, H., Guo, L., Zhang, Q., Chen, Y., Chen, S., Zhang, Z., Peng, J., Chen, J., 2019. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* 568, 259–263. <https://doi.org/10.1038/s41586-019-1057-y>
- MacIntyre, C.R., 2020. Reevaluating the Risk of Smallpox Reemergence. *Mil Med* 185, e952–e957. <https://doi.org/10.1093/milmed/usaa084>

- MacIntyre, C.R., Seccull, A., Lane, J.M., Plant, A., 2006. Development of a Risk-Priority Score for Category A Bioterrorism Agents as an Aid for Public Health Policy. *Military Medicine* 171, 589–594. <https://doi.org/10.7205/MILMED.171.7.589>
- Mackinnon, M.J., Ndila, C., Uyoga, S., Macharia, A., Snow, R.W., Band, G., Rautanen, A., Rockett, K.A., Kwiatkowski, D.P., Williams, T.N., 2016. Environmental Correlation Analysis for Genes Associated with Protection against Malaria. *Mol. Biol. Evol.* 33, 1188–1204. <https://doi.org/10.1093/molbev/msw004>
- MacManes, M.D., 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5. <https://doi.org/10.3389/fgene.2014.00013>
- Maeß, M.B., Wittig, B., Cignarella, A., Lorkowski, S., 2014a. Reduced PMA enhances the responsiveness of transfected THP-1 macrophages to polarizing stimuli. *Journal of Immunological Methods* 402, 76–81. <https://doi.org/10.1016/j.jim.2013.11.006>
- Maeß, M.B., Wittig, B., Lorkowski, S., 2014b. Highly Efficient Transfection of Human THP-1 Macrophages by Nucleofection. *J Vis Exp.* <https://doi.org/10.3791/51960>
- Mahmoud, R.Y., Li, W., Eldomany, R.A., Emara, M., Yu, J., 2016a. The Shigella ProU system is required for osmotic tolerance and virulence. *Virulence* 8, 362–374. <https://doi.org/10.1080/21505594.2016.1227906>
- Mahmoud, R.Y., Stones, D.H., Li, W., Emara, M., El-domany, R.A., Wang, D., Wang, Y., Krachler, A.M., Yu, J., 2016b. The Multivalent Adhesion Molecule SSO1327 plays a key role in Shigella sonnei pathogenesis. *Molecular Microbiology* 99, 658–673. <https://doi.org/10.1111/mmi.13255>
- Major, L.D., Partridge, T.S., Gardner, J., Kent, S.J., Rose, R. de, Suhrbier, A., Schroder, W.A., 2013. Induction of SerpinB2 and Th1/Th2 Modulation by SerpinB2 during Lentiviral Infections In Vivo. *PLOS ONE* 8, e57343. <https://doi.org/10.1371/journal.pone.0057343>
- Man, S.M., Karki, R., Malireddi, R.K.S., Neale, G., Vogel, P., Yamamoto, M., Lamkanfi, M., Kanneganti, T.-D., 2015. The transcription factor IRF1 and guanylate-binding proteins target activation of the AIM2 inflammasome by Francisella infection. *Nature Immunology* 16, 467–475. <https://doi.org/10.1038/ni.3118>
- Mansilla Pareja, M.E., Bongiovanni, A., Lafont, F., Colombo, M.I., 2017. Alterations of the Coxiella burnetii Replicative Vacuole Membrane Integrity and Interplay with the Autophagy Pathway. *Front. Cell. Infect. Microbiol.* 7. <https://doi.org/10.3389/fcimb.2017.00112>
- Marchetti, G., Pinotti, M., Lunghi, B., Casari, C., Bernardi, F., 2012. Functional genetics. *Thrombosis Research, XI ETRO ADVANCED TEACHING COURSE “Thrombosis: a multidisciplinary approach”* September 18-23, 2011, Campobasso, Italy 129, 336–340. <https://doi.org/10.1016/j.thromres.2011.10.028>
- Mares, C.A., Ojeda, S.S., Morris, E.G., Li, Q., Teale, J.M., 2008. Initial Delay in the Immune Response to Francisella tularensis Is Followed by Hypercytokinemia Characteristic of Severe Sepsis and Correlating with Upregulation and Release of Damage-Associated Molecular Patterns. *Infection and Immunity* 76, 3001–3010. <https://doi.org/10.1128/IAI.00215-08>
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinez, J., Malireddi, R.K.S., Lu, Q., Cunha, L.D., Pelletier, S., Gingras, S., Orchard, R., Guan, J.-L., Tan, H., Peng, J., Kanneganti, T.-D., Virgin, H.W., Green, D.R., 2015. Molecular characterization of LC3-associated phagocytosis reveals distinct roles for Rubicon, NOX2 and autophagy proteins. *Nat Cell Biol* 17, 893–906. <https://doi.org/10.1038/ncb3192>

- Maruzs, T., Lőrincz, P., Szatmári, Z., Széplaki, S., Sándor, Z., Lakatos, Z., Puska, G., Juhász, G., Sass, M., 2015. Retromer Ensures the Degradation of Autophagic Cargo by Maintaining Lysosome Function in *Drosophila*. *Traffic* 16, 1088–1107. <https://doi.org/10.1111/tra.12309>
- Masters, T.A., Pontes, B., Viasnoff, V., Li, Y., Gauthier, N.C., 2013. Plasma membrane tension orchestrates membrane trafficking, cytoskeletal remodeling, and biochemical signaling during phagocytosis. *PNAS* 110, 11875–11880. <https://doi.org/10.1073/pnas.1301766110>
- Mastroeni, P., Vazquez-Torres, A., Fang, F.C., Xu, Y., Khan, S., Hormaeche, C.E., Dougan, G., 2000. Antimicrobial Actions of the NADPH Phagocyte Oxidase and Inducible Nitric Oxide Synthase in Experimental Salmonellosis. II. Effects on Microbial Proliferation and Host Survival in Vivo. *Journal of Experimental Medicine* 192, 237–248. <https://doi.org/10.1084/jem.192.2.237>
- Maurin, M., Gyuranecz, M., 2016. Tularaemia: clinical aspects in Europe. *The Lancet Infectious Diseases* 16, 113–124. [https://doi.org/10.1016/S1473-3099\(15\)00355-2](https://doi.org/10.1016/S1473-3099(15)00355-2)
- May, B., 2011. TOMM40 complex translocates proteins from the cytosol to the mitochondrial intermembrane space (No. Reactome release 72). Reactome.
- Mayor-Ruiz, C., Dominguez, O., Fernandez-Capetillo, O., 2017. TrapSeq: An RNA Sequencing-Based Pipeline for the Identification of Gene-Trap Insertions in Mammalian Cells. *Journal of Molecular Biology* 429, 2780–2789. <https://doi.org/10.1016/j.jmb.2017.07.020>
- McCaffrey, R.L., Allen, L.-A.H., 2006. Pivotal Advance: *Francisella tularensis* LVS evades killing by human neutrophils via inhibition of the respiratory burst and phagosome escape. *J Leukoc Biol* 80, 1224–1230. <https://doi.org/10.1189/jlb.0406287>
- McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413, 852–856. <https://doi.org/10.1038/35101614>
- McCurley, N., Mellman, I., 2010. Monocyte-Derived Dendritic Cells Exhibit Increased Levels of Lysosomal Proteolysis as Compared to Other Human Dendritic Cell Populations. *PLOS ONE* 5, e11949. <https://doi.org/10.1371/journal.pone.0011949>
- McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32, W20–W25. <https://doi.org/10.1093/nar/gkh435>
- McInerney, P., Adams, P., Hadi, M.Z., 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International* 2014, e287430. <https://doi.org/10.1155/2014/287430>
- McManus, M.T., Petersen, C.P., Haines, B.B., Chen, J., Sharp, P.A., 2002. Gene silencing using micro-RNA designed hairpins. *RNA* 8, 842–850.
- McVicker, G., Tang, C.M., 2017. Deletion of toxin–antitoxin systems in the evolution of *Shigella sonnei* as a host-adapted pathogen. *Nature Microbiology* 2, 16204. <https://doi.org/10.1038/nmicrobiol.2016.204>
- McWhorter, F.Y., Wang, T., Nguyen, P., Chung, T., Liu, W.F., 2013. Modulation of macrophage phenotype by cell shape. *PNAS* 110, 17253–17258. <https://doi.org/10.1073/pnas.1308887110>

- Mehta, M., Singh, A., 2019. Mycobacterium tuberculosis WhiB3 maintains redox homeostasis and survival in response to reactive oxygen and nitrogen species. *Free Radic Biol Med* 131, 50–58. <https://doi.org/10.1016/j.freeradbiomed.2018.11.032>
- Meinhardt, G., Hass, R., 1995. Differential expression of c-myc, max and mx1 in human myeloid leukemia cells during retrodifferentiation and cell death. *Leukemia Research* 19, 699–705. [https://doi.org/10.1016/0145-2126\(95\)00040-U](https://doi.org/10.1016/0145-2126(95)00040-U)
- Melillo, A.A., Bakshi, C.S., Melendez, J.A., 2010. Francisella tularensis Antioxidants Harness Reactive Oxygen Species to Restrict Macrophage Signaling and Cytokine Production. *J. Biol. Chem.* 285, 27553–27560. <https://doi.org/10.1074/jbc.M110.144394>
- Meunier, E., Broz, P., 2015. Quantification of Cytosolic vs. Vacuolar Salmonella in Primary Macrophages by Differential Permeabilization. *J Vis Exp.* <https://doi.org/10.3791/52960>
- Meunier, E., Dick, M.S., Dreier, R.F., Schürmann, N., Broz, D.K., Warming, S., Roose-Girma, M., Bumann, D., Kayagaki, N., Takeda, K., Yamamoto, M., Broz, P., 2014. Caspase-11 activation requires lysis of pathogen-containing vacuoles by IFN-induced GTPases. *Nature* 509, 366–370. <https://doi.org/10.1038/nature13157>
- Meunier, E., Wallet, P., Dreier, R.F., Costanzo, S., Anton, L., Rühl, S., Dussurgey, S., Dick, M.S., Kistner, A., Rigard, M., Degrandi, D., Pfeffer, K., Yamamoto, M., Henry, T., Broz, P., 2015. Guanylate-binding proteins promote activation of the AIM2 inflammasome during infection with Francisella novicida. *Nature Immunology* 16, 476–484. <https://doi.org/10.1038/ni.3119>
- Miao, E.A., Alpujch-Aranda, C.M., Dors, M., Clark, A.E., Bader, M.W., Miller, S.I., Aderem, A., 2006. Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1 $\beta$  via Ipaf. *Nature Immunology* 7, 569–575. <https://doi.org/10.1038/ni1344>
- Miao, E.A., Leaf, I.A., Treuting, P.M., Mao, D.P., Dors, M., Sarkar, A., Warren, S.E., Wewers, M.D., Aderem, A., 2010. Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. *Nature Immunology* 11, 1136–1142. <https://doi.org/10.1038/ni.1960>
- Michard, C., Yum, L.K., Agaisse, H., 2019. WIPF2 promotes Shigella flexneri actin-based motility and cell-to-cell spread. *Cellular Microbiology* 21, e13098. <https://doi.org/10.1111/cmi.13098>
- Micheva-Viteva, S.N., Shou, Y., Ganguly, K., Wu, T.H., Hong-Geller, E., 2017. PKC- $\eta$ -MARCKS Signaling Promotes Intracellular Survival of Unopsonized Burkholderia thailandensis. *Front Cell Infect Microbiol* 7. <https://doi.org/10.3389/fcimb.2017.00231>
- Michl, J., Pieczonka, M.M., Unkeless, J.C., Silverstein, S.C., 1979. Effects of immobilized immune complexes on Fc- and complement-receptor function in resident and thioglycollate-elicited mouse peritoneal macrophages. *Journal of Experimental Medicine* 150, 607–621. <https://doi.org/10.1084/jem.150.3.607>
- Mick, D.U., Dennerlein, S., Wiese, H., Reinhold, R., Pacheu-Grau, D., Lorenzi, I., Sasarman, F., Weraarpachai, W., Shoubbridge, E.A., Warscheid, B., Rehling, P., 2012. MITRAC Links Mitochondrial Protein Translocation to Respiratory-Chain Assembly and Translational Regulation. *Cell* 151, 1528–1541. <https://doi.org/10.1016/j.cell.2012.11.053>
- Mildner, A., Yona, S., Jung, S., 2013. Chapter Three - A Close Encounter of the Third Kind: Monocyte-Derived Cells, in: Murphy, K.M., Merad, M. (Eds.), *Advances in Immunology, Development and Function of Myeloid Subsets*. Academic Press, pp. 69–103. <https://doi.org/10.1016/B978-0-12-417028-5.00003-X>

- Milewski, M.I., Mickle, J.E., Forrest, J.K., Stafford, D.M., Moyer, B.D., Cheng, J., Guggino, W.B., Stanton, B.A., Cutting, G.R., 2001. A PDZ-binding motif is essential but not sufficient to localize the C terminus of CFTR to the apical membrane. *Journal of Cell Science* 114, 719–726.
- Misawa, T., Takahama, M., Kozaki, T., Lee, H., Zou, J., Saitoh, T., Akira, S., 2013. Microtubule-driven spatial arrangement of mitochondria promotes activation of the NLRP3 inflammasome. *Nature Immunology* 14, 454–460. <https://doi.org/10.1038/ni.2550>
- Misselwitz, B., Dilling, S., Vonaesch, P., Sacher, R., Snijder, B., Schlumberger, M., Rout, S., Stark, M., von Mering, C., Pelkmans, L., Hardt, W.-D., 2011. RNAi screen of *Salmonella* invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Molecular Systems Biology* 7, 474. <https://doi.org/10.1038/msb.2011.7>
- Mitchell, G., Cheng, M.I., Chen, C., Nguyen, B.N., Whiteley, A.T., Kianian, S., Cox, J.S., Green, D.R., McDonald, K.L., Portnoy, D.A., 2018. *Listeria monocytogenes* triggers noncanonical autophagy upon phagocytosis, but avoids subsequent growth-restricting xenophagy. *PNAS* 115, E210–E217. <https://doi.org/10.1073/pnas.1716055115>
- Mitchell, G., Isberg, R.R., 2017. Innate Immunity to Intracellular Pathogens: Balancing Microbial Elimination and Inflammation. *Cell Host Microbe* 22, 166–175. <https://doi.org/10.1016/j.chom.2017.07.005>
- Mitra, S., Dunphy, P.S., Das, S., Zhu, B., Luo, T., McBride, J.W., 2018. Ehrlichia chaffeensis TRP120 Effector Targets and Recruits Host Polycomb Group Proteins for Degradation To Promote Intracellular Infection. *Infection and Immunity* 86. <https://doi.org/10.1128/IAI.00845-17>
- Mitterstiller, A.-M., Haschka, D., Dichtl, S., Nairz, M., Demetz, E., Talasz, H., Soares, M.P., Einwallner, E., Esterbauer, H., Fang, F.C., Geley, S., Weiss, G., 2016. Heme oxygenase 1 controls early innate immune response of macrophages to *Salmonella* Typhimurium infection. *Cellular Microbiology* 18, 1374–1389. <https://doi.org/10.1111/cmi.12578>
- Mizushima, N., Yamamoto, A., Hatano, M., Kobayashi, Y., Kabeya, Y., Suzuki, K., Tokuhiya, T., Ohsumi, Y., Yoshimori, T., 2001. Dissection of Autophagosome Formation Using Apg5-Deficient Mouse Embryonic Stem Cells. *J Cell Biol* 152, 657–668. <https://doi.org/10.1083/jcb.152.4.657>
- Monribot-Villanueva, J., Zurita, M., Vázquez, M., 2017. Developmental transcriptional regulation by SUMOylation, an evolving field. *Genesis* 55. <https://doi.org/10.1002/dvg.23009>
- Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., Valen, E., 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 42, W401–W407. <https://doi.org/10.1093/nar/gku410>
- Morais, V.A., Haddad, D., Craessaerts, K., Bock, P.-J.D., Swerts, J., Vilain, S., Aerts, L., Overbergh, L., Grünwald, A., Seibler, P., Klein, C., Gevaert, K., Verstreken, P., Strooper, B.D., 2014. PINK1 Loss-of-Function Mutations Affect Mitochondrial Complex I Activity via Ndufa10 Ubiquinone Uncoupling. *Science* 344, 203–207. <https://doi.org/10.1126/science.1249161>
- Moratal, C., Raffort, J., Arrighi, N., Rekima, S., Schaub, S., Dechesne, C.A., Chinetti, G., Dani, C., 2018. IL-1 $\beta$ - and IL-4-polarized macrophages have opposite effects on adipogenesis of intramuscular fibro-adipogenic progenitors in humans. *Scientific Reports* 8, 17005. <https://doi.org/10.1038/s41598-018-35429-w>

- Morgens, D.W., Deans, R.M., Li, A., Bassik, M.C., 2016. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nature Biotechnology* 34, 634–636. <https://doi.org/10.1038/nbt.3567>
- Morici, L., Torres, A.G., Titball, R.W., 2019. Novel multi-component vaccine approaches for *Burkholderia pseudomallei*. *Clin Exp Immunol* 196, 178–188. <https://doi.org/10.1111/cei.13286>
- Morikawa, H., Kim, M., Mimuro, H., Punginelli, C., Koyama, T., Nagai, S., Miyawaki, A., Iwai, K., Sasakawa, C., 2010. The bacterial effector Cif interferes with SCF ubiquitin ligase function by inhibiting deneddylation of Cullin1. *Biochemical and Biophysical Research Communications* 401, 268–274. <https://doi.org/10.1016/j.bbrc.2010.09.048>
- Morris, A.P., Cunningham, S.A., Tousson, A., Benos, D.J., Frizzell, R.A., 1994. Polarization-dependent apical membrane CFTR targeting underlies cAMP-stimulated Cl<sup>-</sup> secretion in epithelial cells. *American Journal of Physiology-Cell Physiology* 266, C254–C268. <https://doi.org/10.1152/ajpcell.1994.266.1.C254>
- Morrone, S.R., Matyszewski, M., Yu, X., Delannoy, M., Egelman, E.H., Sohn, J., 2015. Assembly-driven activation of the AIM2 foreign-dsDNA sensor provides a polymerization template for downstream ASC. *Nature Communications* 6, 7827. <https://doi.org/10.1038/ncomms8827>
- Mukhopadhyay, D., Dasso, M., 2017. The SUMO Pathway in Mitosis. *Adv. Exp. Med. Biol.* 963, 171–184. [https://doi.org/10.1007/978-3-319-50044-7\\_10](https://doi.org/10.1007/978-3-319-50044-7_10)
- Muñoz-Planillo, R., Kuffa, P., Martínez-Colón, G., Smith, B.L., Rajendiran, T.M., Núñez, G., 2013. K<sup>+</sup> Efflux Is the Common Trigger of NLRP3 Inflammasome Activation by Bacterial Toxins and Particulate Matter. *Immunity* 38, 1142–1153. <https://doi.org/10.1016/j.immuni.2013.05.016>
- Murray, P.J., 2017. Macrophage Polarization. *Annu. Rev. Physiol.* 79, 541–566. <https://doi.org/10.1146/annurev-physiol-022516-034339>
- Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdts, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T., Locati, M., Mantovani, A., Martinez, F.O., Mege, J.-L., Mosser, D.M., Natoli, G., Saeij, J.P., Schultze, J.L., Shirey, K.A., Sica, A., Suttles, J., Udalova, I., van Ginderachter, J.A., Vogel, S.N., Wynn, T.A., 2014. Macrophage Activation and Polarization: Nomenclature and Experimental Guidelines. *Immunity* 41, 14–20. <https://doi.org/10.1016/j.immuni.2014.06.008>
- Nakamura, N., Lill, J.R., Phung, Q., Jiang, Z., Bakalarski, C., de Mazière, A., Klumperman, J., Schlatter, M., Delamarre, L., Mellman, I., 2014. Endosomes are specialized platforms for bacterial sensing and NOD2 signalling. *Nature* 509, 240–244. <https://doi.org/10.1038/nature13133>
- Nakatogawa, H., Ichimura, Y., Ohsumi, Y., 2007. Atg8, a Ubiquitin-like Protein Required for Autophagosome Formation, Mediates Membrane Tethering and Hemifusion. *Cell* 130, 165–178. <https://doi.org/10.1016/j.cell.2007.05.021>
- Nakaya, T., Sato, M., Hata, N., Asagiri, M., Suemori, H., Noguchi, S., Tanaka, N., Taniguchi, T., 2001. Gene Induction Pathways Mediated by Distinct IRFs during Viral Infection. *Biochemical and Biophysical Research Communications* 283, 1150–1156. <https://doi.org/10.1006/bbrc.2001.4913>
- Nalvarte, I., Damdimopoulos, A.E., Spyrou, G., 2004. Human mitochondrial thioredoxin reductase reduces cytochrome c and confers resistance to complex III inhibition. *Free Radical Biology and Medicine* 36, 1270–1278. <https://doi.org/10.1016/j.freeradbiomed.2004.02.072>

- Napier, B.A., Brubaker, S.W., Sweeney, T.E., Monette, P., Rothmeier, G.H., Gertsvoft, N.A., Puschnik, A., Carette, J.E., Khatri, P., Monack, D.M., 2016. Complement pathway amplifies caspase-11-dependent cell death and endotoxin-induced sepsis severity. *Journal of Experimental Medicine* 213, 2365–2382. <https://doi.org/10.1084/jem.20160027>
- Narendra, D.P., Jin, S.M., Tanaka, A., Suen, D.-F., Gautier, C.A., Shen, J., Cookson, M.R., Youle, R.J., 2010. PINK1 Is Selectively Stabilized on Impaired Mitochondria to Activate Parkin. *PLOS Biology* 8, e1000298. <https://doi.org/10.1371/journal.pbio.1000298>
- National Institute of Allergy and Infectious Diseases (NIAID), 2020. Phase IIa Randomized, Single-Blinded, Placebo-Controlled Clinical Trial of the Reprofiled Drug Auranofin for GI Protozoa (Clinical trial registration No. NCT02736968). [clinicaltrials.gov](https://clinicaltrials.gov).
- National Institute of Allergy and Infectious Diseases (NIAID), 2017. An Open Label, Multiple Dose Study to Evaluate the Pharmacokinetics of Auranofin Following Oral Dose Administration for 7 Days to Healthy Subjects (Clinical trial registration No. NCT02089048). [clinicaltrials.gov](https://clinicaltrials.gov).
- Naujoks, J., Tabeling, C., Dill, B.D., Hoffmann, C., Brown, A.S., Kunze, M., Kempa, S., Peter, A., Mollenkopf, H.-J., Dorhoi, A., Kershaw, O., Gruber, A.D., Sander, L.E., Witzernath, M., Herold, S., Nerlich, A., Hocke, A.C., Driel, I. van, Suttorp, N., Bedoui, S., Hilbi, H., Trost, M., Opitz, B., 2016. IFNs Modify the Proteome of Legionella-Containing Vacuoles and Restrict Infection Via IRG1-Derived Itaconic Acid. *PLOS Pathogens* 12, e1005408. <https://doi.org/10.1371/journal.ppat.1005408>
- Nauseef, W.M., 2019. The phagocyte NOX2 NADPH oxidase in microbial killing and cell signaling. *Current Opinion in Immunology, Allergy and hypersensitivity • Host pathogens* 60, 130–140. <https://doi.org/10.1016/j.coi.2019.05.006>
- Navratil, A.R., Brummett, A.M., Bryan, J.D., Woolard, M.D., 2014. Francisella tularensis LVS Induction of Prostaglandin Biosynthesis by Infected Macrophages Requires Specific Host Phospholipases and Lipid Phosphatases. *Infection and Immunity* 82, 3299–3311. <https://doi.org/10.1128/IAI.02060-14>
- Newton, H.J., Pearson, J.S., Badea, L., Kelly, M., Lucas, M., Holloway, G., Wagstaff, K.M., Dunstone, M.A., Sloan, J., Whisstock, J.C., Kaper, J.B., Robins-Browne, R.M., Jans, D.A., Frankel, G., Phillips, A.D., Coulson, B.S., Hartland, E.L., 2010. The Type III Effectors NleE and NleB from Enteropathogenic E. coli and OspZ from Shigella Block Nuclear Translocation of NF-κB p65. *PLOS Pathogens* 6, e1000898. <https://doi.org/10.1371/journal.ppat.1000898>
- Ngo, C.C., Man, S.M., 2017. Mechanisms and functions of guanylate-binding proteins and related interferon-inducible GTPases: Roles in intracellular lysis of pathogens. *Cellular Microbiology* 19, e12791. <https://doi.org/10.1111/cmi.12791>
- Nguyen, K.D., Qiu, Y., Cui, X., Goh, Y.P.S., Mwangi, J., David, T., Mukundan, L., Brombacher, F., Locksley, R.M., Chawla, A., 2011. Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature* 480, 104–108. <https://doi.org/10.1038/nature10653>
- NIH: National Institute of Allergy and Infectious Diseases, 2016. NIAID Emerging Infectious Diseases/Pathogens [WWW Document]. URL <https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens> (accessed 7.19.17).
- Nnadozie, C.F., Odume, O.N., 2019. Freshwater environments as reservoirs of antibiotic resistant bacteria and their role in the dissemination of antibiotic resistance genes.

- Environmental Pollution 254, 113067.  
<https://doi.org/10.1016/j.envpol.2019.113067>
- Noad, J., von der Malsburg, A., Pathe, C., Michel, M.A., Komander, D., Randow, F., 2017. LUBAC-synthesized linear ubiquitin chains restrict cytosol-invading bacteria by activating autophagy and NF- $\kappa$ B. *Nat Microbiol* 2, 17063.  
<https://doi.org/10.1038/nmicrobiol.2017.63>
- Nottke, A.C., Kim, H.-M., Colaiácovo, M.P., 2017. Wrestling with Chromosomes: The Roles of SUMO During Meiosis. *Adv. Exp. Med. Biol.* 963, 185–196.  
[https://doi.org/10.1007/978-3-319-50044-7\\_11](https://doi.org/10.1007/978-3-319-50044-7_11)
- Nozawa, T., Minowa-Nozawa, A., Aikawa, C., Nakagawa, I., 2017. The STX6-VTI1B-VAMP3 complex facilitates xenophagy by regulating the fusion between recycling endosomes and autophagosomes. *Autophagy* 13, 57–69.  
<https://doi.org/10.1080/15548627.2016.1241924>
- Offenbacher, S., Divaris, K., Barros, S.P., Moss, K.L., Marchesan, J.T., Morelli, T., Zhang, S., Kim, S., Sun, L., Beck, J.D., Laudes, M., Munz, M., Schaefer, A.S., North, K.E., 2016. Genome-wide association study of biologically informed periodontal complex traits offers novel insights into the genetic basis of periodontal disease. *Hum. Mol. Genet.* 25, 2113–2129. <https://doi.org/10.1093/hmg/ddw069>
- Ohya, K., Handa, Y., Ogawa, M., Suzuki, M., Sasakawa, C., 2005. IpgB1 Is a Novel Shigella Effector Protein Involved in Bacterial Invasion of Host Cells ITS ACTIVITY TO PROMOTE MEMBRANE RUFFLING VIA RAC1 AND CDC42 ACTIVATION. *J. Biol. Chem.* 280, 24022–24034. <https://doi.org/10.1074/jbc.M502509200>
- Okada, Y., Kimura, T., Kameoka, M., Kishi, M., Azuma, I., Ikuta, K., 1995. Viral activation from latency during retrodifferentiation of U937 cells exposed to phorbol ester followed by infection with human immunodeficiency virus type 1. *Immunopharmacology* 30, 27–39. [https://doi.org/10.1016/0162-3109\(95\)98646-K](https://doi.org/10.1016/0162-3109(95)98646-K)
- O’Neil, J., 2016. Tackling Drug-Resistant Infections Globally: final report and recommendations. Review on Antimicrobial Resistance.
- Ørstavik, S., Natarajan, V., Taskén, K., Jahnsen, T., Sandberg, M., 1997. Characterization of the Human Gene Encoding the Type  $\alpha$  and Type  $\beta$  cGMP-Dependent Protein Kinase (PRKG1). *Genomics* 42, 311–318. <https://doi.org/10.1006/geno.1997.4743>
- Ott, C.J., Blackledge, N.P., Kerschner, J.L., Leir, S.-H., Crawford, G.E., Cotton, C.U., Harris, A., 2009. Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *PNAS* 106, 19934–19939. <https://doi.org/10.1073/pnas.0900946106>
- Øvstebø, R., Olstad, O.K., Brusletto, B., Møller, A.S., Aase, A., Haug, K.B.F., Brandtzaeg, P., Kierulf, P., 2008. Identification of Genes Particularly Sensitive to Lipopolysaccharide (LPS) in Human Monocytes Induced by Wild-Type versus LPS-Deficient *Neisseria meningitidis* Strains. *Infection and Immunity* 76, 2685–2695.  
<https://doi.org/10.1128/IAI.01625-07>
- Ozawa, T., Natori, Y., Sato, M., Umezawa, Y., 2007. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* 4, 413–419.  
<https://doi.org/10.1038/nmeth1030>
- PacBio, 2020. RNA Sequencing [WWW Document]. PacBio. URL <https://www.pacb.com/applications/rna-sequencing/> (accessed 4.29.20).
- Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., Conklin, D.S., 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* 16, 948–958. <https://doi.org/10.1101/gad.981002>

- Paetzold, S., Lourido, S., Raupach, B., Zychlinsky, A., 2007. Shigella flexneri Phagosomal Escape Is Independent of Invasion. *Infection and Immunity* 75, 4826–4830. <https://doi.org/10.1128/IAI.00454-07>
- Pagie, S., Gérard, N., Charreau, B., 2018. Notch signaling triggered via the ligand DLL4 impedes M2 macrophage differentiation and promotes their apoptosis. *Cell Communication and Signaling* 16, 4. <https://doi.org/10.1186/s12964-017-0214-x>
- Pai, M., Behr, M.A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C.C., Ginsberg, A., Swaminathan, S., Spigelman, M., Getahun, H., Menzies, D., Raviglione, M., 2016. Tuberculosis. *Nature Reviews Disease Primers* 2, 1–23. <https://doi.org/10.1038/nrdp.2016.76>
- Panagi, I., Jennings, E., Zeng, J., Günster, R.A., Stones, C.D., Mak, H., Jin, E., Stapels, D.A.C., Subari, Nur.Z., Pham, T.H.M., Brewer, S.M., Ong, S.Y.Q., Monack, D.M., Helaine, S., Thurston, T.L.M., 2020. Salmonella Effector SteE Converts the Mammalian Serine/Threonine Kinase GSK3 into a Tyrosine Kinase to Direct Macrophage Polarization. *Cell Host Microbe* 27, 41-53.e6. <https://doi.org/10.1016/j.chom.2019.11.002>
- Pandey, A.K., Yang, Y., Jiang, Z., Fortune, S.M., Coulombe, F., Behr, M.A., Fitzgerald, K.A., Sasseti, C.M., Kelliher, M.A., 2009. NOD2, RIP2 and IRF5 Play a Critical Role in the Type I Interferon Response to Mycobacterium tuberculosis. *PLOS Pathogens* 5, e1000500. <https://doi.org/10.1371/journal.ppat.1000500>
- Pankow, S., Bamberger, C., Yates, J.R., 2019. A posttranslational modification code for CFTR maturation is altered in cystic fibrosis. *Sci. Signal.* 12. <https://doi.org/10.1126/scisignal.aan7984>
- Panning, B., 2008. X-chromosome inactivation: the molecular basis of silencing. *Journal of Biology* 7, 30. <https://doi.org/10.1186/jbiol95>
- Papadopoulos, G., Weinberg, E.O., Massari, P., Gibson, F.C., Wetzler, L.M., Morgan, E.F., Genco, C.A., 2013. Macrophage-Specific TLR2 Signaling Mediates Pathogen-Induced TNF-Dependent Inflammatory Oral Bone Loss. *The Journal of Immunology* 190, 1148–1157. <https://doi.org/10.4049/jimmunol.1202511>
- Pappas, G., Panagopoulou, P., Akritidis, N., 2009. Reclassifying bioterrorism risk: Are we preparing for the proper pathogens? *Journal of Infection and Public Health* 2, 55–61. <https://doi.org/10.1016/j.jiph.2009.03.002>
- Park, H.W., Nam, J.H., Kim, J.Y., Namkung, W., Yoon, J.S., Lee, J.-S., Kim, K.S., Venglovecz, V., Gray, M.A., Kim, K.H., Lee, M.G., 2010. Dynamic Regulation of CFTR Bicarbonate Permeability by [Cl<sup>-</sup>]<sub>i</sub> and Its Role in Pancreatic Bicarbonate Secretion. *Gastroenterology* 139, 620–631. <https://doi.org/10.1053/j.gastro.2010.04.004>
- Patel, J.C., Galán, J.E., 2006. Differential activation and function of Rho GTPases during Salmonella–host cell interactions. *J Cell Biol* 175, 453–463. <https://doi.org/10.1083/jcb.200605144>
- Patel, S.J., Wellington, M., Shah, R.M., Ferreira, M.J., 2020. Antibiotic Stewardship in Food-producing Animals: Challenges, Progress, and Opportunities. *Clinical Therapeutics* 42, 1649–1658. <https://doi.org/10.1016/j.clinthera.2020.07.004>
- Paul, B., Kim, H.S., Kerr, M.C., Huston, W.M., Teasdale, R.D., Collins, B.M., 2017. Structural basis for the hijacking of endosomal sorting nexin proteins by Chlamydia trachomatis. *eLife* 6, e22311. <https://doi.org/10.7554/eLife.22311>
- Paulsen, K., Tauber, S., Timm, J., Goelz, N., Dumrese, C., Stolzing, A., Hass, R., Ullrich, O., 2011. The cannabinoid receptors agonist WIN55212-2 inhibits macrophageal differentiation and alters expression and phosphorylation of cell cycle control

- proteins. *Cell Communication and Signaling* 9, 33. <https://doi.org/10.1186/1478-811X-9-33>
- Payá-Milans, M., Olmstead, J.W., Nunez, G., Rinehart, T.A., Staton, M., 2018. Comprehensive evaluation of RNA-seq analysis pipelines in diploid and polyploid species. *Gigascience* 7. <https://doi.org/10.1093/gigascience/giy132>
- Payne, D.J., Gwynn, M.N., Holmes, D.J., Pompliano, D.L., 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* 6, 29–40. <https://doi.org/10.1038/nrd2201>
- Payne, S.M., Finkelstein, R.A., 1977. Detection and differentiation of iron-responsive avirulent mutants on Congo red agar. *Infection and Immunity* 18, 94–98.
- Pechous, R.D., Sivaraman, V., Stasulli, N.M., Goldman, W.E., 2016. Pneumonic Plague: The Darker Side of *Yersinia pestis*. *Trends in Microbiology* 24, 190–197. <https://doi.org/10.1016/j.tim.2015.11.008>
- Peiser, L., Gough, P.J., Kodama, T., Gordon, S., 2000. Macrophage Class A Scavenger Receptor-Mediated Phagocytosis of *Escherichia coli*: Role of Cell Heterogeneity, Microbial Strain, and Culture Conditions In Vitro. *Infection and Immunity* 68, 1953–1963. <https://doi.org/10.1128/IAI.68.4.1953-1963.2000>
- Peng, S., Song, C., Li, H., Cao, X., Ma, Y., Wang, X., Huang, Y., Lan, X., Lei, C., Chaogetu, B., Chen, H., 2019. Circular RNA SNX29 Sponges miR-744 to Regulate Proliferation and Differentiation of Myoblasts by Activating the Wnt5a/Ca<sup>2+</sup> Signaling Pathway. *Molecular Therapy - Nucleic Acids* 16, 481–493. <https://doi.org/10.1016/j.omtn.2019.03.009>
- Perregaux, D., Gabel, C.A., 1994. Interleukin-1 beta maturation and release in response to ATP and nigericin. Evidence that potassium depletion mediated by these agents is a necessary and common feature of their activity. *J. Biol. Chem.* 269, 15195–15203.
- Perrin, A.J., Jiang, X., Birmingham, C.L., So, N.S.Y., Brumell, J.H., 2004. Recognition of Bacteria in the Cytosol of Mammalian Cells by the Ubiquitin System. *Current Biology* 14, 806–811. <https://doi.org/10.1016/j.cub.2004.04.033>
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11, 1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33, 290–295. <https://doi.org/10.1038/nbt.3122>
- Pétrilli, V., Papin, S., Dostert, C., Mayor, A., Martinon, F., Tschopp, J., 2007. Activation of the NALP3 inflammasome is triggered by low intracellular potassium concentration. *Cell Death & Differentiation* 14, 1583–1589. <https://doi.org/10.1038/sj.cdd.4402195>
- Petzke, M.M., Iyer, R., Love, A.C., Spieler, Z., Brooks, A., Schwartz, I., 2016. *Borrelia burgdorferi* induces a type I interferon response during early stages of disseminated infection in mice. *BMC Microbiology* 16, 29. <https://doi.org/10.1186/s12866-016-0644-4>
- Pham, T.H.M., Brewer, S.M., Thurston, T., Massis, L.M., Honeycutt, J., Lugo, K., Jacobson, A.R., Vilches-Moure, J.G., Hamblin, M., Helaine, S., Monack, D.M., 2020. Salmonella-Driven Polarization of Granuloma Macrophages Antagonizes TNF-Mediated Pathogen Restriction during Persistent Infection. *Cell Host & Microbe* 27, 54-67.e5. <https://doi.org/10.1016/j.chom.2019.11.011>
- Pichler, A., Fatouros, C., Lee, H., Eisenhardt, N., 2017. SUMO conjugation – a mechanistic view. *Biomolecular Concepts* 8, 13–36. <https://doi.org/10.1515/bmc-2016-0030>

- Pier, G.B., Grout, M., Zaidi, T., Meluleni, G., Mueschenborn, S.S., Banting, G., Ratcliff, R., Evans, M.J., Colledge, W.H., 1998. Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature* 393, 79–82. <https://doi.org/10.1038/30006>
- Pignataro, D., Francia, S., Zanetta, F., Brenna, G., Brandini, S., Olivieri, A., Torroni, A., Biamonti, G., Montecucco, A., 2017. A missense MT-ND5 mutation in differentiated Parkinson Disease cytoplasmic hybrid induces ROS-dependent DNA Damage Response amplified by DROSHA. *Scientific Reports* 7, 9528. <https://doi.org/10.1038/s41598-017-09910-x>
- Pilla, D.M., Hagar, J.A., Haldar, A.K., Mason, A.K., Degrandi, D., Pfeffer, K., Ernst, R.K., Yamamoto, M., Miao, E.A., Coers, J., 2014. Guanylate binding proteins promote caspase-11-dependent pyroptosis in response to cytoplasmic LPS. *PNAS* 111, 6046–6051. <https://doi.org/10.1073/pnas.1321700111>
- Pinczowski, D., Ekbo, A., Baron, J., Yuen, J., Adami, H.-O., 1994. Risk factors for colorectal cancer in patients with ulcerative colitis: A case-control study. *Gastroenterology* 107, 117–120. [https://doi.org/10.1016/0016-5085\(94\)90068-X](https://doi.org/10.1016/0016-5085(94)90068-X)
- Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M., Vitale, L., 2019. On the length, weight and GC content of the human genome. *BMC Res Notes* 12. <https://doi.org/10.1186/s13104-019-4137-z>
- Pizarro-Cerdá, J., Cossart, P., 2006. Bacterial Adhesion and Entry into Host Cells. *Cell* 124, 715–727. <https://doi.org/10.1016/j.cell.2006.02.012>
- Platnich, J.M., Chung, H., Lau, A., Sandall, C.F., Bondzi-Simpson, A., Chen, H.-M., Komada, T., Trotman-Grant, A.C., Brandelli, J.R., Chun, J., Beck, P.L., Philpott, D.J., Girardin, S.E., Ho, M., Johnson, R.P., MacDonald, J.A., Armstrong, G.D., Muruve, D.A., 2018. Shiga Toxin/Lipopolysaccharide Activates Caspase-4 and Gasdermin D to Trigger Mitochondrial Reactive Oxygen Species Upstream of the NLRP3 Inflammasome. *Cell Reports* 25, 1525-1536.e7. <https://doi.org/10.1016/j.celrep.2018.09.071>
- Poon, I.K.H., Lucas, C.D., Rossi, A.G., Ravichandran, K.S., 2014. Apoptotic cell clearance: basic biology and therapeutic potential. *Nature Reviews Immunology* 14, 166–180. <https://doi.org/10.1038/nri3607>
- Poprac, P., Jomova, K., Simunkova, M., Kollar, V., Rhodes, C.J., Valko, M., 2017. Targeting Free Radicals in Oxidative Stress-Related Human Diseases. *Trends in Pharmacological Sciences* 38, 592–607. <https://doi.org/10.1016/j.tips.2017.04.005>
- Potapov, V., Ong, J.L., 2017. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS One* 12, e0169774. <https://doi.org/10.1371/journal.pone.0169774>
- Poyet, J.-L., Srinivasula, S.M., Tnani, M., Razmara, M., Fernandes-Alnemri, T., Alnemri, E.S., 2001. Identification of Ipaf, a Human Caspase-1-activating Protein Related to Apaf-1. *J. Biol. Chem.* 276, 28309–28313. <https://doi.org/10.1074/jbc.C100250200>
- Preston, G.M., Brodsky, J.L., 2017. The evolving role of ubiquitin modification in endoplasmic reticulum-associated degradation. *Biochem J* 474, 445. <https://doi.org/10.1042/BCJ20160582>
- Princz, A., Tavernarakis, N., 2017. The role of SUMOylation in ageing and senescent decline. *Mech. Ageing Dev.* 162, 85–90. <https://doi.org/10.1016/j.mad.2017.01.002>
- Puig, K.L., Brose, S.A., Zhou, X., Sens, M.A., Combs, G.F., Jensen, M.D., Golovko, M.Y., Combs, C.K., 2017. Amyloid precursor protein modulates macrophage phenotype and diet-dependent weight gain. *Sci Rep* 7, 43725. <https://doi.org/10.1038/srep43725>

- Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., Mekalanos, J.J., 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *PNAS* 104, 15508–15513. <https://doi.org/10.1073/pnas.0706532104>
- Pulavendran, S., Prasanthi, M., Ramachandran, A., Grant, R., Snider, T.A., Chow, V.T.K., Malayer, J.R., Teluguakula, N., 2020. Production of Neutrophil Extracellular Traps Contributes to the Pathogenesis of Francisella tularemia. *Front. Immunol.* 11. <https://doi.org/10.3389/fimmu.2020.00679>
- Qiu, J., Yu, K., Fei, X., Liu, Y., Nakayasu, E.S., Piehowski, P.D., Shaw, J.B., Puvar, K., Das, C., Liu, X., Luo, Z.-Q., 2017. A unique deubiquitinase that deconjugates phosphoribosyl-linked protein ubiquitination. *Cell Res* 27, 865–881. <https://doi.org/10.1038/cr.2017.66>
- Québatte, M., Dehio, C., 2017. Systems-level interference strategies to decipher host factors involved in bacterial pathogen interaction: from RNAi to CRISPRi. *Current Opinion in Microbiology, Antimicrobials \* Bacterial Systems Biology* 39, 34–41. <https://doi.org/10.1016/j.mib.2017.08.002>
- Quero, L., Hanser, E., Manigold, T., Tiaden, A.N., Kyburz, D., 2017. TLR2 stimulation impairs anti-inflammatory activity of M2-like macrophages, generating a chimeric M1/M2 phenotype. *Arthritis Research & Therapy* 19, 245. <https://doi.org/10.1186/s13075-017-1447-1>
- Qureshi, N., Perera, P.-Y., Shen, J., Zhang, G., Lenschat, A., Splitter, G., Morrison, D.C., Vogel, S.N., 2003. The Proteasome as a Lipopolysaccharide-Binding Protein in Macrophages: Differential Effects of Proteasome Inhibition on Lipopolysaccharide-Induced Signaling Events. *The Journal of Immunology* 171, 1515–1525. <https://doi.org/10.4049/jimmunol.171.3.1515>
- Radin, J.N., González-Rivera, C., Frick-Cheng, A.E., Sheng, J., Gaddy, J.A., Rubin, D.H., Algood, H.M.S., McClain, M.S., Cover, T.L., 2014. Role of Connexin 43 in *Helicobacter pylori* VacA-Induced Cell Death. *Infection and Immunity* 82, 423–432. <https://doi.org/10.1128/IAI.00827-13>
- Radtke, A.L., Wilson, J.W., Sarker, S., Nickerson, C.A., 2010. Analysis of Interactions of Salmonella Type Three Secretion Mutants with 3-D Intestinal Epithelial Cells. *PLOS ONE* 5, e15750. <https://doi.org/10.1371/journal.pone.0015750>
- Raffatellu, M., Wilson, R.P., Chessa, D., Andrews-Polymenis, H., Tran, Q.T., Lawhon, S., Khare, S., Adams, L.G., Bäumler, A.J., 2005. SipA, SopA, SopB, SopD, and SopE2 Contribute to Salmonella enterica Serotype Typhimurium Invasion of Epithelial Cells. *Infection and Immunity* 73, 146–154. <https://doi.org/10.1128/IAI.73.1.146-154.2005>
- Ramadass, M., Johnson, J.L., Marki, A., Zhang, J., Wolf, D., Kiosses, W.B., Pestonjamas, K., Ley, K., Catz, S.D., 2019. The trafficking protein JFC1 regulates Rac1-GTP localization at the uropod controlling neutrophil chemotaxis and in vivo migration. *Journal of Leukocyte Biology* 105, 1209–1224. <https://doi.org/10.1002/JLB.1VMA0818-320R>
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* 8, 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Ranganathan, S., Doucet, M., Grassel, C.L., Delaine-Elias, B., Zachos, N.C., Barry, E.M., 2019. Evaluating Shigella flexneri Pathogenesis in the Human Enteroid Model. *Infection and Immunity* 87. <https://doi.org/10.1128/IAI.00740-18>
- Raplee, I.D., Evsikov, A.V., Marín de Evsikova, C., 2019. Aligning the Aligners: Comparison of RNA Sequencing Data Alignment and Gene Expression Quantification Tools for

- Clinical Breast Cancer Research. *Journal of Personalized Medicine* 9, 18.  
<https://doi.org/10.3390/jpm9020018>
- Rath, P., Allen, J.A., Schneider, D.S., 2018. Predicting position along a looping immune response trajectory. *PLOS ONE* 13, e0200147.  
<https://doi.org/10.1371/journal.pone.0200147>
- Rathinam, V.A.K., Jiang, Z., Waggoner, S.N., Sharma, S., Cole, L.E., Waggoner, L., Vanaja, S.K., Monks, B.G., Ganesan, S., Latz, E., Hornung, V., Vogel, S.N., Szomolanyi-Tsuda, E., Fitzgerald, K.A., 2010. The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses. *Nat Immunol* 11, 395–402.  
<https://doi.org/10.1038/ni.1864>
- Rauch, I., Deets, K.A., Ji, D.X., Moltke, J. von, Tenthorey, J.L., Lee, A.Y., Philip, N.H., Ayres, J.S., Brodsky, I.E., Gronert, K., Vance, R.E., 2017. NAIP-NLRC4 Inflammasomes Coordinate Intestinal Epithelial Cell Expulsion with Eicosanoid and IL-18 Release via Activation of Caspase-1 and -8. *Immunity* 46, 649–659.  
<https://doi.org/10.1016/j.immuni.2017.03.016>
- Rayamajhi, M., Zak, D.E., Chavarria-Smith, J., Vance, R.E., Miao, E.A., 2013. Cutting Edge: Mouse NAIP1 Detects the Type III Secretion System Needle Protein. *The Journal of Immunology* 191, 3986–3989. <https://doi.org/10.4049/jimmunol.1301549>
- Rechem, C.V., Black, J.C., Abbas, T., Allen, A., Rinehart, C.A., Yuan, G.-C., Dutta, A., Whetstine, J.R., 2011. The SKP1-Cul1-F-box and Leucine-rich Repeat Protein 4 (SCF-FbxL4) Ubiquitin Ligase Regulates Lysine Demethylase 4A (KDM4A)/Jumonji Domain-containing 2A (JMJD2A) Protein. *J. Biol. Chem.* 286, 30462–30470.  
<https://doi.org/10.1074/jbc.M111.273508>
- Reiman, M., Laan, M., Rull, K., Söber, S., 2017. Effects of RNA integrity on transcript quantification by total RNA sequencing of clinically collected human placental samples. *The FASEB Journal* 31, 3298–3308.  
<https://doi.org/10.1096/fj.201601031RR>
- Ribet, D., Hamon, M., Gouin, E., Nahori, M.-A., Impens, F., Neyret-Kahn, H., Gevaert, K., Vandekerckhove, J., Dejean, A., Cossart, P., 2012. *Listeria monocytogenes* impairs SUMOylation for efficient infection. *Nature* 464, 1192–1195.  
<https://doi.org/10.1038/nature08963>
- Richardson, G.M., Lannigan, J., Macara, I.G., 2015. Does FACS perturb gene expression? *Cytometry Part A* 87, 166–175. <https://doi.org/10.1002/cyto.a.22608>
- Richter, L., Landsverk, O.J.B., Atlasy, N., Bujko, A., Yaqub, S., Horneland, R., Øyen, O., Aandahl, E.M., Lundin, K.E.A., Stunnenberg, H.G., Bækkevold, E.S., Jahnsen, F.L., 2018. Transcriptional profiling reveals monocyte-related macrophages phenotypically resembling DC in human intestine. *Mucosal Immunol* 11, 1512–1523. <https://doi.org/10.1038/s41385-018-0060-1>
- Riedel, S., 2004. Biological warfare and bioterrorism: a historical review. *Proc (Bayl Univ Med Cent)* 17, 400–406.
- Rienks, M., Carai, P., Bitsch, N., Schellings, M., Vanhaverbeke, M., Verjans, J., Cuijpers, I., Heymans, S., Papageorgiou, A., 2017. Sema3A promotes the resolution of cardiac inflammation after myocardial infarction. *Basic Res Cardiol* 112, 42.  
<https://doi.org/10.1007/s00395-017-0630-5>
- Riss, T., Niles, A., Moravec, R., Karassina, N., Vidugiriene, J., 2004. Cytotoxicity Assays: In Vitro Methods to Measure Dead Cells, in: Sittampalam, G.S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C.P., Baell, J., Bejcek, B., Caaveiro, J.M.M., Chung, T.D.Y., Coussens, N.P., Dahlin, J.L., Devanaryan, V., Foley, T.L., Glicksman, M., Hall, M.D., Haas, J.V., Hoare, S.R.J., Inglese, J., Iversen, P.W., Kahl,

- S.D., Kales, S.C., Kirshner, S., Lal-Nag, M., Li, Z., McGee, J., McManus, O., Riss, T., Saradjian, P., Trask, O.J., Weidner, J.R., Wildey, M.J., Xia, M., Xu, X. (Eds.), *Assay Guidance Manual*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD).
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Robert-Gangneux, F., Dardé, M.-L., 2012. Epidemiology of and Diagnostic Strategies for Toxoplasmosis. *Clin. Microbiol. Rev.* 25, 264–296. <https://doi.org/10.1128/CMR.05013-11>
- Roberts, T.L., Idris, A., Dunn, J.A., Kelly, G.M., Burnton, C.M., Hodgson, S., Hardy, L.L., Garceau, V., Sweet, M.J., Ross, I.L., Hume, D.A., Stacey, K.J., 2009. HIN-200 Proteins Regulate Caspase Activation in Response to Foreign Cytoplasmic DNA. *Science* 323, 1057–1060. <https://doi.org/10.1126/science.1169841>
- Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., Taylor, J.M., 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* 13, 484. <https://doi.org/10.1186/1471-2164-13-484>
- Roca, F.J., Ramakrishnan, L., 2013. TNF Dually Mediates Resistance and Susceptibility to Mycobacteria via Mitochondrial Reactive Oxygen Species. *Cell* 153, 521–534. <https://doi.org/10.1016/j.cell.2013.03.022>
- Rodríguez-Mozaz, S., Chamorro, S., Marti, E., Huerta, B., Gros, M., Sànchez-Melsió, A., Borrego, C.M., Barceló, D., Balcázar, J.L., 2015. Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. *Water Research* 69, 234–242. <https://doi.org/10.1016/j.watres.2014.11.021>
- Rodríguez-Pastor, R., Escudero, R., Vidal, D., Mougeot, F., Arroyo, B., Lambin, X., Vila-Coro, A.M., Rodríguez-Moreno, I., Anda, P., Luque-Larena, J.J., 2017. Density-Dependent Prevalence of *Francisella tularensis* in Fluctuating Vole Populations, Northwestern Spain. *Emerg Infect Dis* 23, 1377–1379. <https://doi.org/10.3201/eid2308.161194>
- Rojas, R., van Vlijmen, T., Mardones, G.A., Prabhu, Y., Rojas, A.L., Mohammed, S., Heck, A.J.R., Raposo, G., van der Sluijs, P., Bonifacino, J.S., 2008. Regulation of retromer recruitment to endosomes by sequential action of Rab5 and Rab7. *Journal of Cell Biology* 183, 513–526. <https://doi.org/10.1083/jcb.200804048>
- Roos, D., Boer, M., 2014. Molecular diagnosis of chronic granulomatous disease. *Clin Exp Immunol* 175, 139–149. <https://doi.org/10.1111/cei.12202>
- Rossi, M., Duan, S., Jeong, Y.-T., Horn, M., Saraf, A., Florens, L., Washburn, M.P., Antebi, A., Pagano, M., 2013. Regulation of the CRL4Cdt2 Ubiquitin Ligase and Cell-Cycle Exit by the SCFFbxo11 Ubiquitin Ligase. *Molecular Cell* 49, 1159–1166. <https://doi.org/10.1016/j.molcel.2013.02.004>
- Rudolph, M.G., Weise, C., Miold, S., Hillenbrand, B., Bader, B., Wittinghofer, A., Hardt, W.-D., 1999. Biochemical Analysis of SopE from *Salmonella typhimurium*, a Highly Efficient Guanosine Nucleotide Exchange Factor for RhoGTPases. *J. Biol. Chem.* 274, 30501–30509. <https://doi.org/10.1074/jbc.274.43.30501>
- Russell, D.G., Huang, L., VanderVen, B.C., 2019. Immunometabolism at the interface between macrophages and pathogens. *Nature Reviews Immunology* 19, 291–304. <https://doi.org/10.1038/s41577-019-0124-9>
- Sakharkar, M.K., Chow, V.T.K., Kanguane, P., 2004. Distributions of Exons and Introns in the Human Genome. *In Silico Biology* 4, 387–393.

- Saliba, A.-E., Li, L., Westermann, A.J., Appenzeller, S., Stapels, D.A.C., Schulte, L.N., Helaine, S., Vogel, J., 2016. Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular Salmonella. *Nat Microbiol* 2, 16206. <https://doi.org/10.1038/nmicrobiol.2016.206>
- Saliminejad, K., Khorram Khorshid, H.R., Soleymani Fard, S., Ghaffari, S.H., 2019. An overview of microRNAs: Biology, functions, therapeutics, and analysis methods: SALIMINEJAD ET AL. *J Cell Physiol* 234, 5451–5465. <https://doi.org/10.1002/jcp.27486>
- Salinas, R.E., Ogohara, C., Thomas, M.I., Shukla, K.P., Miller, S.I., Ko, D.C., 2013. A cellular genome-wide association study reveals human variation in microtubule stability and a role in inflammatory cell death. *MBoC* 25, 76–86. <https://doi.org/10.1091/mbc.e13-06-0294>
- Salzberg, S.L., 2018. Open questions: How many genes do we have? *BMC Biology* 16, 94. <https://doi.org/10.1186/s12915-018-0564-x>
- Sandstrom, A., Mitchell, P.S., Goers, L., Mu, E.W., Lesser, C.F., Vance, R.E., 2019. Functional degradation: A mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes. *Science* 364. <https://doi.org/10.1126/science.aau1330>
- Sanjurjo, L., Aran, G., Téllez, É., Amézaga, N., Armengol, C., López, D., Prats, C., Sarrias, M.-R., 2018. CD5L Promotes M2 Macrophage Polarization through Autophagy-Mediated Upregulation of ID3. *Front. Immunol.* 9. <https://doi.org/10.3389/fimmu.2018.00480>
- Santic, M., Asare, R., Skrobonja, I., Jones, S., Kwaik, Y.A., 2008. Acquisition of the Vacuolar ATPase Proton Pump and Phagosome Acidification Are Essential for Escape of *Francisella tularensis* into the Macrophage Cytosol. *Infection and Immunity* 76, 2671–2677. <https://doi.org/10.1128/IAI.00185-08>
- Santos, A.L., Preta, G., 2018. Lipids in the cell: organisation regulates function. *Cell. Mol. Life Sci.* 75, 1909–1927. <https://doi.org/10.1007/s00018-018-2765-4>
- Sawada, N., Itoh, H., Yamashita, J., Doi, K., Inoue, M., Masatsugu, K., Fukunaga, Y., Sakaguchi, S., Sone, M., Yamahara, K., Yurugi, T., Nakao, K., 2001. cGMP-dependent protein kinase phosphorylates and inactivates RhoA. *Biochem. Biophys. Res. Commun.* 280, 798–805. <https://doi.org/10.1006/bbrc.2000.4194>
- Schindler, C., Chen, Y., Pu, J., Guo, X., Bonifacino, J.S., 2015. EARP is a multisubunit tethering complex involved in endocytic recycling. *Nature Cell Biology* 17, 639–650. <https://doi.org/10.1038/ncb3129>
- Schiroli, G., Conti, A., Ferrari, S., della Volpe, L., Jacob, A., Albano, L., Beretta, S., Calabria, A., Vavassori, V., Gasparini, P., Salataj, E., Ndiaye-Lobry, D., Brombin, C., Chaumeil, J., Montini, E., Merelli, I., Genovese, P., Naldini, L., Di Micco, R., 2019. Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response. *Cell Stem Cell* 24, 551-565.e8. <https://doi.org/10.1016/j.stem.2019.02.019>
- Schmid-Burgk, J.L., Gaidt, M.M., Schmidt, T., Ebert, T.S., Bartok, E., Hornung, V., 2015. Caspase-4 mediates non-canonical activation of the NLRP3 inflammasome in human myeloid cells. *European Journal of Immunology* 45, 2911–2917. <https://doi.org/10.1002/eji.201545523>
- Schnupf, P., Sansonetti, P.J., 2019. Shigella Pathogenesis: New Insights through Advanced Methodologies. *Microbiology Spectrum* 7. <https://doi.org/10.1128/microbiolspec.BAI-0023-2019>
- Schroder, W.A., Gardner, J., Le, T.T., Duke, M., Burke, M.L., Jones, M.K., McMANUS, D.P., Suhrbier, A., 2010. SerpinB2 deficiency modulates Th1/Th2 responses after

- schistosome infection. *Parasite Immunology* 32, 764–768.  
<https://doi.org/10.1111/j.1365-3024.2010.01241.x>
- Schroder, W.A., Hirata, T.D., Le, T.T., Gardner, J., Boyle, G.M., Ellis, J., Nakayama, E., Pathirana, D., Nakaya, H.I., Suhrbier, A., 2019a. SerpinB2 inhibits migration and promotes a resolution phase signature in large peritoneal macrophages. *Sci Rep* 9, 12421. <https://doi.org/10.1038/s41598-019-48741-w>
- Schroder, W.A., Le, T.T., Gardner, J., Andrews, R.K., Gardiner, E.E., Callaway, L., Suhrbier, A., 2019b. SerpinB2 deficiency in mice reduces bleeding times via dysregulated platelet activation. *Platelets* 30, 658–663. <https://doi.org/10.1080/09537104.2018.1535702>
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., Ragg, T., 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 7, 3. <https://doi.org/10.1186/1471-2199-7-3>
- Seaman, M.N.J., 2018. Retromer and the cation-independent mannose 6-phosphate receptor—Time for a trial separation? *Traffic* 19, 150–152. <https://doi.org/10.1111/tra.12542>
- Segura, E., Touzot, M., Bohineust, A., Cappuccio, A., Chiochia, G., Hosmalin, A., Dalod, M., Soumelis, V., Amigorena, S., 2013. Human Inflammatory Dendritic Cells Induce Th17 Cell Differentiation. *Immunity* 38, 336–348. <https://doi.org/10.1016/j.immuni.2012.10.018>
- Seidman, K.J., Barsuk, J.H., Johnson, R.F., Weyhenmeyer, J.A., 1996. Differentiation of NG108-15 neuroblastoma cells by serum starvation or dimethyl sulfoxide results in marked differences in angiotensin II receptor subtype expression. *J. Neurochem.* 66, 1011–1018. <https://doi.org/10.1046/j.1471-4159.1996.66031011.x>
- Seixas, E., Escrevente, C., Seabra, M.C., Barral, D.C., 2018. Rab GTPase regulation of bacteria and protozoa phagocytosis occurs through the modulation of phagocytic receptor surface expression. *Sci Rep* 8, 12998. <https://doi.org/10.1038/s41598-018-31171-5>
- Selle, A., Ullrich, O., Harnacke, K., Hass, R., 2007. Retrodifferentiation and rejuvenation of senescent monocytic cells requires PARP-1. *Experimental Gerontology* 42, 554–562. <https://doi.org/10.1016/j.exger.2006.12.004>
- Sellin, M.E., Müller, A.A., Felmy, B., Dolowschiak, T., Diard, M., Tardivel, A., Maslowski, K.M., Hardt, W.-D., 2014. Epithelium-Intrinsic NAIP/NLRC4 Inflammasome Drives Infected Enterocyte Expulsion to Restrict Salmonella Replication in the Intestinal Mucosa. *Cell Host & Microbe* 16, 237–248. <https://doi.org/10.1016/j.chom.2014.07.001>
- Seo, J.-W., Yang, E.-J., Yoo, K.-H., Choi, I.-H., 2015. Macrophage Differentiation from Monocytes Is Influenced by the Lipid Oxidation Degree of Low Density Lipoprotein [WWW Document]. *Mediators of Inflammation*. <https://doi.org/10.1155/2015/235797>
- Serbina, N.V., Salazar-Mather, T.P., Biron, C.A., Kuziel, W.A., Pamer, E.G., 2003. TNF/iNOS-Producing Dendritic Cells Mediate Innate Immune Defense against Bacterial Infection. *Immunity* 19, 59–70. [https://doi.org/10.1016/S1074-7613\(03\)00171-7](https://doi.org/10.1016/S1074-7613(03)00171-7)
- Syednasrollah, F., Laiho, A., Elo, L.L., 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* 16, 59–70. <https://doi.org/10.1093/bib/bbt086>
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., Zhang, F., 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 343, 84–87. <https://doi.org/10.1126/science.1247005>

- Sharma, S., Kelly, T.K., Jones, P.A., 2010. Epigenetics in cancer. *Carcinogenesis* 31, 27–36. <https://doi.org/10.1093/carcin/bgp220>
- Shaw, G., Morse, S., Ararat, M., Graham, F.L., 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *The FASEB Journal* 16, 869–871. <https://doi.org/10.1096/fj.01-0995fje>
- Shi, J., Zhao, Y., Wang, Y., Gao, W., Ding, J., Li, P., Hu, L., Shao, F., 2014. Inflammatory caspases are innate immune receptors for intracellular LPS. *Nature* 514, 187–192. <https://doi.org/10.1038/nature13683>
- Shi, L., Chowdhury, S.M., Smallwood, H.S., Yoon, H., Mottaz-Brewer, H.M., Norbeck, A.D., McDermott, J.E., Clauss, T.R.W., Heffron, F., Smith, R.D., Adkins, J.N., 2009. Proteomic Investigation of the Time Course Responses of RAW 264.7 Macrophages to Infection with *Salmonella enterica*. *Infection and Immunity* 77, 3227–3233. <https://doi.org/10.1128/IAI.00063-09>
- Shih, P.-Y., Lee, S.-P., Chen, Y.-K., Hsueh, Y.-P., 2014. Cortactin-binding protein 2 increases microtubule stability and regulates dendritic arborization. *J Cell Sci* 127, 3521–3534. <https://doi.org/10.1242/jcs.149476>
- Shimada, K., Crother, T.R., Karlin, J., Dagvadorj, J., Chiba, N., Chen, S., Ramanujan, V.K., Wolf, A.J., Vergnes, L., Ojcius, D.M., Rentsendorj, A., Vargas, M., Guerrero, C., Wang, Y., Fitzgerald, K.A., Underhill, D.M., Town, T., Arditi, M., 2012. Oxidized Mitochondrial DNA Activates the NLRP3 Inflammasome during Apoptosis. *Immunity* 36, 401–414. <https://doi.org/10.1016/j.immuni.2012.01.009>
- Shinoda, H., Shannon, M., Nagai, T., 2018. Fluorescent Proteins for Investigating Biological Events in Acidic Environments. *Int J Mol Sci* 19. <https://doi.org/10.3390/ijms19061548>
- Shirey, K.A., Cole, L.E., Keegan, A.D., Vogel, S.N., 2008. *Francisella tularensis* Live Vaccine Strain Induces Macrophage Alternative Activation as a Survival Mechanism. *The Journal of Immunology* 181, 4159–4167. <https://doi.org/10.4049/jimmunol.181.6.4159>
- Shuai-Cheng, W., Ben-Dong, F., Xiu-Ling, C., Jian-Qing, S., Yun-Xing, F., Zhen-Qiang, C., Dao-Xiu, X., Zong-Mei, W., 2016. Subinhibitory concentrations of phloretin repress the virulence of *Salmonella typhimurium* and protect against *Salmonella typhimurium* infection. *Antonie van Leeuwenhoek* 109, 1503–1512. <https://doi.org/10.1007/s10482-016-0752-z>
- Sidik, S.M., Salsman, J., Delleire, G., Rohde, J.R., 2015. *Shigella* Infection Interferes with SUMOylation and Increases PML-NB Number. *PLOS ONE* 10, e0122585. <https://doi.org/10.1371/journal.pone.0122585>
- Silva, M.T., 2012. Classical Labeling of Bacterial Pathogens According to Their Lifestyle in the Host: Inconsistencies and Alternatives. *Front Microbiol* 3. <https://doi.org/10.3389/fmicb.2012.00071>
- Silva-Gomes, S., Appelberg, R., Larsen, R., Soares, M.P., Gomes, M.S., 2013. Heme Catabolism by Heme Oxygenase-1 Confers Host Resistance to *Mycobacterium* Infection. *Infection and Immunity* 81, 2536–2545. <https://doi.org/10.1128/IAI.00251-13>
- Simonetti, B., Paul, B., Chaudhari, K., Weeratunga, S., Steinberg, F., Gorla, M., Heesom, K.J., Bashaw, G.J., Collins, B.M., Cullen, P.J., 2019. Molecular identification of a BAR domain-containing coat complex for endosomal recycling of transmembrane proteins. *Nature Cell Biology* 21, 1219–1233. <https://doi.org/10.1038/s41556-019-0393-3>

- Singh, A., Periasamy, S., Malik, M., Bakshi, C.S., Stephen, L., Ault, J.G., Mannella, C.A., Sellati, T.J., 2017. Necroptotic debris including damaged mitochondria elicits sepsis-like syndrome during late-phase tularemia. *Cell Death Discovery* 3, 1–12. <https://doi.org/10.1038/cddiscovery.2017.56>
- Singh, S., Wu, X., Ljosa, V., Bray, M.-A., Piccioni, F., Root, D.E., Doench, J.G., Boehm, J.S., Carpenter, A.E., 2015. Morphological Profiles of RNAi-Induced Gene Knockdown Are Highly Reproducible but Dominated by Seed Effects. *PLOS ONE* 10, e0131370. <https://doi.org/10.1371/journal.pone.0131370>
- Sjöstrand, M., Carow, B., Nyberg, W.A., Covacu, R., Rottenberg, M.E., Espinosa, A., 2020. TRIM21 controls Toll-like receptor 2 responses in bone-marrow-derived macrophages. *Immunology* 159, 335–343. <https://doi.org/10.1111/imm.13157>
- Smith-Unna, R., Bournsnel, C., Patro, R., Hibberd, J.M., Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. <https://doi.org/10.1101/gr.196469.115>
- Soneson, C., Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* 14, 91. <https://doi.org/10.1186/1471-2105-14-91>
- Soscia, S.J., Kirby, J.E., Washicosky, K.J., Tucker, S.M., Ingelsson, M., Hyman, B., Burton, M.A., Goldstein, L.E., Duong, S., Tanzi, R.E., Moir, R.D., 2010. The Alzheimer's Disease-Associated Amyloid  $\beta$ -Protein Is an Antimicrobial Peptide. *PLOS ONE* 5, e9505. <https://doi.org/10.1371/journal.pone.0009505>
- Sossey-Alaoui, K., Pluskota, E., Szpak, D., Plow, E.F., 2019. The Kindlin2-p53-SerpinB2 signaling axis is required for cellular senescence in breast cancer. *Cell Death Dis* 10, 539. <https://doi.org/10.1038/s41419-019-1774-z>
- Sousa, S., Cabanes, D., Bournès, L., Lecuit, M., Sansonetti, P., Tran-Van-Nhieu, G., Cossart, P., 2007. Src, cortactin and Arp2/3 complex are required for E-cadherin-mediated internalization of *Listeria* into cells. *Cell. Microbiol.* 9, 2629–2643. <https://doi.org/10.1111/j.1462-5822.2007.00984.x>
- Sperandio, B., Regnault, B., Guo, J., Zhang, Z., Stanley, S.L., Sansonetti, P.J., Pédrón, T., 2008. Virulent *Shigella flexneri* subverts the host innate immune response through manipulation of antimicrobial peptide gene expression. *J Exp Med* 205, 1121–1132. <https://doi.org/10.1084/jem.20071698>
- Stapels, D.A.C., Hill, P.W.S., Westermann, A.J., Fisher, R.A., Thurston, T.L., Saliba, A.-E., Blommestein, I., Vogel, J., Helaine, S., 2018. Salmonella persists undermine host immune defenses during antibiotic treatment. *Science* 362, 1156–1160. <https://doi.org/10.1126/science.aat7148>
- Starr, T., Bauler, T.J., Malik-Kale, P., Steele-Mortimer, O., 2018. The phorbol 12-myristate-13-acetate differentiation protocol is critical to the interaction of THP-1 macrophages with *Salmonella Typhimurium*. *PLOS ONE* 13, e0193601. <https://doi.org/10.1371/journal.pone.0193601>
- Starr, T., Sun, Y., Wilkins, N., Storrie, B., 2010. Rab33b and Rab6 are Functionally Overlapping Regulators of Golgi Homeostasis and Trafficking. *Traffic* 11, 626–636. <https://doi.org/10.1111/j.1600-0854.2010.01051.x>
- St-Denis, N., Gupta, G.D., Lin, Z.Y., Gonzalez-Badillo, B., Veri, A.O., Knight, J.D.R., Rajendran, D., Couzens, A.L., Currie, K.W., Tkach, J.M., Cheung, S.W.T., Pelletier, L., Gingras, A.-C., 2016. Phenotypic and Interaction Profiling of the Human Phosphatases Identifies Diverse Mitotic Regulators. *Cell Reports* 17, 2488–2501. <https://doi.org/10.1016/j.celrep.2016.10.078>

- Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M., Barthel, M., Kremer, M., Chaffron, S., Macpherson, A.J., Buer, J., Parkhill, J., Dougan, G., Mering, C. von, Hardt, W.-D., 2007. Salmonella enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. *PLOS Biology* 5, e244. <https://doi.org/10.1371/journal.pbio.0050244>
- Stepanenko, A.A., Dmitrenko, V.V., 2015. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene* 569, 182–190. <https://doi.org/10.1016/j.gene.2015.05.065>
- Stevanin, T.M., Poole, R.K., Demoncheaux, E.A.G., Read, R.C., 2002. Flavohemoglobin Hmp Protects Salmonella enterica Serovar Typhimurium from Nitric Oxide-Related Killing by Human Macrophages. *Infection and Immunity* 70, 4399–4405. <https://doi.org/10.1128/IAI.70.8.4399-4405.2002>
- Stévenin, V., Chang, Y.-Y., Toquin, Y.L., Duchateau, M., Gianetto, Q.G., Luk, C.H., Salles, A., Sohst, V., Matondo, M., Reiling, N., Enninga, J., 2019. Dynamic Growth and Shrinkage of the Salmonella-Containing Vacuole Determines the Intracellular Pathogen Niche. *Cell Reports* 29, 3958–3973.e7. <https://doi.org/10.1016/j.celrep.2019.11.049>
- Stevenson, K., McVey, A.F., Clark, I.B.N., Swain, P.S., Pilizota, T., 2016. General calibration of microbial growth in microplate readers. *Scientific Reports* 6, 38828. <https://doi.org/10.1038/srep38828>
- Stewart, M.K., Cummings, L.A., Johnson, M.L., Berezow, A.B., Cookson, B.T., 2011. Regulation of phenotypic heterogeneity permits Salmonella evasion of the host caspase-1 inflammatory response. *PNAS* 108, 20742–20747. <https://doi.org/10.1073/pnas.1108963108>
- Storek, K.M., Gertszolf, N.A., Ohlson, M.B., Monack, D.M., 2015. cGAS and Ifi204 Cooperate To Produce Type I IFNs in Response to Francisella Infection. *The Journal of Immunology* 194, 3236–3245. <https://doi.org/10.4049/jimmunol.1402764>
- Su, X.-D., Shuai, Y., 2020. Macromolecules and Antibody-Based Drugs, in: Xu, J. (Ed.), Regulation of Cancer Immune Checkpoints: Molecular and Cellular Mechanisms and Therapy, *Advances in Experimental Medicine and Biology*. Springer, Singapore, pp. 485–530. [https://doi.org/10.1007/978-981-15-3266-5\\_20](https://doi.org/10.1007/978-981-15-3266-5_20)
- Sulahian, T.H., Imrich, A., DeLoid, G., Winkler, A.R., Kobzik, L., 2008. Signaling pathways required for macrophage scavenger receptor-mediated phagocytosis: analysis by scanning cytometry. *Respiratory Research* 9, 59. <https://doi.org/10.1186/1465-9921-9-59>
- Sun, J., Jin, L., He, T., Wei, Z., Liu, X., Zhu, L., Li, X., 2020. Antibiotic resistance genes (ARGs) in agricultural soils from the Yangtze River Delta, China. *Science of The Total Environment* 740, 140001. <https://doi.org/10.1016/j.scitotenv.2020.140001>
- Sun, L., Wu, J., Du, F., Chen, X., Chen, Z.J., 2013. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science* 339, 786–791. <https://doi.org/10.1126/science.1232458>
- Sun, Q., Yong, X., Sun, X., Yang, F., Dai, Z., Gong, Y., Zhou, L., Zhang, X., Niu, D., Dai, L., Liu, J.-J., Jia, D., 2017. Structural and functional insights into sorting nexin 5/6 interaction with bacterial effector IncE. *Signal Transduct Target Ther* 2, 17030. <https://doi.org/10.1038/sigtrans.2017.30>
- Surks, H.K., Mochizuki, N., Kasai, Y., Georgescu, S.P., Tang, K.M., Ito, M., Lincoln, T.M., Mendelsohn, M.E., 1999. Regulation of myosin phosphatase by a specific interaction with cGMP-dependent protein kinase I $\alpha$ . *Science* 286, 1583–1587.

- Suvarnapunya, A.E., Stein, M.A., 2005. DNA base excision repair potentiates the protective effect of Salmonella Pathogenicity Island 2 within macrophages. *Microbiology*, 151, 557–567. <https://doi.org/10.1099/mic.0.27555-0>
- Suzuki, S., Mimuro, H., Kim, M., Ogawa, M., Ashida, H., Toyotome, T., Franchi, L., Suzuki, M., Sanada, T., Suzuki, T., Tsutsui, H., Núñez, G., Sasakawa, C., 2014. Shigella IpaH7.8 E3 ubiquitin ligase targets glomulin and activates inflammasomes to demolish macrophages. *PNAS* 111, E4254–E4263. <https://doi.org/10.1073/pnas.1324021111>
- Swanson, K.V., Deng, M., Ting, J.P.-Y., 2019. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nature Reviews Immunology* 19, 477–489. <https://doi.org/10.1038/s41577-019-0165-0>
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C. von, 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607–D613. <https://doi.org/10.1093/nar/gky1131>
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* 45, D362–D368. <https://doi.org/10.1093/nar/gkw937>
- Tabata, H., Morita, H., Kaji, H., Tohyama, K., Tohyama, Y., 2020. Syk facilitates phagosome-lysosome fusion by regulating actin-remodeling in complement-mediated phagocytosis. *Sci Rep* 10, 22086. <https://doi.org/10.1038/s41598-020-79156-7>
- Tacconelli, E., Magrini, N., 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. WHO.
- Tahoun, A., Mahajan, S., Paxton, E., Malterer, G., Donaldson, D.S., Wang, D., Tan, A., Gillespie, T.L., O’Shea, M., Roe, A.J., Shaw, D.J., Gally, D.L., Lengeling, A., Mabbott, N.A., Haas, J., Mahajan, A., 2012. Salmonella Transforms Follicle-Associated Epithelial Cells into M Cells to Promote Intestinal Invasion. *Cell Host & Microbe* 12, 645–656. <https://doi.org/10.1016/j.chom.2012.10.009>
- Tang-Huau, T.-L., Segura, E., 2019. Human in vivo-differentiated monocyte-derived dendritic cells. *Seminars in Cell & Developmental Biology*, SI: Human dendritic cells 86, 44–49. <https://doi.org/10.1016/j.semcd.2018.02.018>
- Tannahill, G.M., Curtis, A.M., Adamik, J., Palsson-McDermott, E.M., McGettrick, A.F., Goel, G., Frezza, C., Bernard, N.J., Kelly, B., Foley, N.H., Zheng, L., Gardet, A., Tong, Z., Jany, S.S., Corr, S.C., Haneklaus, M., Caffrey, B.E., Pierce, K., Walmsley, S., Beasley, F.C., Cummins, E., Nizet, V., Whyte, M., Taylor, C.T., Lin, H., Masters, S.L., Gottlieb, E., Kelly, V.P., Clish, C., Auron, P.E., Xavier, R.J., O’Neill, L. a. J., 2013. Succinate is an inflammatory signal that induces IL-1 $\beta$  through HIF-1 $\alpha$ . *Nature* 496, 238–242. <https://doi.org/10.1038/nature11986>
- Tarique, A.A., Logan, J., Thomas, E., Holt, P.G., Sly, P.D., Fantino, E., 2015. Phenotypic, Functional, and Plasticity Features of Classical and Alternatively Activated Human Macrophages. *Am J Respir Cell Mol Biol* 53, 676–688. <https://doi.org/10.1165/rcmb.2015-0012OC>
- Tarran, R., Button, B., Picher, M., Paradiso, A.M., Ribeiro, C.M., Lazarowski, E.R., Zhang, L., Collins, P.L., Pickles, R.J., Fredberg, J.J., Boucher, R.C., 2005. Normal and Cystic Fibrosis Airway Surface Liquid Homeostasis THE EFFECTS OF PHASIC SHEAR STRESS AND VIRAL INFECTIONS. *J. Biol. Chem.* 280, 35751–35759. <https://doi.org/10.1074/jbc.M505832200>

- Tauseef, M., Knezevic, N., Chava, K.R., Smith, M., Sukriti, S., Gianaris, N., Obukhov, A.G., Vogel, S.M., Schraufnagel, D.E., Dietrich, A., Birnbaumer, L., Malik, A.B., Mehta, D., 2012. TLR4 activation of TRPC6-dependent calcium signaling mediates endotoxin-induced lung vascular permeability and inflammation. *J Exp Med* 209, 1953–1968. <https://doi.org/10.1084/jem.20111355>
- Tedesco, S., De Majo, F., Kim, J., Trenti, A., Trevisi, L., Fadini, G.P., Bolego, C., Zandstra, P.W., Cignarella, A., Vitiello, L., 2018. Convenience versus Biological Significance: Are PMA-Differentiated THP-1 Cells a Reliable Substitute for Blood-Derived Macrophages When Studying in Vitro Polarization? *Front. Pharmacol.* 9. <https://doi.org/10.3389/fphar.2018.00071>
- The Aurum Institute NPC, 2019. A Ph2 Randomized Trial to Evaluate the Safety Preliminary Efficacy and Biomarker Response of Host Directed Therapies Added to Rifabutin-modified Standard Therapy in Adults With Drug-Sensitive Smear-Positive Pulmonary TB (Clinical trial registration No. study/NCT02968927). [clinicaltrials.gov](https://clinicaltrials.gov).
- The BBC, 2017. Antibiotic resistance very “risky” [WWW Document]. BBC News. URL <http://www.bbc.co.uk/news/av/uk-41607281/dame-sally-davies-antibiotic-resistance-very-risky> (accessed 10.26.17).
- The Gene Ontology Consortium, 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 47, D330–D338. <https://doi.org/10.1093/nar/gky1055>
- The Gene Ontology Consortium, Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. *Nat Genet* 25, 25–29. <https://doi.org/10.1038/75556>
- The Human Protein Atlas, 2017. The Human Protein Atlas [WWW Document]. URL <http://www.proteinatlas.org/> (accessed 7.20.17).
- The UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49, D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- The Uniprot Consortium, 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47, D506–D515. <https://doi.org/10.1093/nar/gky1049>
- The Uniprot Consortium, 2015. UniProt: a hub for protein information. *Nucl. Acids Res.* 43, D204–D212. <https://doi.org/10.1093/nar/gku989>
- Thomas, M.P., Liu, X., Whangbo, J., McCrossan, G., Sanborn, K.B., Basar, E., Walch, M., Lieberman, J., 2015. Apoptosis Triggers Specific, Rapid, and Global mRNA Decay with 3' Uridylated Intermediates Degraded by DIS3L2. *Cell Rep* 11, 1079–1089. <https://doi.org/10.1016/j.celrep.2015.04.026>
- Thomas, P.D., 2017. The Gene Ontology and the Meaning of Biological Function, in: Dessimoz, C., Škunca, N. (Eds.), *The Gene Ontology Handbook, Methods in Molecular Biology*. Springer, New York, NY, pp. 15–24. [https://doi.org/10.1007/978-1-4939-3743-1\\_2](https://doi.org/10.1007/978-1-4939-3743-1_2)
- Thornbrough, J.M., Gopinath, A., Hundley, T., Worley, M.J., 2016. Human Genome-Wide RNAi Screen for Host Factors That Facilitate Salmonella Invasion Reveals a Role for Potassium Secretion in Promoting Internalization. *PLOS ONE* 11, e0166916. <https://doi.org/10.1371/journal.pone.0166916>
- Thornbrough, J.M., Hundley, T., Valdivia, R., Worley, M.J., 2012. Human Genome-Wide RNAi Screen for Host Factors That Modulate Intracellular Salmonella Growth. *PLOS ONE* 7, e38097. <https://doi.org/10.1371/journal.pone.0038097>

- Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H.A., Alm, T., Asplund, A., Björk, L., Breckels, L.M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C.M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J.M., Sivertsson, Å., Sjöstedt, E., Skogs, M., Stadler, C., Sullivan, D.P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Pontén, F., Feilitzten, K. von, Lilley, K.S., Uhlén, M., Lundberg, E., 2017. A subcellular map of the human proteome. *Science* 356. <https://doi.org/10.1126/science.aal3321>
- Thurston, T.L.M., Matthews, S.A., Jennings, E., Alix, E., Shao, F., Shenoy, A.R., Birrell, M.A., Holden, D.W., 2016. Growth inhibition of cytosolic Salmonella by caspase-1 and caspase-11 precedes host cell death. *Nat Commun* 7, 1–15. <https://doi.org/10.1038/ncomms13292>
- Tian, X., Gan, H., Zeng, Y., Zhao, H., Tang, R., Xia, Y., 2018. Inhibition of semaphorin-3a suppresses lipopolysaccharide-induced acute kidney injury. *J Mol Med* 96, 713–724. <https://doi.org/10.1007/s00109-018-1653-6>
- Tommasi, R., Brown, D.G., Walkup, G.K., Manchester, J.I., Miller, A.A., 2015. ESKAPEing the labyrinth of antibacterial discovery. *Nat Rev Drug Discov* 14, 529–542. <https://doi.org/10.1038/nrd4572>
- Travassos, L.H., Carneiro, L.A.M., Ramjeet, M., Hussey, S., Kim, Y.-G., Magalhães, J.G., Yuan, L., Soares, F., Chea, E., Le Bourhis, L., Boneca, I.G., Allaoui, A., Jones, N.L., Nuñez, G., Girardin, S.E., Philpott, D.J., 2010. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunology* 11, 55–62. <https://doi.org/10.1038/ni.1823>
- Tretina, K., Park, E.-S., Maminska, A., MacMicking, J.D., 2019. Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J Exp Med* 216, 482–500. <https://doi.org/10.1084/jem.20182031>
- Tripal, P., Bauer, M., Naschberger, E., Mörtinger, T., Hohenadl, C., Cornali, E., Thurau, M., Stürzl, M., 2007. Unique Features of Different Members of the Human Guanylate-Binding Protein Family. *Journal of Interferon & Cytokine Research* 27, 44–52. <https://doi.org/10.1089/jir.2007.0086>
- Tsai, S.-C., Seto, E., 2002. Regulation of Histone Deacetylase 2 by Protein Kinase CK2. *J. Biol. Chem.* 277, 31826–31833. <https://doi.org/10.1074/jbc.M204149200>
- Tsakiridis, A., Tzouanacou, E., Rahman, A., Colby, D., Axton, R., Chambers, I., Wilson, V., Forrester, L., Brickman, J.M., 2009. Expression-independent gene trap vectors for random and targeted mutagenesis in embryonic stem cells. *Nucl. Acids Res.* 37, e129–e129. <https://doi.org/10.1093/nar/gkp640>
- Turco, M.Y., Furia, L., Dietze, A., Diaz, L.F., Ronzoni, S., Sciallo, A., Simeone, A., Constam, D., Faretta, M., Lanfrancone, L., 2012. Cellular Heterogeneity During Embryonic Stem Cell Differentiation to Epiblast Stem Cells Is Revealed by the ShcD/RaLP Adaptor Protein. *STEM CELLS* 30, 2423–2436. <https://doi.org/10.1002/stem.1217>
- Tusiimire, J., Wallace, J., Woods, N., Dufton, M.J., Parkinson, J.A., Abbott, G., Clements, C.J., Young, L., Park, J.K., Jeon, J.W., Ferro, V.A., Watson, D.G., 2016. Effect of Bee Venom and Its Fractions on the Release of Pro-Inflammatory Cytokines in PMA-Differentiated U937 Cells Co-Stimulated with LPS. *Vaccines* 4, 11. <https://doi.org/10.3390/vaccines4020011>
- Uc, A., Giriappa, R., Meyerholz, D.K., Griffin, M., Ostedgaard, L.S., Tang, X.X., Abu-El-Haija, Marwa, Stoltz, D.A., Ludwig, P., Pezzulo, A., Abu-El-Haija, Maisam, Taft, P., Welsh, M.J., 2012. Pancreatic and biliary secretion are both altered in cystic fibrosis pigs.

- Am J Physiol Gastrointest Liver Physiol 303, G961–G968.  
<https://doi.org/10.1152/ajpgi.00030.2012>
- Udofa, E.A., Stringer, B.W., Gade, P., Mahony, D., Buzza, M.S., Kalvakolanu, D.V., Antalis, T.M., 2013. The transcription factor C/EBP- $\beta$  mediates constitutive and LPS-inducible transcription of murine SerpinB2. *PLoS One* 8, e57855.  
<https://doi.org/10.1371/journal.pone.0057855>
- Uemura, T., Waguri, S., 2020. Emerging roles of Golgi/endosome-localizing monomeric clathrin adaptors GGAs. *Anat Sci Int* 95, 12–21. <https://doi.org/10.1007/s12565-019-00505-2>
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigartyo, C.A.-K., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., Feilitzén, K. von, Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Heijne, G. von, Nielsen, J., Pontén, F., 2015. Tissue-based map of the human proteome. *Science* 347. <https://doi.org/10.1126/science.1260419>
- Vaandrager, A.B., Smolenski, A., Tilly, B.C., Houtsmuller, A.B., Ehlert, E.M., Bot, A.G., Edixhoven, M., Boomaars, W.E., Lohmann, S.M., de Jonge, H.R., 1998. Membrane targeting of cGMP-dependent protein kinase is required for cystic fibrosis transmembrane conductance regulator Cl<sup>-</sup> channel activation. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1466–1471.
- Vaandrager, A.B., Tilly, B.C., Smolenski, A., Schneider-Rasp, S., Bot, A.G., Edixhoven, M., Scholte, B.J., Jarchau, T., Walter, U., Lohmann, S.M., Poller, W.C., de Jonge, H.R., 1997. cGMP stimulation of cystic fibrosis transmembrane conductance regulator Cl<sup>-</sup> channels co-expressed with cGMP-dependent protein kinase type II but not type I $\beta$ . *J. Biol. Chem.* 272, 4195–4200.
- Valdés López, J.F., Urcuqui-Inchima, S., 2018. Synergism between phorbol-12-myristate-13-acetate and vitamin D3 in the differentiation of U937 cells to monocytes and macrophages. *Morphologie, 100E CONGRÈS DE L'ASSOCIATION DES MORPHOLOGISTES 23ES JOURNÉES DU COLLÈGE DES HISTOLOGISTES, EMBRYOLOGISTES ET CYTOGÉNÉTIENS, CAEN 22-24 MARS 2018* 102, 205–218. <https://doi.org/10.1016/j.morpho.2018.06.001>
- van Roy, F., Berx, G., 2008. The cell-cell adhesion molecule E-cadherin. *Cell. Mol. Life Sci.* 65, 3756–3788. <https://doi.org/10.1007/s00018-008-8281-1>
- Vasse, G.F., Kühn, P.T., Zhou, Q., Bhusari, S.A., Reker-Smit, C., Melgert, B.N., van Rijn, P., 2018. Collagen morphology influences macrophage shape and marker expression in vitro. *Journal of Immunology and Regenerative Medicine* 1, 13–20.  
<https://doi.org/10.1016/j.regen.2018.01.002>
- Vázquez, C.L., Rodgers, A., Herbst, S., Coade, S., Gronow, A., Guzman, C.A., Wilson, M.S., Kanzaki, M., Nykjaer, A., Gutierrez, M.G., 2016. The proneurotrophin receptor sortilin is required for Mycobacterium tuberculosis control by macrophages. *Scientific Reports* 6, 1–13. <https://doi.org/10.1038/srep29332>
- Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Holden, D.W., Lucia, S.M., Dinauer, M.C., Mastroeni, P., Fang, F.C., 2000. Salmonella Pathogenicity Island 2-Dependent Evasion of the Phagocyte NADPH Oxidase. *Science* 287, 1655–1658.  
<https://doi.org/10.1126/science.287.5458.1655>
- Vehmas, A., Lieu, J., Pardo, C.A., McArthur, J.C., Gartner, S., 2004. Amyloid precursor protein expression in circulating monocytes and brain macrophages from patients

- with HIV-associated cognitive impairment. *J. Neuroimmunol.* 157, 99–110.  
<https://doi.org/10.1016/j.jneuroim.2004.08.035>
- Vestal, D.J., Gorbacheva, V.Y., Sen, G.C., 2000. Different Subcellular Localizations for the Related Interferon-Induced GTPases, MuGBP-1 and MuGBP-2: Implications for Different Functions? *Journal of Interferon & Cytokine Research* 20, 991–1000.  
<https://doi.org/10.1089/10799900050198435>
- Vij, N., Mazur, S., Zeitlin, P.L., 2009. CFTR Is a Negative Regulator of NFκB Mediated Innate Immune Response. *PLOS ONE* 4, e4664.  
<https://doi.org/10.1371/journal.pone.0004664>
- Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., Assmus, H.E., Andrade-Navarro, M.A., Wanker, E.E., 2011. A Directed Protein Interaction Network for Investigating Intracellular Signal Transduction. *Science Signaling* 4, rs8–rs8.  
<https://doi.org/10.1126/scisignal.2001699>
- Vogel, D.Y.S., Glim, J.E., Stavenuiter, A.W.D., Breur, M., Heijnen, P., Amor, S., Dijkstra, C.D., Beelen, R.H.J., 2014. Human macrophage polarization in vitro: Maturation and activation methods compared. *Immunobiology* 219, 695–703.  
<https://doi.org/10.1016/j.imbio.2014.05.002>
- Volf, J., Havlickova, H., Hradecka, H., Ondrackova, P., Matiasovic, J., Faldyna, M., Rychlik, I., 2010. Epidemiology and interaction of *Salmonella enterica* serovar Derby, Infantis and Typhimurium with porcine alveolar macrophages. *Veterinary Microbiology* 146, 105–110. <https://doi.org/10.1016/j.vetmic.2010.04.031>
- Vosse, E. van de, Ali, S., Visser, A.W. de, Surjadi, C., Widjaja, S., Vollaard, A.M., Dissel, J.T. van, 2005. Susceptibility to typhoid fever is associated with a polymorphism in the cystic fibrosis transmembrane conductance regulator (CFTR). *Hum Genet* 118, 138–140. <https://doi.org/10.1007/s00439-005-0005-0>
- Wan, L., Lin, H.-J., Huang, C.-C., Chen, Y.-C., Hsu, Y.-A., Lin, C.-H., Lin, H.-C., Chang, C.-Y., Huang, S.-H., Lin, J.-M., Liu, F.-T., 2016. Galectin-12 enhances inflammation by promoting M1 polarization of macrophages and reduces insulin sensitivity in adipocytes. *Glycobiology* 26, 732–744. <https://doi.org/10.1093/glycob/cww013>
- Wan, P., Zhang, Q., Liu, W., Jia, Y., Ai, S., Wang, T., Wang, W., Pan, P., Yang, G., Xiang, Q., Huang, S., Yang, Q., Zhang, Wei, Liu, F., Tan, Q., Zhang, Wen, Wu, K., Liu, Y., Wu, J., 2019. Cullin1 binds and promotes NLRP3 ubiquitination to repress systematic inflammasome activation. *The FASEB Journal* 33, 5793–5807.  
<https://doi.org/10.1096/fj.201801681R>
- Wandel, M.P., Pathe, C., Werner, E.I., Ellison, C.J., Boyle, K.B., von der Malsburg, A., Rohde, J., Randow, F., 2017. GBPs Inhibit Motility of *Shigella flexneri* but Are Targeted for Degradation by the Bacterial Ubiquitin Ligase IpaH9.8. *Cell Host & Microbe* 22, 507–518.e5. <https://doi.org/10.1016/j.chom.2017.09.007>
- Wang, C., Bauckman, K.A., Ross, A.S.B., Symington, J.W., Ligon, M.M., Scholtes, G., Kumar, A., Chang, H.-W., Twentyman, J., Fashemi, B.E., Xavier, R.J., Mysorekar, I.U., 2018. A non-canonical autophagy-dependent role of the ATG16L1T300A variant in urothelial vesicular trafficking and uropathogenic *Escherichia coli* persistence. *Autophagy* 15, 527–542. <https://doi.org/10.1080/15548627.2018.1535290>
- Wang, J., Yazdani, S., Han, A., Schapira, M., 2020. The Druggable Genome as Seen from the Protein Data Bank. *bioRxiv* 2020.02.06.937730.  
<https://doi.org/10.1101/2020.02.06.937730>
- Wang, L., Zhao, X., Xia, X., Zhu, C., Zhang, H., Qin, W., Xu, Y., Hang, B., Sun, Y., Chen, S., Jiang, J., Zhang, G., Hu, J., 2019. Inhibitory Effects of Antimicrobial Peptide JH-3 on *Salmonella enterica* Serovar Typhimurium Strain CVCC541 Infection-Induced

- Inflammatory Cytokine Release and Apoptosis in RAW264.7 Cells. *Molecules* 24, 596. <https://doi.org/10.3390/molecules24030596>
- Wang, Ting, Moreno-Vinasco, L., Ma, S.-F., Zhou, T., Shimizu, Y., Sammani, S., Epshtein, Y., Watterson, D.M., Dudek, S.M., Garcia, J.G.N., 2014. Nonmuscle Myosin Light Chain Kinase Regulates Murine Asthmatic Inflammation. *Am J Respir Cell Mol Biol* 50, 1129–1135. <https://doi.org/10.1165/rcmb.2013-0434OC>
- Wang, T., Wei, J.J., Sabatini, D.M., Lander, E.S., 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* 343, 80–84. <https://doi.org/10.1126/science.1246981>
- Wang, X., Guo, J., Fei, E., Mu, Y., He, S., Che, X., Tan, J., Xia, K., Zhang, Z., Wang, G., Tang, B., 2014. BAG5 Protects against Mitochondrial Oxidative Damage through Regulating PINK1 Degradation. *PLOS ONE* 9, e86276. <https://doi.org/10.1371/journal.pone.0086276>
- Wang, X., Pesakhov, S., Weng, A., Kafka, M., Gocek, E., Nguyen, M., Harrison, J.S., Danilenko, M., Studzinski, G.P., 2014. ERK 5/MAPK pathway has a major role in  $1\alpha,25\text{-(OH)}_2$  vitamin D<sub>3</sub>-induced terminal differentiation of myeloid leukemia cells. *The Journal of Steroid Biochemistry and Molecular Biology, Proceedings of the 16th Vitamin D Workshop* 144, 223–227. <https://doi.org/10.1016/j.jsbmb.2013.10.002>
- Wang, Y., Shi, P., Chen, Q., Huang, Z., Zou, D., Zhang, J., Gao, X., Lin, Z., 2019. Mitochondrial ROS promote macrophage pyroptosis by inducing GSDMD oxidation. *J Mol Cell Biol* 11, 1069–1082. <https://doi.org/10.1093/jmcb/mjz020>
- Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D., Morris, Q., 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38, W214–W220. <https://doi.org/10.1093/nar/gkq537>
- Watanabe, T., Asano, N., Fichtner-Feigl, S., Gorelick, P.L., Tsuji, Y., Matsumoto, Y., Chiba, T., Fuss, I.J., Kitani, A., Strober, W., 2010. NOD1 contributes to mouse host defense against *Helicobacter pylori* via induction of type I IFN and activation of the ISGF3 signaling pathway. *J Clin Invest* 120, 1645–1662. <https://doi.org/10.1172/JCI39481>
- Watson, J.L., Sanchez-Garrido, J., Goddard, P.J., Torraca, V., Mostowy, S., Shenoy, A.R., Clements, A., 2019. Shigella sonnei O-Antigen Inhibits Internalization, Vacuole Escape, and Inflammasome Activation. *mBio* 10. <https://doi.org/10.1128/mBio.02654-19>
- Watters, K., Palmenberg, A.C., 2018. CDHR3 extracellular domains EC1-3 mediate rhinovirus C interaction with cells and as recombinant derivatives, are inhibitory to virus infection. *PLOS Pathogens* 14, e1007477. <https://doi.org/10.1371/journal.ppat.1007477>
- Wehmas, L.C., Wood, C.E., Chorley, B.N., Yauk, C.L., Nelson, G.M., Hester, S.D., 2019. Enhanced Quality Metrics for Assessing RNA Derived From Archival Formalin-Fixed Paraffin-Embedded Tissue Samples. *Toxicol Sci* 170, 357–373. <https://doi.org/10.1093/toxsci/kfz113>
- Wehner, A.B., Abdesslem, H., Dickendesher, T.L., Imai, F., Yoshida, Y., Giger, R.J., Pierchala, B.A., 2016. Semaphorin 3A is a retrograde cell death signal in developing sympathetic neurons. *Development* 143, 1560–1570. <https://doi.org/10.1242/dev.134627>
- Weigele, B.A., Orchard, R.C., Jimenez, A., Cox, G.W., Alto, N.M., 2017. A systematic exploration of the interactions between bacterial effector proteins and host cell

- membranes. *Nature Communications* 8, 532. <https://doi.org/10.1038/s41467-017-00700-7>
- Wen, H., Lei, Y., Eun, S.-Y., P. -Y. Ting, J., 2010. Plexin-A4–semaphorin 3A signaling is required for Toll-like receptor– and sepsis-induced cytokine storm. *J Exp Med* 207, 2943–2957. <https://doi.org/10.1084/jem.20101138>
- Wen, Q., Liu, J., Kang, R., Zhou, B., Tang, D., 2019. The release and activity of HMGB1 in ferroptosis. *Biochemical and Biophysical Research Communications* 510, 278–283. <https://doi.org/10.1016/j.bbrc.2019.01.090>
- Wessling-Resnick, M., 2015. Nramp1 and Other Transporters Involved in Metal Withholding during Infection. *J. Biol. Chem.* 290, 18984–18990. <https://doi.org/10.1074/jbc.R115.643973>
- West, A.P., Brodsky, I.E., Rahner, C., Woo, D.K., Erdjument-Bromage, H., Tempst, P., Walsh, M.C., Choi, Y., Shadel, G.S., Ghosh, S., 2011. TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* 472, 476–480. <https://doi.org/10.1038/nature09973>
- Westermann, A.J., Vogel, J., 2018. Host-Pathogen Transcriptomics by Dual RNA-Seq, in: Arluison, V., Valverde, C. (Eds.), *Bacterial Regulatory RNA: Methods and Protocols, Methods in Molecular Biology*. Springer, New York, NY, pp. 59–75. [https://doi.org/10.1007/978-1-4939-7634-8\\_4](https://doi.org/10.1007/978-1-4939-7634-8_4)
- WHO, 2019. WHO model list of essential medicines - 21st list, 2019.
- WHO, 2014. Antimicrobial resistance: global report on surveillance 2014. WHO.
- WHO, 2004. Public health response to biological and chemical weapons: WHO guidance (2004).
- Williams, C.R., Baccarella, A., Parrish, J.Z., Kim, C.C., 2016. Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics* 17, 103. <https://doi.org/10.1186/s12859-016-0956-2>
- Willyard, C., 2018. New human gene tally reignites debate. *Nature* 558, 354–355. <https://doi.org/10.1038/d41586-018-05462-w>
- Willyard, C., 2017. The drug-resistant bacteria that pose the greatest health threats. *Nature* 543, 15–15. <https://doi.org/10.1038/nature.2017.21550>
- Wilsmann-Theis, D., Koch, S., Mindnich, C., Bonness, S., Schnautz, S., Bubnoff, D. von, Bieber, T., 2013. Generation and functional analysis of human TNF- $\alpha$ /iNOS-producing dendritic cells (Tip-DC). *Allergy* 68, 890–898. <https://doi.org/10.1111/all.12172>
- Wingett, S., Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and... | F1000Research [WWW Document]. URL <https://f1000research.com/articles/7-1338/v2> (accessed 2.3.20).
- Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsois, R.M., Roth, J.R., Bäumlner, A.J., 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467, 426–429. <https://doi.org/10.1038/nature09415>
- Wooldridge, A.A., MacDonald, J.A., Erdodi, F., Ma, C., Borman, M.A., Hartshorne, D.J., Haystead, T.A.J., 2004. Smooth muscle phosphatase is regulated in vivo by exclusion of phosphorylation of threonine 696 of MYPT1 by phosphorylation of Serine 695 in response to cyclic nucleotides. *J. Biol. Chem.* 279, 34496–34504. <https://doi.org/10.1074/jbc.M405957200>
- Wu, J., Li, J., Salcedo, R., Mivechi, N.F., Trinchieri, G., Horuzsko, A., 2012. The Proinflammatory Myeloid Cell Receptor TREM-1 Controls Kupffer Cell Activation

- and Development of Hepatocellular Carcinoma. *Cancer Res* 72, 3977–3986. <https://doi.org/10.1158/0008-5472.CAN-12-0938>
- Wu, J., Pugh, R., Laughlin, R.C., Andrews-Polymenis, H., McClelland, M., Bäumlner, A.J., Adams, L.G., 2014. High-throughput Assay to Phenotype *Salmonella enterica* Typhimurium Association, Invasion, and Replication in Macrophages. *J Vis Exp*. <https://doi.org/10.3791/51759>
- Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., Chen, Z.J., 2013. Cyclic GMP-AMP Is an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science* 339, 826–830. <https://doi.org/10.1126/science.1229963>
- Wu, L.-G., Hamid, E., Shin, W., Chiang, H.-C., 2014. Exocytosis and Endocytosis: Modes, Functions, and Coupling Mechanisms. *Annu Rev Physiol* 76, 301–331. <https://doi.org/10.1146/annurev-physiol-021113-170305>
- Wu, S., Ye, Z., Liu, X., Zhao, Y., Xia, Y., Steiner, A., Petrof, E.O., Claud, E.C., Sun, J., 2010. *Salmonella typhimurium* infection increases p53 acetylation in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol* 298, G784–G794. <https://doi.org/10.1152/ajpgi.00526.2009>
- Wu, S.-C., Chu, X.-L., Su, J.-Q., Cui, Z.-Q., Zhang, L.-Y., Yu, Z.-J., Wu, Z.-M., Cai, M.-L., Li, H.-X., Zhang, Z.-J., 2018. Baicalin protects mice against *Salmonella typhimurium* infection via the modulation of both bacterial virulence and host response. *Phytomedicine* 48, 21–31. <https://doi.org/10.1016/j.phymed.2018.04.063>
- Wullaert, A., Verstrepren, L., Huffel, S.V., Adib-Conquy, M., Cornelis, S., Kreike, M., Haegman, M., Bakkouri, K.E., Sanders, M., Verhelst, K., Carpentier, I., Cavaillon, J.-M., Heynink, K., Beyaert, R., 2007. LIND/ABIN-3 Is a Novel Lipopolysaccharide-inducible Inhibitor of NF- $\kappa$ B Activation. *J. Biol. Chem.* 282, 81–90. <https://doi.org/10.1074/jbc.M607481200>
- Xia, L., Nordman, T., Olsson, J.M., Damdimopoulos, A., Björkhem-Bergman, L., Nalvarte, I., Eriksson, L.C., Arnér, E.S.J., Spyrou, G., Björnstedt, M., 2003. The Mammalian Cytosolic Selenoenzyme Thioredoxin Reductase Reduces Ubiquinone A NOVEL MECHANISM FOR DEFENSE AGAINST OXIDATIVE STRESS. *J. Biol. Chem.* 278, 2141–2146. <https://doi.org/10.1074/jbc.M210456200>
- Xiao, W., Wang, R.-S., Handy, D.E., Loscalzo, J., 2018. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxid Redox Signal* 28, 251–272. <https://doi.org/10.1089/ars.2017.7216>
- Xu, J., Luo, F., Zhang, Z., Xue, L., Wu, X.-S., Chiang, H.-C., Shin, W., Wu, L.-G., 2013. SNARE Proteins Synaptobrevin, SNAP-25, and Syntaxin Are Involved in Rapid and Slow Endocytosis at Synapses. *Cell Reports* 3, 1414–1421. <https://doi.org/10.1016/j.celrep.2013.03.010>
- Yachie, A., Niida, Y., Wada, T., Igarashi, N., Kaneda, H., Toma, T., Ohta, K., Kasahara, Y., Koizumi, S., 1999. Oxidative stress causes enhanced endothelial cell injury in human heme oxygenase-1 deficiency. *J Clin Invest* 103, 129–135.
- Yang, C.-S., Shin, D.-M., Kim, K.-H., Lee, Z.-W., Lee, C.-H., Park, S.G., Bae, Y.S., Jo, E.-K., 2009. NADPH Oxidase 2 Interaction with TLR2 Is Required for Efficient Innate Immune Responses to Mycobacteria via Cathelicidin Expression. *The Journal of Immunology* 182, 3696–3705. <https://doi.org/10.4049/jimmunol.0802217>
- Yang, J., Zhang, E., Liu, F., Zhang, Y., Zhong, M., Li, Y., Zhou, D., Chen, Y., Cao, Y., Xiao, Y., He, B., Yang, Y., Sun, Y., Lu, M., Yan, H., 2014. Flagellins of *Salmonella Typhi* and Nonpathogenic *Escherichia coli* Are Differentially Recognized through the NLRC4 Pathway in Macrophages. *JIN* 6, 47–57. <https://doi.org/10.1159/000351476>

- Yang, J., Zhao, Y., Shi, J., Shao, F., 2013. Human NAIP and mouse NAIP1 recognize bacterial type III secretion needle protein for inflammasome activation. *PNAS* 110, 14408–14413. <https://doi.org/10.1073/pnas.1306376110>
- Yang, T., Wang, R., Zhang, Jianzhong, Bao, C., Zhang, Juling, Li, R., Chen, X., Wu, S., Wen, J., Wei, S., Li, H., Cai, H., Yang, X., Zhao, Y., 2020. Mechanism of berberine in treating *Helicobacter pylori* induced chronic atrophic gastritis through IRF8-IFN- $\gamma$  signaling axis suppressing. *Life Sciences* 248, 117456. <https://doi.org/10.1016/j.lfs.2020.117456>
- Yeung, A.T.Y., Choi, Y.H., Lee, A.H.Y., Hale, C., Ponstingl, H., Pickard, D., Goulding, D., Thomas, M., Gill, E., Kim, J.K., Bradley, A., Hancock, R.E.W., Dougan, G., 2019. A Genome-Wide Knockout Screen in Human Macrophages Identified Host Factors Modulating *Salmonella* Infection. *mBio* 10. <https://doi.org/10.1128/mBio.02169-19>
- Yong, X., Zhao, L., Deng, W., Sun, H., Zhou, X., Mao, L., Hu, W., Shen, X., Sun, Q., Billadeau, D.D., Xue, Y., Jia, D., 2020. Mechanism of cargo recognition by retromer-linked SNX-BAR proteins. *PLOS Biology* 18, e3000631. <https://doi.org/10.1371/journal.pbio.3000631>
- Yoshida, K., 2003. Fibroblast cell shape and adhesion in vitro is altered by overexpression of the 7a and 7b isoforms of protocadherin 7, but not the 7c isoform. *Cell. Mol. Biol. Lett.* 8, 735–741.
- Yoshida, K., Watanabe, M., Kato, H., Dutta, A., Sugano, S., 1999. BH-protocadherin-c, a member of the cadherin superfamily, interacts with protein phosphatase 1 alpha through its intracellular domain. *FEBS Letters* 460, 93–98. [https://doi.org/10.1016/S0014-5793\(99\)01309-5](https://doi.org/10.1016/S0014-5793(99)01309-5)
- You, B.R., Park, W.H., 2016. Auranofin induces mesothelioma cell death through oxidative stress and GSH depletion. *Oncology Reports* 35, 546–551. <https://doi.org/10.3892/or.2015.4382>
- Yu, A., Zhou, R., Xia, B., Dang, W., Yang, Z., Chen, X., 2020. NAMPT maintains mitochondria content via NRF2-PPAR $\alpha$ /AMPK $\alpha$  pathway to promote cell survival under oxidative stress. *Cellular Signalling* 66, 109496. <https://doi.org/10.1016/j.cellsig.2019.109496>
- Yu, J., Rossi, R., Hale, C., Goulding, D., Dougan, G., 2009. Interaction of Enteric Bacterial Pathogens with Murine Embryonic Stem Cells. *Infect. Immun.* 77, 585–597. <https://doi.org/10.1128/IAI.01003-08>
- Yue, H.H., Leng, N., Wu, Z.B., Li, H.M., Li, X.Y., Zhu, P., 2009. Expression of CD147 on phorbol-12-myristate-13-acetate (PMA)-treated U937 cells differentiating into foam cells. *Archives of Biochemistry and Biophysics* 485, 30–34. <https://doi.org/10.1016/j.abb.2009.01.023>
- Yun, C., Wang, Y., Mukhopadhyay, D., Backlund, P., Kolli, N., Yergey, A., Wilkinson, K.D., Dasso, M., 2008. Nucleolar protein B23/nucleophosmin regulates the vertebrate SUMO pathway through SENP3 and SENP5 proteases. *J Cell Biol* 183, 589–595. <https://doi.org/10.1083/jcb.200807185>
- Zhan, S., Wang, T., Ge, W., Li, J., 2018. Multiple roles of Ring 1 and YY1 binding protein in physiology and disease. *J Cell Mol Med* 22, 2046–2054. <https://doi.org/10.1111/jcmm.13503>
- Zhang, J., Wang, X., Xu, L., Zhang, Z., Wang, F., Tang, X., 2020. Investigation of Potential Genetic Biomarkers and Molecular Mechanism of Ulcerative Colitis Utilizing Bioinformatics Analysis. *Biomed Res Int* 2020. <https://doi.org/10.1155/2020/4921387>
- Zhang, K., Riba, A., Nietschke, M., Torow, N., Repnik, U., Pütz, A., Fulde, M., Dupont, A., Hensel, M., Hornef, M., 2018. Minimal SPI1-T3SS effector requirement for

- Salmonella enterocyte invasion and intracellular proliferation in vivo. *PLOS Pathogens* 14, e1006925. <https://doi.org/10.1371/journal.ppat.1006925>
- Zhang, L., Button, B., Gabriel, S.E., Burkett, S., Yan, Y., Skiadopoulos, M.H., Dang, Y.L., Vogel, L.N., McKay, T., Mengos, A., Boucher, R.C., Collins, P.L., Pickles, R.J., 2009. CFTR Delivery to 25% of Surface Epithelial Cells Restores Normal Rates of Mucus Transport to Human Cystic Fibrosis Airway Epithelium. *PLOS Biology* 7, e1000155. <https://doi.org/10.1371/journal.pbio.1000155>
- Zhang, Q., Chao, T., Patil, V.S., Qin, Y., Tiwari, S.K., Chiou, J., Dobin, A., Tsai, C., Li, Z., Dang, J., Gupta, S., Urdahl, K., Nizet, V., Gingeras, T.R., Gaulton, K.J., Rana, T.M., 2019. The long noncoding RNA ROCK1 regulates inflammatory gene expression. *EMBO J* 38. <https://doi.org/10.15252/embj.2018100041>
- Zhang, X.M., Wang, T., Hu, P., Li, B., Liu, H., Cheng, Y.-F., 2019. SERPINB2 overexpression inhibited cell proliferation, invasion and migration, led to G2/M arrest, and increased radiosensitivity in nasopharyngeal carcinoma cells. *J Radiat Res* 60, 318–327. <https://doi.org/10.1093/jrr/rrz003>
- Zhang, Xu, Shi, H., Wu, J., Zhang, Xuewu, Sun, L., Chen, C., Chen, Z.J., 2013. Cyclic GMP-AMP Containing Mixed Phosphodiester Linkages Is An Endogenous High-Affinity Ligand for STING. *Molecular Cell* 51, 226–235. <https://doi.org/10.1016/j.molcel.2013.05.022>
- Zhang, Y., Mühlen, S., Oates, C.V., Pearson, J.S., Hartland, E.L., 2016. Identification of a Distinct Substrate-binding Domain in the Bacterial Cysteine Methyltransferase Effectors NleE and OspZ \*. *Journal of Biological Chemistry* 291, 20149–20162. <https://doi.org/10.1074/jbc.M116.734079>
- Zhao, A., Yang, Z., Sun, R., Grinchuk, V., Netzel-Arnett, S., Anglin, I.E., Driesbaugh, K.H., Notari, L., Bohl, J.A., Madden, K.B., Urban, J.F., Antalis, T.M., Shea-Donohue, T., 2013. SerpinB2 Is Critical to Th2 Immunity against Enteric Nematode Infection. *The Journal of Immunology* 190, 5779–5787. <https://doi.org/10.4049/jimmunol.1200293>
- Zhao, X., Zhuang, S., Chen, Y., Boss, G.R., Pilz, R.B., 2005. Cyclic GMP-dependent Protein Kinase Regulates CCAAT Enhancer-binding Protein  $\beta$  Functions through Inhibition of Glycogen Synthase Kinase-3. *J. Biol. Chem.* 280, 32683–32692. <https://doi.org/10.1074/jbc.M505486200>
- Zhao, Y., Yang, J., Shi, J., Gong, Y.-N., Lu, Q., Xu, H., Liu, L., Shao, F., 2011. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* 477, 596–600. <https://doi.org/10.1038/nature10510>
- Zhou, J., Li, C., Yu, A., Jie, S., Du, X., Liu, T., Wang, W., Luo, Y., 2019. Bioinformatics analysis of differentially expressed genes involved in human developmental chondrogenesis. *Medicine (Baltimore)* 98. <https://doi.org/10.1097/MD.0000000000016240>
- Zhou, R., Yazdi, A.S., Menu, P., Tschopp, J., 2011. A role for mitochondria in NLRP3 inflammasome activation. *Nature* 469, 221–225. <https://doi.org/10.1038/nature09663>
- Zhou, W., Wei, W., Sun, Y., 2013. Genetically engineered mouse models for functional studies of SKP1-CUL1-F-box-protein (SCF) E3 ubiquitin ligases. *Cell Research* 23, 599–619. <https://doi.org/10.1038/cr.2013.44>
- Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., Wei, W., 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491. <https://doi.org/10.1038/nature13166>

- Zhu, B., Das, S., Mitra, S., Farris, T.R., McBride, J.W., 2017. Ehrlichia chaffeensis TRP120 Moonlights as a HECT E3 Ligase Involved in Self- and Host Ubiquitination To Influence Protein Interactions and Stability for Intracellular Survival. *Infection and Immunity* 85. <https://doi.org/10.1128/IAI.00290-17>
- Zilio, N., Eifler-Olivi, K., Ulrich, H.D., 2017. Functions of SUMO in the Maintenance of Genome Stability. *Adv. Exp. Med. Biol.* 963, 51–87. [https://doi.org/10.1007/978-3-319-50044-7\\_4](https://doi.org/10.1007/978-3-319-50044-7_4)
- Zink, D., Amaral, M.D., Englmann, A., Lang, S., Clarke, L.A., Rudolph, C., Alt, F., Luther, K., Braz, C., Sadoni, N., Rosenecker, J., Schindelhauer, D., 2004. Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol* 166, 815–825. <https://doi.org/10.1083/jcb.200404107>
- Zunino, R., Braschi, E., Xu, L., McBride, H.M., 2009. Translocation of SenP5 from the Nucleoli to the Mitochondria Modulates DRP1-dependent Fission during Mitosis. *J. Biol. Chem.* 284, 17783–17795. <https://doi.org/10.1074/jbc.M901902200>

# Appendix A INVESTIGATING THE SUITABILITY OF DIFFERENTIATION STRATEGIES OF U937 CELLS FOR SCREENING

This appendix describes in more detail work summarised in Chapter 4.

## A.1. Differentiation Strategy #1

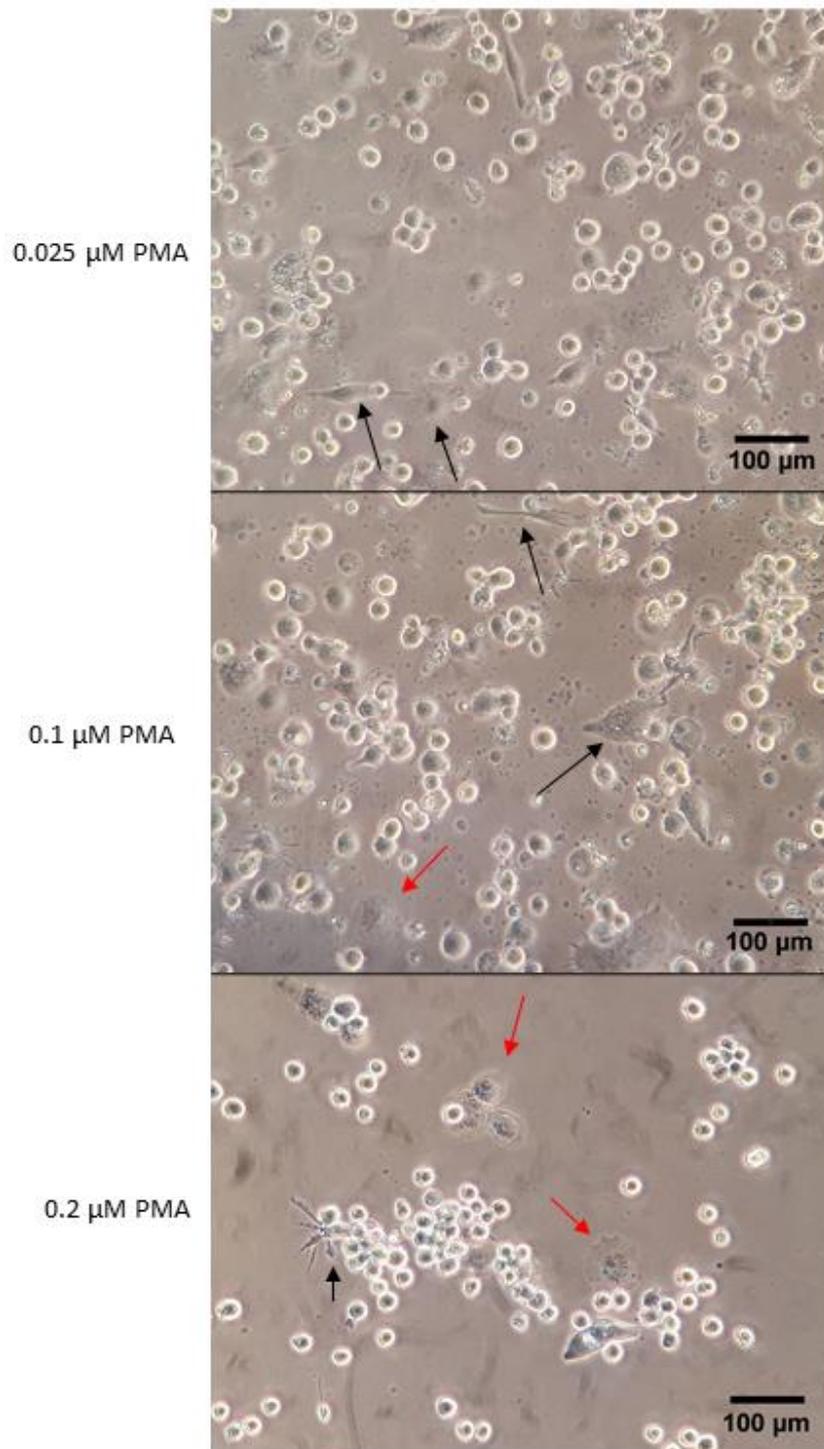
A decision was made to test different concentrations of one of the most used drugs, PMA. Characteristics such as adherence and loss of proliferative ability are indicators of differentiation (Chanput et al., 2015; Cooper, 2000; Starr et al., 2018), although a variety of cell morphologies and phenotypes have been observed in the literature. Indeed, differences in culturing prior to differentiation (Aldo et al., 2013), differences in PMA concentration, treatment time, and resting period can all change the apparent phenotype (Maeß et al., 2014a; Starr et al., 2018).

U937 cells were treated with 0.025, 0.1 (mostly commonly used e.g. (Boukes and van de Venter, 2012; Gillies et al., 2012; Tusiimire et al., 2016)) and 0.2  $\mu\text{M}$  (highest reliable concentration seen in the literature (Daigneault et al., 2010; Yue et al., 2009)) of PMA with varying concentrations of FCS (10%, 1% and 0%) on Day 0 (Figure A.1). Lower concentrations of FCS were tested because differentiation of other types of cell line have benefited from serum starvation (Seidman et al., 1996). The use of collagen as a surface for attachment was also tested with PMA and 10% FCS as this has been shown to improve attachment with other types of cell line (Vasse et al., 2018).



**Table A.1 PMA treatment promotes adherence and changes in morphology in U937 cells.**  
*Differentiation strategy #1: U937 cells were treated with PMA at the concentrations noted and incubated for 3 days at 37°C 5% CO<sub>2</sub>. Treated cells were then washed to remove any cells and drug remaining in suspension (10% FCS) or not (1% and 0% FCS) before observing under the microscope. Percentages given are rough estimates based on visual inspection of remaining cells after washing (if performed). The results of one experiment are described.*

Concentration of PMA ( $\mu\text{M}$ )	Starting fetal calf serum (FCS) concentration			
	10% FCS		1% FCS	0% FCS
	No collagen	collagen	No collagen	No collagen
0	No attachment, circular	No attachment, circular, a few have one or two protrusions	No attachment, circular, do not appear healthy	No attachment, circular, do not appear healthy
0.025	>90% are attached, normal size and shape, in clusters. <10% cells are much larger, thin and long; others were circular, with thin filopodia-like protrusions, Remaining are normal in size, shape and not attached	Most are attached, normal size and shape, in clusters. No larger ones seen.	Not tested	Not tested
0.1	Similar to 0.025 $\mu\text{M}$ , more variety of shapes seen in larger adhered cells	Not tested	Some clustering, normal size and shape, though a few have one or two filopodia. A few are attached, some variation in morphology	Very little clustering, normal size and shape. Not clear if any are attached.
0.2	Greater number of large adhered cells with altered morphology and more variety of morphology seen (compared to 0.025 $\mu\text{M}$ , similar to 0.1 $\mu\text{M}$ ).	Not tested	More frequent clustering (than 0.1 $\mu\text{M}$ ), most are normal in size and shape. A few are attached, with some variation in morphology	Very little clustering of 3-5 cells seen, a small number are attached, do not look very healthy



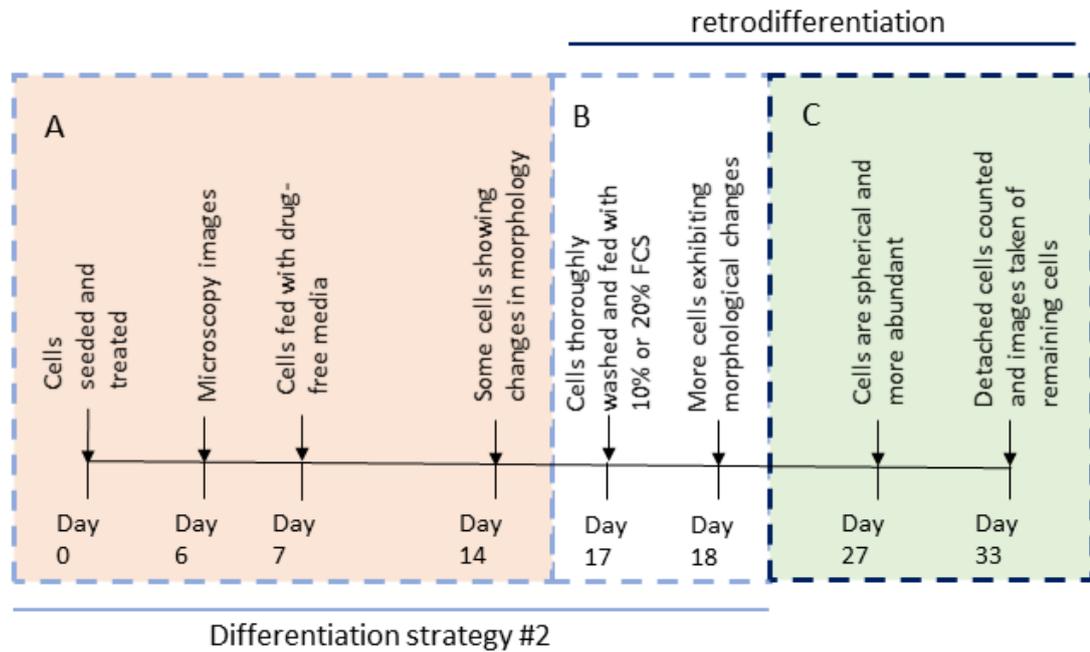
**Figure A.2 PMA treatment of U937 cells promotes a variety of morphological changes, creating a heterogeneous population of attached cells. Differentiation strategy #1:** Micrographs of U937 cells treated with PMA in RPMI and 10% FCS for 3 days at the stated concentrations, gently washed and rested for 3 days in medium without PMA and a lower concentration of FCS (1%). All the cells in the plane of view are attached to the plate, with many in clusters. Arrows point to different shapes of attached cells often with filopodia (black arrows), especially noting the faint cells (red arrows, reminiscent of classically activated macrophages) at 0.2  $\mu\text{M}$ . All images are from one experiment.

## A.2. Differentiation Strategy #2

Although morphological changes are not definitive of the transition to a 'macrophage-like' state, it does give an indication of what proportion of cells might be expected to demonstrate other 'macrophage-like' characteristics (McWhorter et al., 2013). Even at 0.2  $\mu\text{M}$  PMA, the number of cells with the desired macrophage-like morphology was estimated at less than 1% of the treated population, although all cells were adhered indicating some degree differentiation. To identify methods of improving the proportion of cells with macrophage-like morphology in the treated population, additional differentiation drugs at multiple concentrations were tested, as well as a higher concentration of PMA. Lower FCS concentrations in the previous strategy reduced the capacity for differentiation, so all future treatments were performed at 10% FCS. U937 cells were seeded on Day 0 at  $5 \times 10^5$ /well in a 24 well plate and treated with 0.1, 0.2, 0.25  $\mu\text{M}$  PMA, 0.1, 0.4, 0.8  $\mu\text{M}$  VD3 or 0.1, 0.5, ATRA. This time, cells were given 6 days in differentiation media to further promote changes (Figure A.3 for timeline) before micrographs were taken (

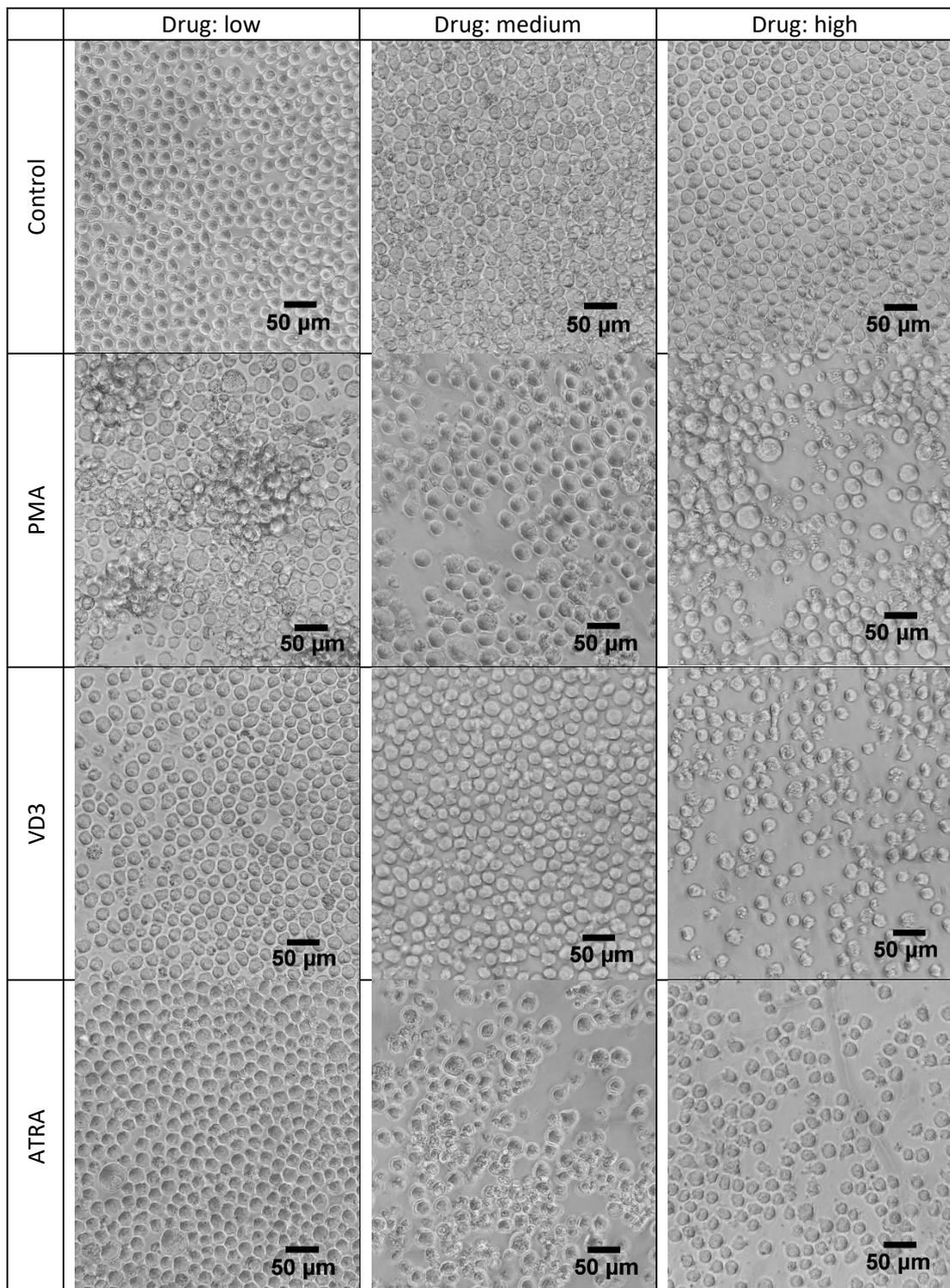
Figure A.4). At 0.1, 0.2 and 0.25  $\mu\text{M}$  of PMA, the U937 cells had attached to the plate and formed clusters (example in

Figure A.4, PMA low dose). At 0.1  $\mu\text{M}$  PMA cell proliferation was inhibited and some cell death was seen at 0.2  $\mu\text{M}$ , this increased at 0.25  $\mu\text{M}$  (notable due to the cell debris). ATRA also partially inhibited cell proliferation at 0.5 and 1  $\mu\text{M}$ ; although it was not clear whether the cells had become adherent or not. VD3 produced no obvious visible changes.



**Figure A.3 Timeline for differentiation strategy #2- comparing effects of VD3, PMA and ATRA; secondly for retrodifferentiation of PMA treated cells.** *The timeline for differentiation strategy #2 encompassed A and B, key features of the retrodifferentiation procedure can be found in B and C. A, Day 0, U937 cells were seeded at  $5 \times 10^5$ /well in a 24 well plate and treated with 0.1, 0.2, 0.25  $\mu\text{M}$  PMA, 0.1, 0.4, 0.8  $\mu\text{M}$  1,25-dihydroxyvitamin D3 (VD3) or 0.1, 0.5, 1  $\mu\text{M}$  All-trans-retinoic acid (ATRA) at 10% FCS. Day 6, microscopy images were taken (*

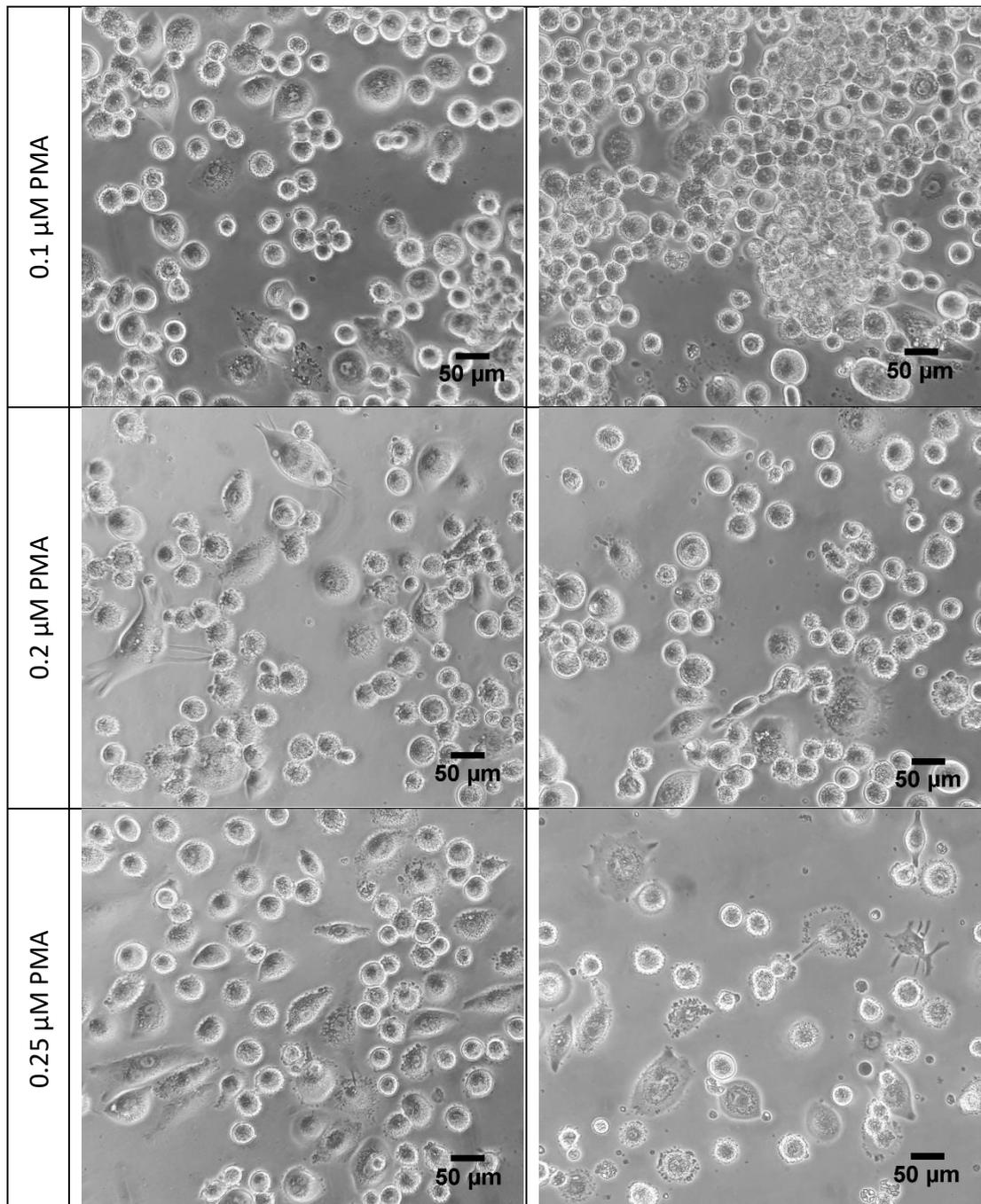
*Figure A.4). Day 7, drug-containing media was removed and replaced with drug-free media. Day 14, some cells began to show changes in morphology. B, Day 17, PMA-treated cells were washed and media containing 10% or 20% FCS was added. Day 18, more cells displayed changes in morphology in PMA-treated wells (Figure A.5). C, Day 27, all PMA-treated cells were circular and more abundant. Day 33, PMA-treated wells were washed to separate detached cells and these were counted (Table 4.1), images were taken of the remaining adherent cells (Figure 4.3). Experiment performed once.*



**Figure A.4 U937 cells continuously treated with PMA cluster together but do not change shape.** *Differentiation strategy #2: U937 cells were seeded at  $5 \times 10^5$ / well in a 24 well plate and treated continuously with 0.1 (low), 0.2 (medium), 0.25 (high)  $\mu\text{M}$  PMA, 0.1 (low), 0.4 (medium), 0.8 (high)  $\mu\text{M}$  1,25-dihydroxyvitamin D3 (VD3) or 0.1 (low), 0.5 (medium), 1 (high)  $\mu\text{M}$  All-trans-retinoic acid (ATRA) at 10% foetal calf serum. The control wells (top row) used for comparison were untreated (left) and vehicle controls at the highest concentration used in drug treatment (3% ethanol (middle) and 0.01% DMSO (right)). On day 6 the cells were imaged with a bright field microscope. All images are from one experiment.*

As cells treated with VD3 or ATRA demonstrated little to no differentiation characteristics, only PMA treated cells were observed further. After taking microscopy images (

Figure A.4), the PMA differentiation media was replaced one day later (day 7) with drug-free media (a drug-free 'resting' stage of a few days has been recommended to complete differentiation) (Daigneault et al., 2010). After a total of 14 days, some cells were flatter with reduced light diffraction, while others had filopodia or were spindle shaped, just as seen at 0.2 and 0.25  $\mu\text{M}$  PMA in strategy #1. Wells were washed to clear off floating cells to improve future microscopy images. Cells exhibiting morphological changes became more abundant by day 18, upon which several micrographs were taken and two micrographs per condition are presented (Figure A.5). When cells were rested, 0.25  $\mu\text{M}$  PMA produced a higher proportion of cells with altered morphology (spindle shaped, flattened pancake shaped, filopodia) than at other concentrations, but also resulted in considerable cell death. In contrast, 0.2  $\mu\text{M}$  PMA treated and rested cells demonstrated lower levels of cell death - comparable to 0.1  $\mu\text{M}$ , but had more cells with altered morphology than at 0.1  $\mu\text{M}$ . Altered morphology was consistently seen after resting, but the proportion of cells demonstrating these features were still low; thus, the proportion of flattened pancake shaped cells was not yet sufficient for experimental use.



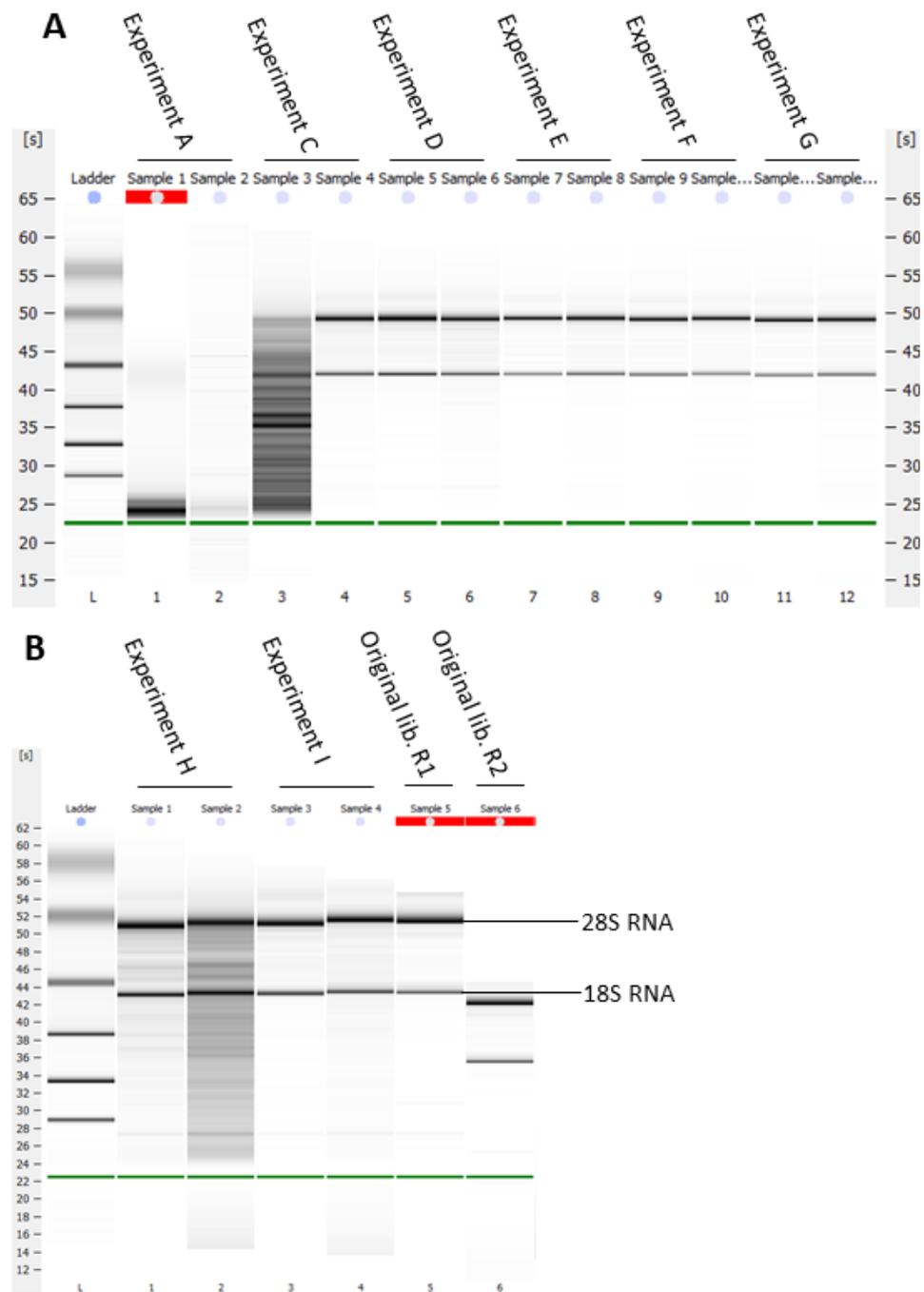
**Figure A.5 PMA-treated U937 cells undergo diverse morphological changes after a subsequent 'rest' period of culturing without PMA.** *Differentiation strategy #2: U937 cells were seeded at  $5 \times 10^5$ / well in a 24 well plate and treated with 0.1, 0.2, or 0.25  $\mu\text{M}$  PMA for 7 days before replacing with media without PMA. The cells were imaged on day 18 by bright field microscopy. For each condition two example micrographs are presented – taken from one experiment.*

## **Appendix B** QUALITY ASSESSMENT OF TOTAL RNA SAMPLES DERIVED FROM STM AND *F. TULARENSIS* U937 SCREENS

### **B.1.** RNA extracted from STM U937 screen samples: Quality assessment and processing for next generation sequencing (NGS)

The concentration and quality of the extracted RNA was measured on a 2100 Bioanalyzer (Figure B.1, Table B.1). The RNA from both samples in experiment A and the control sample in experiment C were degraded, as observed by the complete lack rRNA bands in the first two and lower running of the bands in the latter (indicating a shorter fragment of RNA). Therefore, these experiments were not analysed further. Experiment B is not represented in the following figures and tables due to a technical error resulting in no RNA. The ratio of 28S and 18S band intensities for experiments D, E, F and G were all close to 2 and the RNA integrity numbers (RINs) were all  $\geq 9$ ; both of which are indicators of good quality (Schroeder et al., 2006). The range of RIN numbers is 1 (degraded) to 10 (completely intact). To replace experiments A and C, 2 further experiments (H and I) were performed and the resulting RNA measured (Figure B.1B, Table B.1). Unfortunately, the RNA quality of H infected was not as good, with a noisy gel lane and a RIN of 5.3. A threshold RIN of 5.5 is suggested as meaningful for qRT-PCR data (Schroeder et al., 2006), though a standardised RIN cutoff has not been identified (Gallego Romero et al., 2014; Reiman et al., 2017). Nevertheless, this experiment pair was retained for further analysis.

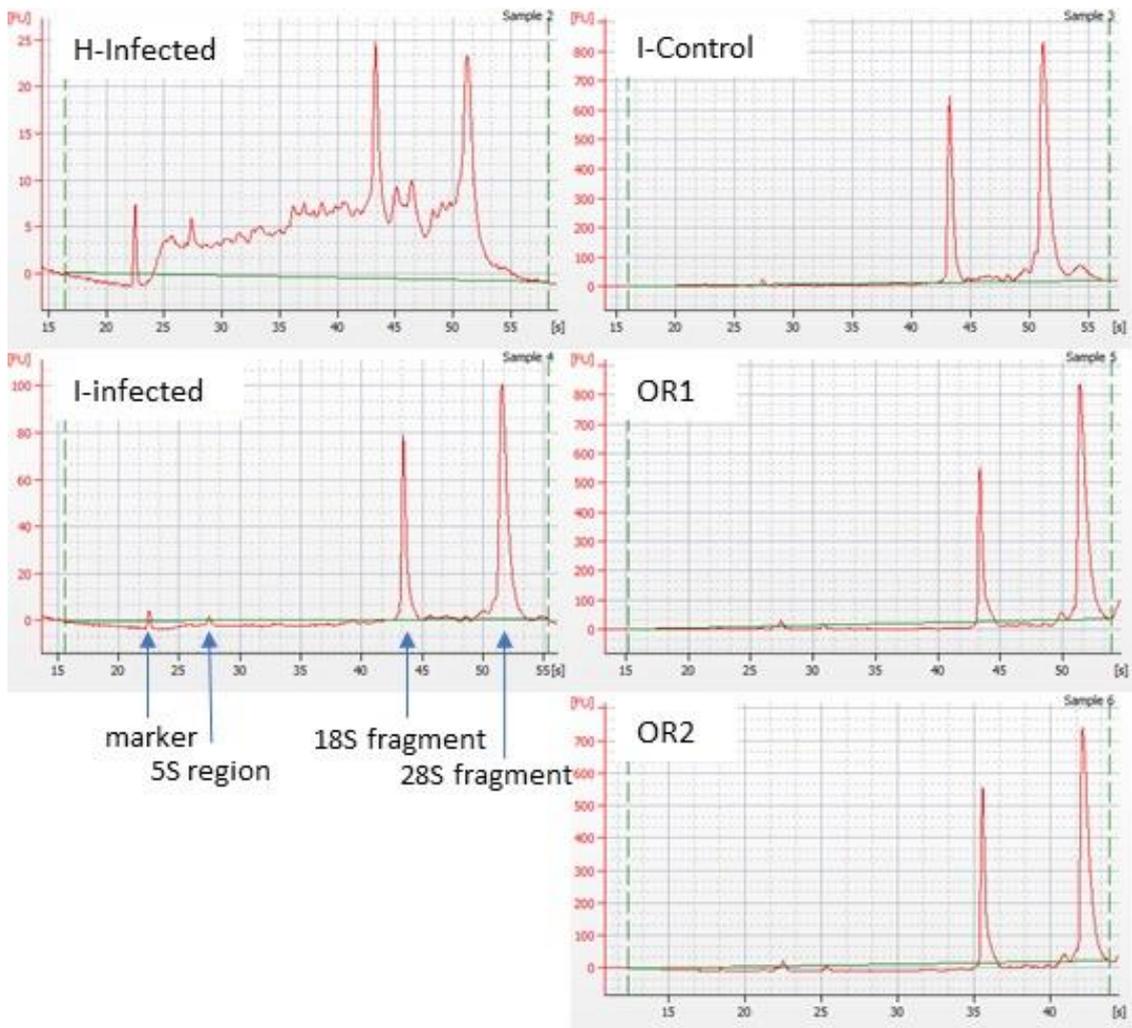
RINs were not calculated for Original Library (Original Lib.) replicates 1 (R1) and 2 (R2) as the marker for these wells was not called; however, the quality could still be estimated using the ratio of 28S and 18S rRNA intensities (Figure B.1B, Table B.1). The RNA quality of Original Lib. R1 was relatively good (28S and 18S rRNA intensity of 2.15) but the quality of Original Lib. R2 RNA appears to have suffered; the bands are clearly much lower than in other lanes (Figure B.1B). The bioanalyzer reports the 28S:18S rRNA intensity as 0.0 but perhaps this is due to the bands being in a different place. If the electropherograms (Figure B.2) are analysed directly, they show distinct 18S and 28S fragment peaks for Original Lib. R1 and Original Lib. R2 comparable to electropherograms for other samples with high RIN numbers (e.g. I-Control and I-Infected). The lack of reported RIN numbers and odd band placement are likely to due to a combination of low marker intensity and high RNA concentration.



**Figure B.1 Selection of infection-Control pair replicates that produced the highest quality RNA.** RNA extraction samples from experiments D, E, F, G, H and I were of good quality. The thick upper band is human 28S RNA and the lower band is human 18S RNA. Artificial nano RNA gel from Agilent Bioanalyzer. The red bars above the gel lanes indicates a warning given by the analysis programme, in (A) sample 1 the warning indicated that the concentration was too high to give an accurate RNA integrity number (RIN). In (B) samples 5 and 6, the marker was not called, preventing the programme from calculating the RIN.

**Table B.1 Quality of total RNA extracted from chosen samples is sufficient for RNA-Seq.** RNA concentration (ng/ $\mu$ L), ribosomal RNA (rRNA) ratio and RIN numbers calculated by a 2100 Bioanalyzer. Total RNA for each sample was diluted in nuclease free water to normalise the overall concentration of RNA to the sample with the lowest RNA concentration and thus normalise the depth of subsequent sequencing. This step was performed prior to cDNA synthesis. The volume of RNA used for normalisation per sample is listed in the far righthand column. The letter at the beginning of the sample name refers to the experiment, n=6 for Infected and control samples, n=2 for Original Library (OR) samples.

Figure B.1 position	Sample	concentration (ng/ $\mu$ L)	rRNA Ratio [28s / 18s]	RIN	Volume ( $\mu$ L, up to 11 $\mu$ L) to normalise to I infected
Gel A, 5	D control	853	1.88	10.0	2.45
Gel A, 6	D infected	610	1.80	9.7	3.43
Gel A, 7	E control	935	2.15	10.0	2.24
Gel A, 8	E infected	555	1.97	10.0	3.77
Gel A, 9	F control	1780	1.95	10.0	1.17
Gel A, 10	F infected	324	2.20	9.9	6.45
Gel A, 11	G control	1380	2.12	10.0	1.52
Gel A, 12	G infected	340.7	1.99	9.9	6.13
Gel B, 1	H control	3420	1.83	8.2	0.61
Gel B, 2	H infected	356	1.31	5.3	5.87
Gel B, 3	I control	1860	1.74	10.0	1.12
Gel B, 4	I infected	190	1.88	10.0	11.0
Gel B, 5	OR1	1330	2.15	N/A	1.57
Gel B, 6	OR2	1200	0.00	N/A	1.74



**Figure B.2** The RNA quality of OR1 and OR2 is good. The RNA electropherograms of H-Infected, I-Control, I-Infected, Original lib. R1 (OR1) and Original lib. R2 (OR2) from an Agilent 2100 Bioanalyzer. RIN values are calculated based on the ratio between the 18S and 28S fragment peaks, the background noise and the RNA marker. A sample containing background noise and a reduced 28S fragment peak (H-Infected) is presented as an example with lower quality (RIN of 5.3); as well as samples with high quality RIN values with different concentrations (I-Control, I-Infected, both RIN of 10).

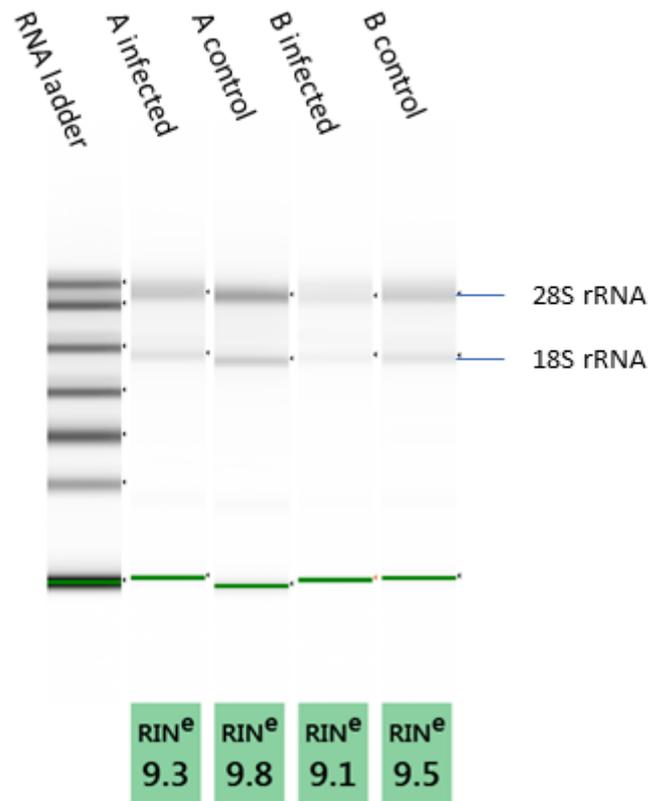
Table B.2 states the concentration and quality (as measured on a nanodrop) of the purified RACE PCR products from the STM U937 screen (Chapter 4) ready for shipment to Glasgow Polyomics for sequencing library construction and next generation sequencing.

**Table B.2 DNA concentration and quality of purified GT screen RACE PCR material prior to DNA library preparation and sequencing.** *DNA concentration and quality measured by nanodrop prior to sending samples to Glasgow Polyomics for DNA library preparation and sequencing. The letter at the beginning of the sample name refers to the experiment, n=6 for Infected and control samples, n=2 for Original Library (OR) samples.*

<b>Sample</b>	<b>Concentration (ng/ <math>\mu</math>L)</b>	<b>260/280 ratio</b>	<b>260/230 ratio</b>
D control	494	1.85	2.20
D infected	573	1.85	2.20
E control	525	1.85	2.19
E infected	493	1.86	2.23
F control	441	1.85	2.20
F infected	437	1.85	2.24
G control	461	1.86	2.24
G infected	453	1.86	2.20
H control	437	1.85	2.19
H infected	471	1.86	2.20
I control	484	1.86	2.23
I infected	455	1.86	2.19
Original Lib. R1	423	1.85	2.24
Original Lib. R2	467	1.86	2.23

## **B.2.** RNA extracted from *F. tularensis* LVS U937 screen samples: Quality assessment and processing for next generation sequencing (NGS)

The concentration and quality of the total RNA extracted from *F. tularensis* LVS U937 screen samples was measured on a TapeStation (Figure B.3). The ratio of 28S and 18S band intensities for experiments A, B, C and D were quite variable but the RINs were all  $\geq 9$ , indicating high quality. RIN numbers, where available, are widely held as more accurate than 28S :18S intensity ratios (Schroeder et al., 2006).



Sample	RIN <sup>e</sup>	28S/18S (Area)	Concentration [ng/μl]
A infected	9.3	1.6	418
A control	9.8	2.2	200
B infected	9.1	2	805
B control	9.5	2.4	407
C infected	9.3	1.7	21.6
C control	9.2	1.0	42.8
D infected	9.4	1.4	36
D control	9.4	1.3	55

**Figure B.3 U937- *F. tularensis* LVS GT library screen RNA samples are good quality.** RNA concentrations and quality values (RIN<sup>e</sup>, 28S/18S ratio, a ratio of 2 is good) obtained from an Agilent TapeStation system. The 28S rRNA band ideally should be twice as intense as the 18S rRNA band, which indicates no RNA degradation. All RIN numbers indicate near-perfect quality RNA (RIN of 10 is undigested). Letters in sample name indicate a specific experiment ( $n=3$  for infected and control samples,  $n=1$  for each passage control).

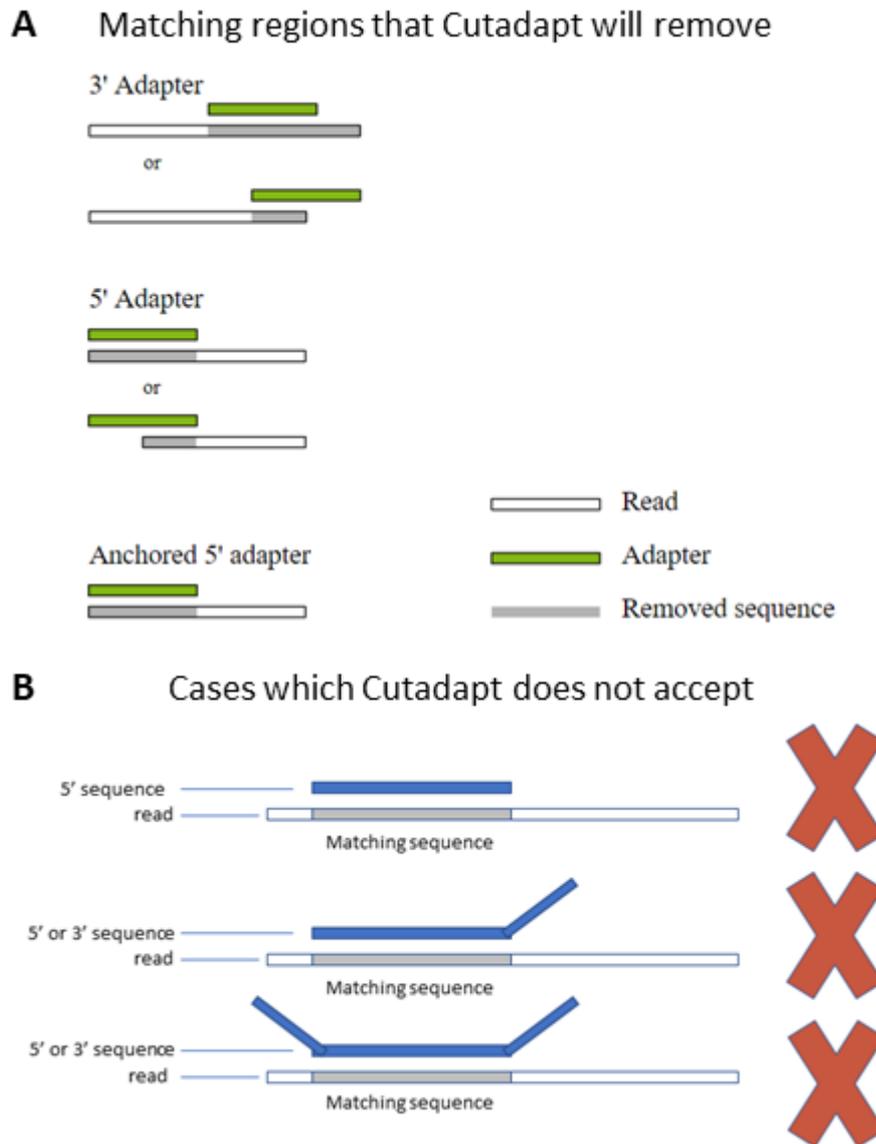
## Appendix C INVESTIGATION OF SUITABLE GT NGS PRE-ALIGNMENT PROCEDURES

The following section describes in more detail work summarised in Chapter 5, Section 5.2.2.

### C.1. Cutting

As previously identified in Section 5.2.1.4, GT cassette and cDNA synthesis primer sequences comprise a considerable proportion of the raw sequencing files. Indeed, based on the summary of the potential consequences of fragmentation in Figure 5.1, a substantial proportion of the reads were expected to contain GT cassette sequences, primers, or GT cassette - mRNA spanning reads or primer. Exogenous GT cassette - mRNA spanning reads with a significant GT cassette region are unlikely to align, therefore the information contained is lost. With this in mind, GT cassette sequences should ideally be removed from sequence reads before alignment; however, this should be weighed against the loss of potentially mappable sequences. A pertinent aspect of Cutadapt is that surviving single and paired reads are separated. After some investigation, it was found that adding orphaned reads back into the paired output (after Cutadapt or Trimmomatic) was not possible using Galaxy.

To begin, the reverse complement of the sequencing primers were put into Cutadapt Galaxy Version 1.16.6 (Martin, 2011) on [usegalaxy.org](http://usegalaxy.org) to cut them from the 3' end of the sequences in sample D-control (used as a representative sample for this optimisation stage). Additionally, the sequences for the cDNA synthesis primer and GT vector exon were provided for removal from the 5' end, as well as the respective reverse complements from the 3' end. If a sequence is provided to cut from the 3' end, Cutadapt can match it to the middle of a read (before the 3' end of the sequence), cutting off the matching sequence as well as any following sequence, match it to the end of the sequence, or match the edge of the provided sequence to the 3' edge of the read (Figure C.1A). If sequences are provided for the 5' end of the read, Cutadapt will only match it if the whole sequence matches the read or if the edge of the sequence matches the 5' edge of the read. Cutadapt will not match a 5' sequence to the middle of the read or part of a 5' or 3' sequence to the middle of a read (Figure C.1B).



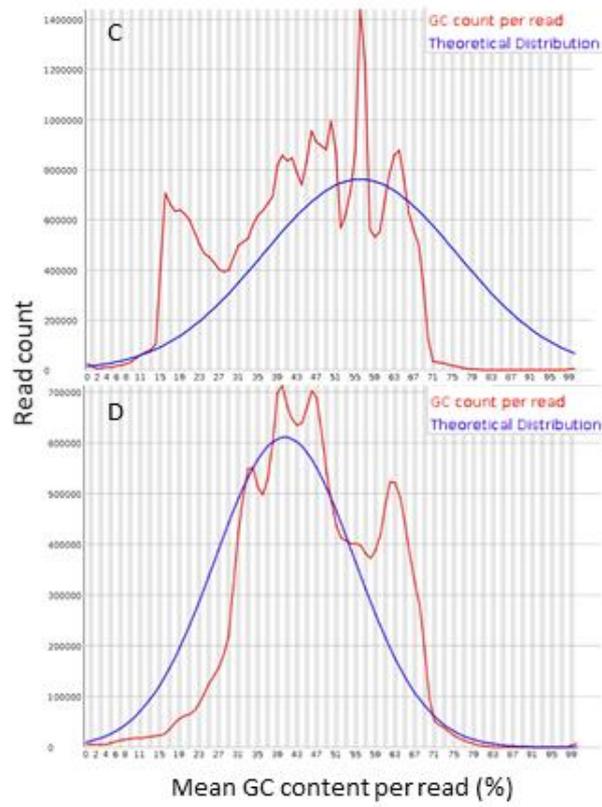
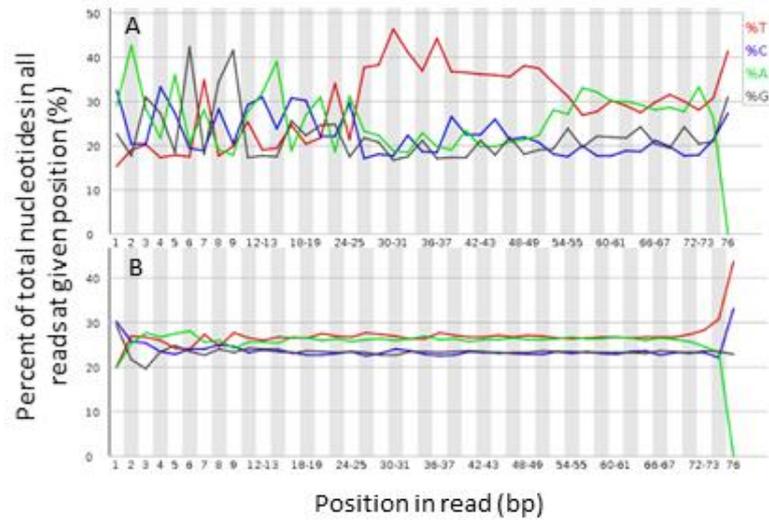
**Figure C.1** Cases in which Cutadapt will remove or not remove a matching region of a read. Cases for which a region that matches a supplied sequence (adapter) will be removed, **A**. Cases for which a region that matches a supplied sequence (adapter) will not be removed, **B**. **A** has been adapted from Galaxy Main <https://usegalaxy.org/Cutadapt> documentation summary.

FastQC reports were prepared for the surviving forward and reverse read output from Cutadapt to identify further potential GT cassette sequences to be removed in an iterative process. Most of the overrepresented sequences in this iteration were versions of the GT vector sequences and 3' RACE1 primer, a couple matched sequences in the human genome (using the Blat tool from the UCSC genome browser) (Kent, 2002) and some failed to match any sequence on NCBI Blast (McGinnis and Madden, 2004). The known GT vector and 3' RACE1 primer sequences identified in forward and reverse samples were cut from the ends of the respective sequences using Cutadapt again. Following this second round, a further FastQC report was prepared for the forward and reverse files and the overrepresented

sequences were fed into a third Cutadapt. The FastQC overrepresented sequences module only reported a list of short sequence strings following the third Cutadapt; therefore, the Cutadapt output was fed into Trimmomatic to trim the reads to a minimum average Phred of 20 and filtered by size as described in Figure 5.7B. FastQC reports for the forward and reverse files were prepared to check the quality as well as any remaining overrepresented sequences. Sequences newly identified as overrepresented matched GT cassette sequences, 3' RACE1 and cDNA synthesis primers in both forward and reverse files. Thus 3 more rounds of Cutadapt were performed iteratively (six in total), with FastQC reports after each feeding identified overrepresented sequences into subsequent rounds of Cutadapt. Further FastQC reports identified no known overrepresented sequences, so a final round of Trimmomatic was performed with the same settings (the surviving paired sequence files are known as (Cut x3-Trim) x2). The final FastQC reports performed on the forward and reverse files identified further overrepresented sequences - in all cases these matched the human genome (UCSC genome browser and NCBI nucleotide Blast) and are likely the desired target host defence/susceptibility genes. The sequence length distribution and per sequence quality scores (measured by FastQC) improved after each round of Trimmomatic; the vast majority of reads had a length of 75 bases, with a small number varying in length between 75 and 20 bases. The mean sequence Phred score after Trimmomatic was 20 or greater. The per base sequence content and GC content distributions also moved closer to the ideal distribution for RNA-Seq data following the iterative rounds of Cutadapt and Trimmomatic (Figure C.2). Comparison of the per base sequence content between the raw (Figure C.2A) and (Cut x3 Trim) x2 (Figure C.2B) forward files shows that the overlapping spiked peaks representing each base had evened out into straighter horizontal lines as expected for RNA-Seq data without vector contamination. The guanine line increases in percentage at the end of the sequence (73-76 bases); which could be due to the addition of multiple guanines added to the end of the read after a drop in signal (a known, inherent technical fault). The NextSeq 500 sequencer uses 2-colour sequencing chemistry and codes guanine as black (or the absence of signal). Contrary to this hypothesis, the symptomatic increase in Phred score after a drop in quality was not present in the reads prior to Trimmomatic quality trimming (data not shown); therefore, it is more likely to be caused by unwanted vector sequences. Furthermore, after the iterative cutting and trimming was performed on the reads, the uptick in the percentage of guanine at the end of the reads was gone (**Figure C.2B**), suggesting that the removal of vector sequences also improved this abnormality.

Comparison of the GC content distribution between raw (Figure C.2C) and (Cut x3 Trim) x2 (Figure C.2D) forward files shows that the broad hump seen in the raw file as well as in the other raw files (Figure 5.5) has narrowed, albeit with 3 peaks between 33-49% and a further peak at 53%. The peaks between 33-49% GC and the small bumps in the per base sequence content distribution in the (Cut x3 Trim) x2 files indicate that the sequence diversity might not be as high as in an ideal RNA-Seq run. This is expected, given the number of mutation events in each of the U937 GT libraries was predicted to fall between 1,000-10,000 (Dr Benjamin Pickard, 2018 personal communication, 29 June). The peak at 53% GC content (previously hypothesised as STM sequence contamination, Figure 5.5) was retained between raw and (Cut x3 Trim) x2 files; and thus is likely to represent low sequence diversity. Additionally, no overrepresented sequences were identified by NCBI nucleotide

Blast as being of *Salmonella* in origin, definitely excluding this explanation. The complete (Cut x3 Trim) x2 length filtering procedure will be known as pre-mapping procedure two.



**Figure C.2 Both per base sequence content and GC count distribution approach expected values after iterative cutting and trimming.** *The per base sequence content **A, B** and GC content, **C, D**, distribution for an example dataset (D-Control reverse) before (raw sequence QC) **A** and **C**, and after, **B** and **D**, quality processing with Cutadapt and Trimmomatic, described here as (Cut x3 Trim) x2. In **A** and **B**, nucleotide composition along the length of the read (plotted on the x axis) is indicated as a percentage of total nucleotides in all reads in the sample by red (thymine), blue (cytosine), green (adenine) and grey (guanine) lines. In **B**, the number of reads (read count) is plotted on the y axis against the average (mean) GC content per read given as a percentage on the x axis. The GC content distribution of the sample is shown in red and the theoretical distribution which (anticipates a binomial distribution for all reads) is shown in blue. Graphs were made with FastQC report on Galaxy.*

## C.2. Simply trimming retains more mappable reads than (Cut x3-Trim) x2

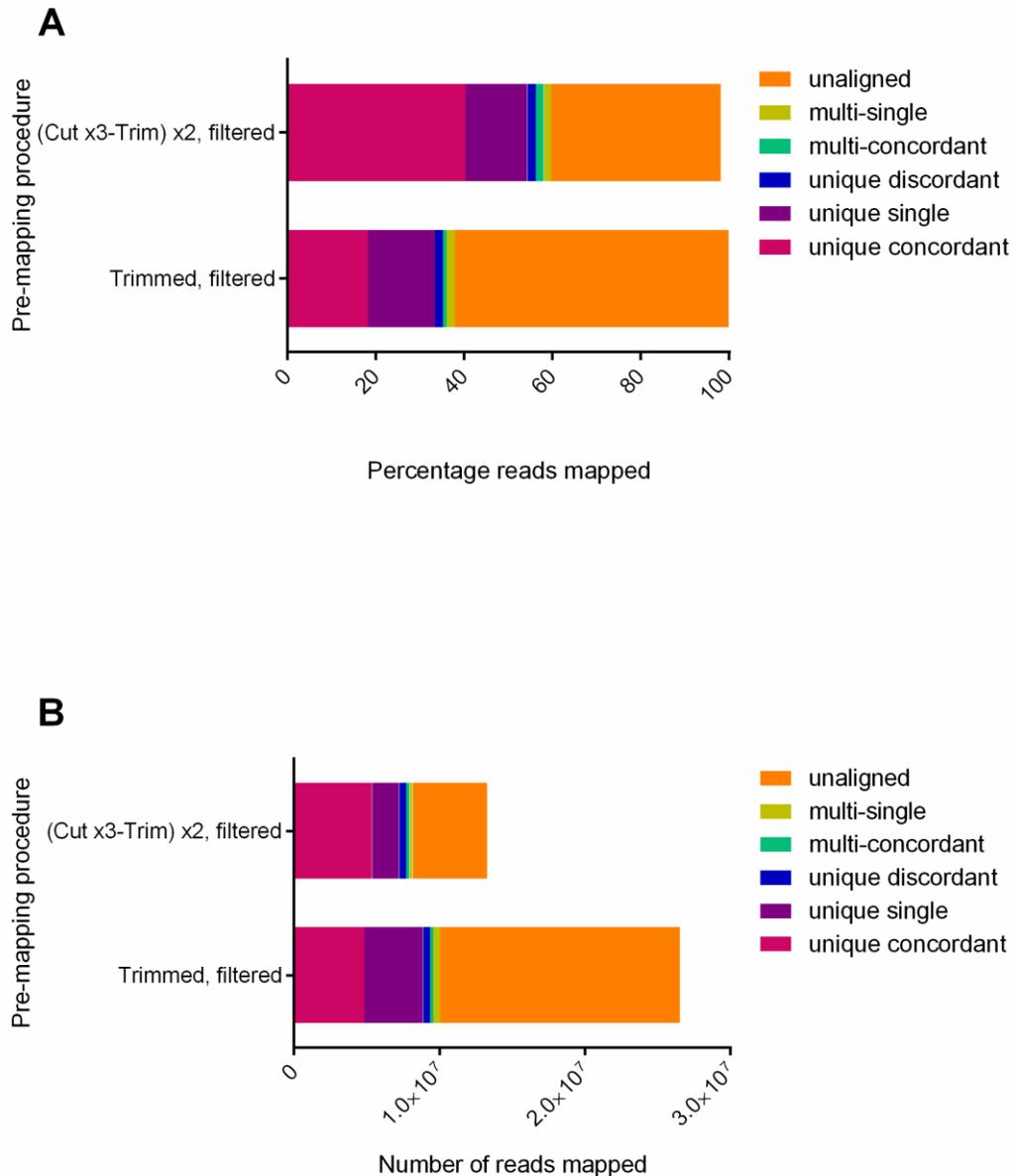
mRNA Sequencing reads need to be aligned to an official human transcriptome or genome to identify the gene they originated from. As the reads in this set of experiments are derived from mRNA produced by a GT cassette spliced to a human exon (or exons) of a gene, this may require novel transcript discovery, which requires use of the full human genome. Mapping transcripts to a reference genome requires a read alignment tool that can deal with intron length gaps in the sequence, known as a Gapped mapper (Conesa et al., 2016). The tool must also deal with mammalian genome levels of size and splicing complexity, as well as being available on Galaxy. Until 2016 TopHat was one of the most popular read alignment tools for RNA-Seq data but has since become deprecated on usegalaxy.org as it has not been updated for 4 years (Center for Computational Biology at Johns Hopkins University, 2016). HISAT2, another splice-aware read alignment tool (D. Kim et al., 2019; Pertea et al., 2016) is faster (Kim et al., 2015) and more accurate than TopHat (Baruzzo et al., 2017; Kim et al., 2015). It is available on Galaxy Main (one of the public Galaxy web servers) and was previously used by a collaborator to analyse a similar project (J. Yap Kean Yi 2019, personal communication 2<sup>nd</sup> August) and thus highly relevant technical advice was available.

At this stage, a paired end read will only align to part of a gene, unless the gene is comparatively short, as the DNA fragments the read pairs were sequenced from were only 300-350 bp long (Figure 5.1). For example, human transcripts can be as short as 19 bases, or extend to more than 2,000,000 bases (Kent et al., 2002). As previously noted, removing the exogenous sequences from the sequence reads would likely make some read pairs more mappable, but at the cost of losing potentially useful singly mapped reads. To determine which pre-mapping procedure generates the most mappable reads, the sample D-Control was subjected to both procedures before alignment onto the human genome (hg 38, hg 38) with HISAT2 (settings in Chapter 2). The HISAT2 output statistics were compared by percentage (C.3A) and number (C.3B) of reads mapped.

To decipher the read alignment output, some technical terms must first be defined. Concordant is the term used to describe the situation when the forward read maps upstream of the reverse complement of the reverse read or the reverse complement of the

forward read maps downstream of the reverse read; that is, the paired ends map onto the same gene in the expected orientation and separation. Discordant refers to pairs of reads that align uniquely but outwith the paired-end constraints. Single-mapped reads are those that map without the read partner. In order of preference, HISAT2 looks for alignments that are concordant, then discordant, then single-read. Potential read alignments are additionally scored based on canonical or non-canonical splice sites, the length of the intervening gap (in the middle of a read or between reads) as well as several other settings. The highest scoring alignments are reported. Multi-mapped reads map equally well to multiple regions of the genome, while uniquely mapped reads have one reported alignment. Multiple alignments are only reported when the alignments are concordant, or a single read aligns in multiple locations. In contrast unique alignments of any type are reported. Multi-mapped reads occur due to repetitive sequence elements as well as read matches to genetically similar genes. The literature advises that multi-mapped reads should be kept as they form an appreciable proportion of the mappable sequences (Conesa et al., 2016).

In the HISAT2 output the total percentage - 60% - of aligned fragments (including all mapped reads) is much greater in the (Cut x3-Trim) x2 version compared to the simple trimmed version - 38% (**C.3A**). This confirms that a large proportion of the unmappable reads were partially or wholly GT cassette sequences. Indeed, a greater number of paired reads mapped concordantly in the (Cut x3-Trim) x2 version (approximately 5.3 million read pairs) (**C.3B**) than the simply trimmed version (approximately 4.8 million read pairs); indicating that some GT cassette-human exon reads were recovered and mapped. However, the total number of mapped fragments was lower for the (Cut x3-Trim) x2 version (approximately 8.2 million) than the simply trimmed version (approximately 10.1 million). The number of unique single-mapped reads retained in the simply trimmed version was substantially greater than the combination of unique single-mapped reads and paired reads that became concordantly mappable in the (Cut x3-Trim) x2 version. One would expect inaccurately aligned reads to have multiple equally possible alignments and inversely that most uniquely aligned reads are accurately aligned. Extensive loss of sequence reads alters differential gene expression (and representation) output if the distribution of these reads within the genome is biased. The distribution within the genome of the uniquely aligned single mapped reads retained by the simply trimmed method is unknown; therefore, it is better to retain these reads if they can be successfully incorporated in the subsequent analysis. Therefore, the simpler approach of trimming and length filtering appears to be the best pre-mapping procedure.



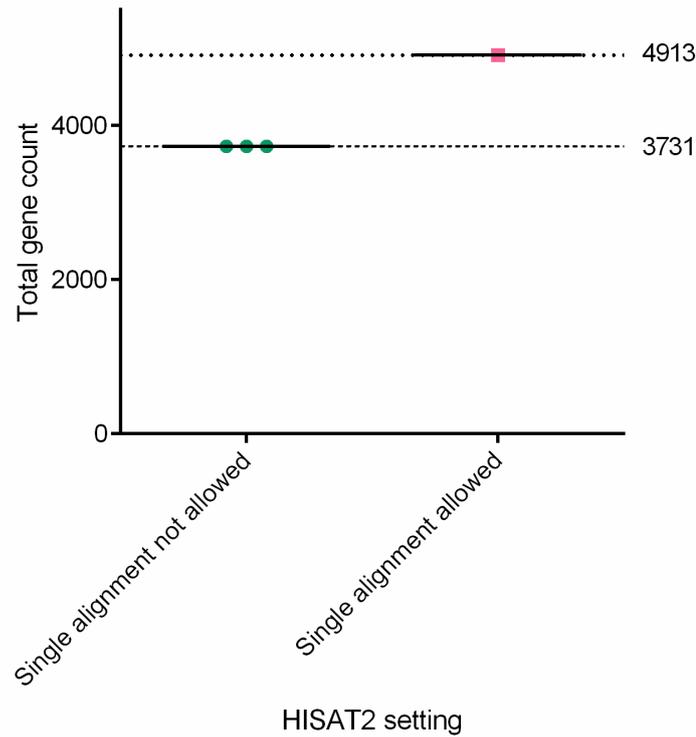
**Figure C.3** The percentage of mapped reads was higher in the (Cut x3-Trim) x2, filtered sample, but the total number of reads was lower, hence trimming and filtering seems to be the best choice. A representative paired sample was either cut, trimmed and filtered - (Cut x3-Trim) x2, filtered - or only trimmed and filtered before mapping the surviving paired reads in HISAT2. The percentage (A) and number (B) of mapped reads for each pre-mapping procedure is presented in stacked bars; the colours are representative of different types of alignment (or unaligned). “Multi-” refer to reads that map to multiple regions of the genome, in contrast unique reads map to a single location. Concordance is when both reads in a pair map according to the paired end rules applied by HISAT2. Discordant refers to pairs of reads that align uniquely but outwith the paired-end constraints. Single refers to pairs where only one read can be mapped. Experiment performed once.

### **C.3.** StringTie successfully incorporates **single-mapped** reads and paired end reads during transcript assembly and enumeration

To determine the proportion of transcripts per trapped gene in infected samples compared to control samples, fragments matching each gene must be assembled into transcripts and the number of transcripts counted (this step). StringTie is one of many tools that assembles transcripts found in each experimental sample and counts the number of fragments mapped to each transcript and each gene. StringTie was chosen for this purpose as it was available on Galaxy Main and had been previously used by a collaborator to analyse a similar project (J. Yap Kean Yi 2019, personal communication 2<sup>nd</sup> August).

To confirm which pre-mapping procedure is best, knowledge of whether single-mapped reads are incorporated into the transcript assembly and gene counts generated by StringTie is essential. To reiterate from Section 5.2.2.2, single-mapped reads refer to the read that mapped to the genome within a read pair.

To determine this, a representative sample (Original library sample 1) was trimmed and filtered, run through HISAT2 with or without the single read alignment preference and subsequently run with StringTie (settings described in Chapter 2). As part of this test, HISAT2 was run without the single alignment preference three times (each with a separate instance of StringTie). This was performed to capture any variation in the gene count resulting from equally well mapped randomly reported multi-mapped alignments. HISAT2 uses a random number generator to decide what alignments to report when fragments map equally well to multiple loci. The StringTie gene count output reports how many transcripts could be assembled for each gene, thus, all genes with one or transcripts were counted as represented. The inclusion of single read alignments in the HISAT2 run resulted in 1,200 more genes represented in the gene count data, an increase of more than 30% (Figure C.4). Additionally, no variation was seen in the gene count data between HISAT2-StringTie replicated runs on the same sample, therefore, this was not a concern. To summarise, the second pre-mapping option (simple trimming and filtering), retains more mappable reads; the biggest differing subgroup of alignments are the single mapped reads, which can (like other types of alignment) be incorporated into the downstream analysis.



**Figure C.4 StringTie makes use of single-mapped reads to produce gene counts.** *StringTie* total number of genes with 1 or more counts from HISAT2 runs with (technical  $n=3$ ) and without single alignments (technical  $n=1$ ). Q20 trimmed and 20 base length filtered original library sample 1 used in HISAT2 runs. By default, HISAT2 attempts to find alignments for both reads individually (termed mixed behaviour in programme preferences) if no concordant or discordant alignments can be found. Single-mapped reads refer to the read that mapped to the human genome within a read pair. Experiment performed once.