A high-content library screen of putative effectors from Crohn's Disease associated bacterium Adherent Invasive *Escherichia coli* LF82

Adam Collins, BSc (Hons)

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2016

Chapter 1: Contents

Abstract i

Acknowledgements ii

Abbreviations iii

Chapter 1: Introduction 1

1.1 Crohn's disease 2

1.1.1 Discovery and definition 2

1.1.2 CD aetiology 3

1.1.3 Adherent Invasive E. coli (AIEC) 9

1.1.4 Autophagy 17

1.2 Secretion systems 34

1.2.1 The type 3 secretion system 36

1.2.2 The type 6 secretion system 40

1.2.3 Outer membrane vesicles 48

1.3 Bacterial virulence factors and pathogenicity islands 53

1.3.1 Identification and purpose 53

1.3.2 Key examples of bacterial virulence factors and PAIs 55

1.4 High-content screens 59

1.5 Aims 64

Chapter 2: Materials and methods 67

2.1 Materials 68

2.1.1 Tables 68

2.2 Methods 80

2.2.1 Bacterial growth and storage conditions 80

2.2.2 Bacterial manipulation 81

2.2.3 Mammalian cell culture 85

2.2.4 Mammalian cell manipulation 86

2.2.5 Nucleic acid manipulation 90

2.2.6 Nucleic acid purification 99

2.2.7 Gel electrophoresis 102

2.2.8 Gene deletion strategy and assays 103

2.2.9 AIEC genomic library creation and screening 109

2.2.10 Seamless Ligation in Cell Extract (SLiCE) 112

Chapter 3: The identification, cloning and transfection of a subset of LF82 CDSs 117

3.1 Adherent invasive E. coli (AIEC) 118

3.1.1 Background 118

3.2 Aims 122

3.3 Workflow of LF82 CDS selection and cloning for library creation 123

3.4 E. coli genome comparisons 124

3.4.1 E. coli serotype selection 124

3.4.2 InParanoid genome analysis 124

3.4.3 Analysis of LF82 unique genes 131

3.5 Preparation of LF82 library for analysis 134

3.5.1 Preparation of LF82 unique genes for SLiCE 134

3.5.2 Seamless ligation in cell extract (SLiCE) 139

3.5.3 Transformation of library DNA 141

3.5.4 Confirmation of successful LF82 unique gene insertion into pCMV-mEm-4GS .. 143

3.6 Revisiting SLiCE failures for the creation of Set 4 147

3.7 Creation of screen reporter cell line 148

3.7.1 HeLa mCherry-LC3 creation 148

3.8 Screen of LF82 unique gene library 149

3.8.1 Transfection of plasmid library for screen 149

3.8.2 Screen image collection 154

3.9 Future perspectives 154

3.10 Summary 157

Chapter 4: Development of an automated pipeline for the analysis of GFP-tagged proteins in a HeLa cell population 159

4.1 High-content techniques for cell imaging 160

4.1.1 Background 160

4.2 Aims 162

4.3 Screen pipeline development 163

4.3.1 Software for image analysis 163

4.4 CellProfiler pipeline 165

4.5 Analysis of screen data 171

4.5.1 CellProfiler Analyst (Analyst) data analysis 171
4.5.2 R data analysis 174
4.5.3 Detection of autophagosomes 181
4.5.4 Identification of genes of interest from screen 195
4.5.5 Detection of nuclear localisation 196
4.5.6 Cell perimeter 214
4.5.7 Nuclei perimeter and area 218
4.5.8 GFP/LC3 and GFP/Actin 229
4.5.9 Radial zone 4 237
4.6 Future perspectives 244
4.7 Summary 244
Chapter 5: Analysis of candidate effectors 247
5.1 Background 248
5.1.1 Genes selected for deletion 248
5.2 Aims 252
5.3 Gene deletion 253
5.3.1 Deletion strategies 253
5.3.2 Gene deletion using the no-SCAR system 255
5.4 p314 phenotype analysis 262
5.4.1 Phenotype confirmation 263
5.4.2 Co-localisation assays 263
5.4.3 Detection of secretion 272
5.5 Invasion assays 276
5.5.1 Invasion assay optimisation 276
5.5.2 Reverse transcriptase PCR (RT-PCR) 281
5.6 Future perspectives 283
5.7 Summary 285
Chapter 6: Conclusions and future perspectives 287
6.1 Research overview and discussion 288
6.1.2 Future work 295
6.1.3 Summary 297
References 299
Appendix I: Primers for LF82 unique genes 325

Appendix II: LF82 unique genes..... 335

Appendix III: Layout of screen transfection plates 343

Appendix IV: Screen analysis graphs 347

6.2 Radial zone 2 347

6.3 Radial zone 3 350

Abstract

Adherent invasive *E. coli* (AIEC) is a recently discovered *E. coli* pathovar which has been closely associated with Crohn's disease (CD). CD is a form of inflammatory bowel disease (IBD) which has seen increasing prevalence in recent years, however, its aetiology still remains elusive. The initiation of CD is thought to involve a complex interplay of factors, including: genetics, environmental factors, and the presence of key members of the gut flora, such as AIEC.

While AIEC can be found in both healthy individuals and CD sufferers, the latter are believed to display a range of genetic mutations which collectively allow the entry of AIEC into host epithelial cells, and prevent bacterial removal by autophagic mechanisms. Two of the most prevalent mutations observed in CD patients include mutations in CAECAM6, a host cell receptor used by AIEC, and mutations in NOD2, an autophagy-associated protein.

Once inside host cells, AIEC appears to stimulate an inflammatory response. This, along with the persistent nature of AIEC, is thought to result in chronic inflammation of the gut. Despite mounting evidence for the role of AIEC in CD, this pathovar remains understudied, with its method of invasion and effects upon host cells still poorly understood.

To help characterise and determine specific genes utilised during invasion, AIEC strain LF82 was compared to other *E. coli* to identify coding sequences (CDSs) distinctive to LF82. From these, a library of expression vectors was created for the expression of each CDS fused to GFP. A high-content phenotypic screen (HCS) was then developed to examine the effects of LF82 CDS expression upon HeLa cells. This data was used to identify CDSs of interest, with particular focus on putative effectors. Preliminary experiments for characterisation of hits from this screen were then performed.

This study resulted in the screening of hundreds of putative LF82 effectors using an HCS. This system was able to identify individual CDSs of interest, which may play a role in CD initiation and/or progression.

Acknowledgements

Firstly, I would like to thank Alan for his scientific expertise, his friendship, and for his lab-DIY skills. Your thirst for knowledge encouraged me on those many occasions where things did not seem to quite go to plan. Similarly, thanks are given to Bill Wickstead, whose knowledge of R, Python and bioinformatics saved me many hours of difficulty on numerous occasions.

Thanks are also extended to Jonathon, Richard, Augusto and Rob. Despite the brief time that I had the pleasure of working with you in the Huett lab, your assistance in experimental endeavours, thought provoking comments, and provision of merriment were a welcome aid – and at times, also a welcome distraction – to the progression of my PhD. Thank you to all the members of D119 for your ideas, and helpful discussion in lab meetings, as well as in the office.

To my family who have supported me throughout my undergraduate degree, and my PhD – thank you for your support, both financially and emotionally. Without you, I do not know where I would be! Thank you to my Grandma, Joan, for taking out that genetics library book so that you could talk to me and understand better just what it was that I did day to day!

And finally, thank you to Sophie, for always reassuring me when times were hard, sharing excitement when long overdue experiments finally succeeded, and for tolerating my early alarms for the past 3 and a half years!

Abbreviations

AIEC	Adherent invasive *E. coli*
ATP	Adenosine-5'-triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base-pair
°C	Degrees centigrade
CFU	Colony forming units
CPEC	Circular polymerase extension cloning
Δ	Delta – representing gene deletion
Hoechst	Used here in reference to Hoechst 33342
DNA	Deoxyribonucleic acid
DSB	Double strand break
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme
eGFP	Enhanced Green Fluorescent Protein
FISH	Fluorescent *in situ* hybridisation
g	Gram
Gent	Gentamicin
GFP	Green fluorescent protein
h	Hour/s
HCS	High-content screen
kb	Kilobase
kDa	kilodalton
L	Litre
LF82	Adherent invasive *E. coli* strain LF82
M	Molar
Mb	megabase
mCherry	mCherry Fluorescent protein
µ	Micro
min	Minute/s

mL	Millilitre
mm	Millimetre
MOI	Multiplicity of infection
MW	Molecular weight
nm	Nanometres
no-SCAR	Scarless Cas9 Assisted Recombineering
nt	Nucleotide/s
OD_{600}	Optical density at 600nm
ORF	Open reading frame
PAMP	Pathogen-associated molecular pattern
PAI	Pathogenicity island
PCR	Polymerase Chain Reaction
pH	Negative $\log_{10}$ of activity of hydrogen ion
ROS	Reactive oxygen species
Rpm	Revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
s	Second
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNA	Small-guide RNA
ssDNA	Single stranded DNA
Tet	Tetracycline
T#SS	Type # secretion system (e.g. T6SS)
V	Volts
g	Unit of gravitational acceleration force equal to 9.8
X^{100}	Substance 'X' at a concentration of 100 mg/ml
X^{R}	Resistance to substance ‘X’

Chapter 1: Introduction

1.1 Crohn's disease

1.1.1 Discovery and definition

Inflammatory bowel disease (IBD) is a chronic condition that causes inflammation of the gastrointestinal tract (Fakhoury et al. 2014). Crohn's disease (CD) and ulcerative colitis (UC) are two major clinical subtypes of IBD. Together, CD and UC are estimated to affect 1.5 million Americans and 2.2 million people in Europe, although estimates vary (Ananthakrishnan 2015). CD is characterised by chronic discontinuous, transmural inflammation and is usually restricted to the distal ileum and colon, but can affect anywhere from the mouth to the anus. UC is a relapsing, continuous, circumferential inflammatory disease, typically restricted to the colon; it frequently causes bloody diarrhoea during more severe relapses (Neurath 2014).

The term 'ulcerative colitis' was first coined for a subtype of IBD by Sir Samuel Wilks in 1859, and it was not until 1932 that Burrill Bernard Crohn identified Crohn's disease as a subtype of IBD, distinct from UC (Mulder et al. 2014; Crohn et al. 1952). Since then, epidemiological studies have suggested an upward trend in both CD and UC incidence worldwide (Molodecky et al. 2012). This is concerning, given that both CD and UC result in an increased likelihood of mortality compared to the general population, with CD sufferers seeing a 50% increase and UC sufferers a 10% increase in long-term mortality (Jess et al. 2013). More recently, the increase in mortality seen in IBD patients has been attributed to colorectal cancer. However, other complications such as *C. difficile* acquisition during treatment are often overlooked, calling this finding into question (Kassam et al. 2014; Ricciardi et al. 2009; Loftus 2004b).

The economic cost of treatment for CD and UC sufferers is high – particularly for CD – with estimated costs in the UK reaching £3902 per person per year for CD, and £2964 per person per year for UC (Petryszyn & Witczak 2016; Rencz et al. 2015; Stone 2012; Odes 2008). While treatment options are

becoming more refined, particularly with CD and 'top-down' treatment strategies (where intensive therapeutic treatment is encouraged from initial diagnosis and gradually tapered if and when symptoms improve), their costs remain high (Amezaga & Van Assche 2016). More effective and expensive therapeutic treatment agents that target tumour necrosis factor (TNF)-α (see Figure 1.1, later), such as infliximab, have been approved for use with CD since 1998, and since 2006 for UC (Odes 2008). However, long-term health care costs are difficult to calculate, and vary depending upon disease severity and which surgical interventions are required, if any (Petryszyn & Paradowski 2015).

Given the economic costs, mortality rates, and increasing incidence of IBD, there is growing need for an understanding of IBD aetiology. This is particularly true with CD due to higher associated treatment costs and heightened mortality rates in comparison to UC.

1.1.2 CD aetiology

CD, like UC, is an idiopathic disease with a multitude of associated 'risk factors' but an elusive aetiology. CD has been linked to various genetic, environmental, immunological and microbial risk factors, with a general consensus that disease is caused by a complex interplay of these factors, rather than simply one alone (van Assche et al. 2010).

1.1.2.1 Genetic and immunological factors

One of the first CD susceptibility genes identified was the nucleotide-binding oligomerization domain 2 encoding gene, *NOD2* (Hugot et al. 2001; Ogura et al. 2001). Individuals homozygous for *NOD2* mutations have a 20-40x increased likelihood of developing CD, whilst heterozygous individuals have a 2-4x increased risk (Bhullar et al. 2014). NOD2 is expressed primarily by macrophages and dendritic cells, but also by intestinal epithelial cells, and is involved in the recognition of muramyl di-peptide (MDP) – a core bacterial

cell wall component (Girardin et al. 2003; Lala et al. 2003; Y Ogura et al. 2001). Maekawa et al. (2016) have recently solved the structure of rabbit NOD2, and note that the mutations associated with CD are distributed throughout the protein, and are predicted to affect oligomer formation and/or ligand binding.

NOD2 is part of the Nod-like receptor (NLR) family and, as such, contains a centrally located nucleotide binding domain, and multiple leucine rich repeats (LRRs) at its C-terminus (Inohara et al. 2005). NLRs are responsible for intracellular detection of so-called pathogen-associated molecular patterns (PAMPs) and subsequent initiation of the formation of an inflammasome, and thus initiate the inflammatory response (Wen et al. 2013). The most common mutations seen in CD-associated *NOD2* are a frameshift polymorphism (Leu1007fsinsC), and two single nucleotide polymorphisms (Arg702Trp and Gly908Arg) (Ek et al. 2014; Hampe et al. 2001; Hugot et al. 2001; Ogura et al. 2001). However, the frameshift variant of *NOD2* appears to be the most reproducible CD-associated genetic risk factor, with it routinely identified across multiple populations (Serbati et al. 2015; Vavassori et al. 2004).

With inflammation being a defining characteristic of CD, the discovery of *NOD2* involvement in CD may initially seem counterintuitive, as a misfolded or truncated NOD2 may be unable to stimulate an inflammatory response. However, given the role of wild-type NOD2 in the regulation of cellular defences via the detection of MDP, it is thought that mutations in NOD2 result in a cell's inability to combat pathogenic attack. While dysregulation of the inflammatory response in epithelial cells may be problematic, in other cells such as macrophages, an insufficient response to internalised bacteria may prove to be disastrous for the regulation of pathogens within the gut (Vazeille et al. 2015; Lapaquette et al. 2010).

Since the identification of *NOD2* as a risk factor, more recent research has failed to identify more strongly associated individual genetic loci (Martinek et al. 2015). However, there are now over 160 distinct risk alleles that have been identified in relation to CD susceptibility (Jostins et al. 2012). Of these, there are several loci in which single nucleotide polymorphisms are consistently observed, including *Irgm*, *Lrrk2* and *Atg16l1* (Adolph et al. 2013; Gardet & Xavier 2012). Both *Irgm* and *Atg16l1* have roles in the innate defence and management of intracellular invasive organisms such as *Salmonella enterica* serovar Typhimurium (*S.* Typhoimurium) and adherent invasive *E.* coli (AIEC), while *Llrk2* has also shown a tentative connection with cellular immunity via the regulation of reactive oxygen species (ROS) (Gardet & Xavier 2012; McCarroll et al. 2008; Kuballa et al. 2008). Mutations in *Atg16l1* are of particular interest as they appear to affect the release of antimicrobial factors into the gut lumen by Paneth cells, aiding bacterial survival and possibly resulting in the localised overgrowth of members of the intestinal flora (Klionsky 2009).

1.1.2.2 Environmental factors

Many environmental factors have been suggested to play a role in CD, including smoking, antibiotic exposure, diet, and stress (Ananthakrishnan 2015). Perhaps the most interesting of these is smoking, with smokers twice as likely to suffer from CD, and yet less likely to suffer from UC (Mahid et al. 2006; Thomas et al. 2000). While it remains unclear exactly what component of cigarette smoke affects IBD, nicotine has been suggested as the most likely candidate, with its presence shown to result in a decreased production of proinflammatory molecules such as IL-2, IL-8 and TNF-α (Wang et al. 2003; van Dijk et al. 1998).

The use of antibiotics in early infancy appears to increase the likelihood of an individual being diagnosed with IBD in later childhood – particularly CD (Shaw et al. 2010). The early life flora of the gut is thought to influence the

development of the immune system, and has been linked to a wide range of diseases and disorders such as obesity, autoimmune diseases, certain cancers and diabetes (Brito & Alm 2016; Cernada et al. 2016; Soderborg et al. 2016; Koleva et al. 2015). While environmental challenges experienced may determine opportunistic residence of some microbes in the gut, the resulting microbial diversity (or lack thereof) and presence of key species may also play a role in CD aetiology and disease progression.

1.1.2.3 Microbial factors

Johne's disease (JD) is a chronic intestinal inflammatory disease that affects ruminants and is caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP) (Dalton et al. 2014). Similarities between JD and CD have been frequently drawn in the past, leading to the suggestion of MAP as a potential causative factor of CD (Loftus 2004a). However, while MAP has been partially implicated in the instigation of CD, involvement appears to be highly complex, and likely relatively minor (Manuc et al. 2016; Sartor 2005).

In recent years, evidence from both genetic and environmental studies has supported the notion that gut microbes play a role in CD. Several studies have, for example, identified increased levels of *E. coli* and other members of the Enterobacteriaceae in the gut microbiota, and an increased prevalence of *Campylobacter concisus* in intestinal biopsies compared to healthy controls (Morgan et al. 2012; Zhang et al. 2009; Michaud et al. 1998). Other studies have noted a reduced diversity of the intestinal microbiota, as well as a higher likelihood of *Clostridium difficile* infection (Trifan et al. 2014; Manichanh et al. 2006).

The sites of inflammation in CD appear to contain some of the highest concentrations of bacteria in the gut, potentially implicating a wide range of the estimated 10^{14} microbes and their ~1100 different species present (Qin et al. 2010). It is therefore no surprise that antibiotics have been suggested for

treatment and prophylactic use in CD. However, limited success has been observed with this strategy, and the efficacy of antibiotics as part of the treatment regime remains unclear (Scribano & Prantera 2013).

The implication of gut microbes in CD raises questions about host immune responses. As described above, genetic risk factors support the notion that CD sufferers have a compromised immune system, lacking the ability to defend against some microbes. This concept has been discussed in the literature, with multiple studies providing evidence to support the theory (Manuc et al. 2016). While this may seem initially counterintuitive, with increased levels of inflammation potentially suggesting an over-active immune response, the theory suggests that bacterial colonisation occurs in an initially immunodeficient individual, which causes constant and prolonged stimulation of the immune system due to the patient's inability to remove the bacteria (Marks et al. 2010; Hayee et al. 2010).

A range of cells have been implicated in IBD inflammation, including intestinal epithelial cells, M-cells and some leukocytes. It is suggested that bacterial antigen presentation – which occurs largely in in Peyer's patches – may lead to the localised inflammation seen in CD through an overexpression of interleukins (IL) – particularly IL-12 (Manuc et al. 2016; C. Jung et al. 2010a). Furthermore, as mentioned above, dysbiosis caused by the effects of dysfunctional Paneth cells, may also play a role. The production of multiple pro and anti-inflammatory cytokines – including interferon-gamma (IFN-γ), IL-4, IL-5, IL-6, IL-10, IL-12 and IL-13 – as well as differential Th1 or Th2 responses, and the involvement of multiple cell types, creates a complex web of immunoregulation that is yet to be fully untangled (Wallace et al. 2014).

Evidence suggests that colonisation of the gut by key bacterial strains may impact CD initiation, progression and prognosis. Prevention of gut colonisation may therefore be used as a measure to prevent CD or reduce its severity. Carcinoembryonic antigen–related cell adhesion molecule 6

(CEACAM6) is a receptor molecule found on ileal epithelial cells, and appears to promote colonisation of the ileal mucosa by certain forms of *E. coli*; the so-called adherent invasive *E. coli* (AIEC) (Keenan et al. 2014; Barnich et al. 2007). CEACAM6 may therefore be used as a therapeutic target in the future to help restrict undesired bacterial colonisation. Furthermore, research suggests that the use of probiotics containing *Lactobacillus casei* or non-pathogenic *E. coli* strains can inhibit the adhesion, invasion, and colonisation of bacteria associated with CD, such as AIEC (Ingrassia et al. 2005; Boudeau et al. 1999).

Despite arguments for the role of an expanding list of microorganisms in CD, Adherent Invasive *E. coli* (AIEC) has perhaps been observed most prominently, especially in recent years (Bosca-Watts 2015; Barnich & Darfeuille-Michaud 2007). It has therefore been the focus of many recent studies, including the research presented here. Despite the strong association of microbes with CD, however, it is clear that genetic, environmental and bacterial factors all play a role in the development of CD. A proposed overview of this process can be seen in Figure 1.1., which describes the involvement of AIEC, CEACAM6, and host genetic susceptibility factors in CD progression.

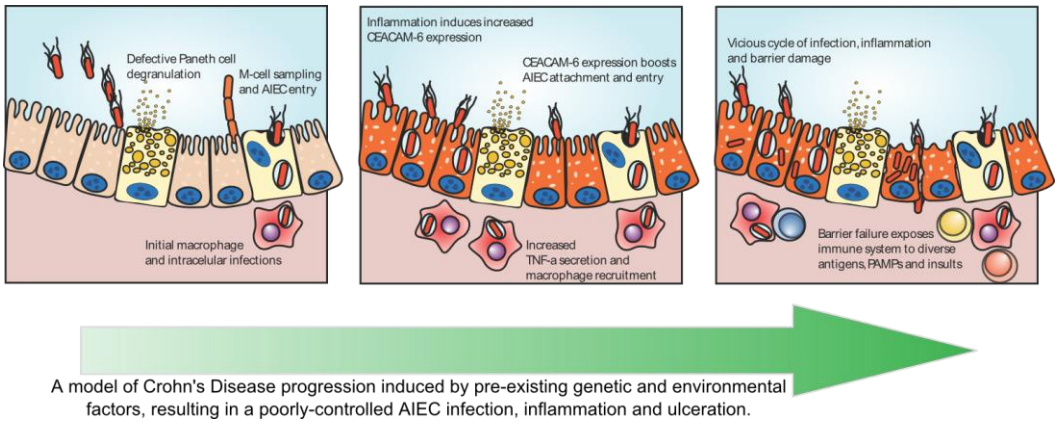


Figure 1.1. **A proposed model for the progression of CD**. Left: AIEC is present as a commensal in the gut. Host susceptibility factors (such as defects in *Atg16l1*) can result in defective Paneth cell degranulation, which may result in localised dysbiosis. AIEC capitalises on this and invades M-cells via long polar fimbriae interaction. Macrophages in the lamina propria attempt to phagocytose any bacteria which cross the epithelial barrier.
Middle: The intracellular presence of AIEC in macrophages results in an increase in TNF-α production, recruiting more macrophages to the site. Inflammatory responses to the bacteria also result in localised inflammation and an increase in CEACAM6 production, through which AIEC gains entry to epithelial cells after interaction with pili. Right: AIEC can survive within both epithelial and macrophage cells, resulting in a vicious cycle of infection and inflammation. Figure courtesy of Dr. Alan Huett (University of Nottingham, UK).

1.1.3 Adherent Invasive *E. coli* (AIEC)

1.1.3.1 E. coli as a commensal and pathogen

E. coli is the most common aerobic Gram negative species identified in the 'normal' gut microbiota, and typically exists as a purely commensal microorganism (Nataro & Kaper 1998). However, many strains of *E. coli* are capable of causing disease which can range from causing bouts of diarrhoea to life threatening illness. These pathogenic variants (known as pathovars) have been subdivided into several groups, such as uropathogenic *E. coli* (UPEC), enterohaemorrhagic *E. coli* (EHEC) and enteroinvasive *E. coli* (EIEC) (Croxen & Finlay 2010).

Perhaps most notable, is the production of shiga toxin by a subgroup of the EHEC – the so-called shiga toxin-producing *E. coli* (STEC). Shiga toxin

production can result in the development of haemolytic uremic syndrome (HUS), which can cause kidney failure and death. Strain O157:H7 of the STEC is most commonly associated with HUS in North America and Europe, and is a particularly virulent member of the STEC (Croxen & Finlay 2010). The STEC have seen extensive research since the 1970s due to their associated with disease (Lee et al. 2007). Since then, other pathovars of *E. coli* have been described, such as adherent invasive *E. coli* (AIEC), which was recovered from the human gastrointestinal tract (Boudeau et al. 1999).

1.1.3.2 Identification of *E. coli* associated with Crohn's disease

Some of the highest concentrations of bacteria in the human gastrointestinal tract are present in the colon and distal ileum – regions which are commonly associated with IBDs (Agus et al. 2014). The variety of bacteria present in these regions may help to explain the difficulties seen in identifying bacteria associated with CD – if any. However, over the last few decades, an association between *E. coli* and CD has been frequently observed across numerous populations, using a multitude of techniques (Table 1.1) (Agus et al. 2014). This has led to the identification of a new subtype of *E. coli,* which appears to strongly adhere to and invade epithelial cells more than typical commensal *E. coli* strains. A new pathovar of *E. coli* was therefore proposed: adherent invasive *E. coli* (AIEC) (Michaud et al. 1998).

The term 'adherent invasive *E. coli*' was first coined by Professor Arlette Darfeuille-Michaud in 1999, where the organism was shown to be linked with CD (Boudeau et al. 1999). Since then, extensive research has supported this initial observation, with numerous studies detecting an increased presence of AIEC in CD patients in comparison to healthy controls (Agus et al. 2014). Despite this strong connection, however, the detection of AIEC in CD patients can vary from 36% to 65%, suggesting that AIEC is not solely responsible for the disease (Barnich & Darfeuille-Michaud 2007; Darfeuille-Michaud et al. 2004). Furthermore, the presence of AIEC in 5-6% of healthy individuals,

although at a lower rate, also supports the idea that AIEC alone is unable to cause CD (Bosca-Watts 2015; Darfeuille-Michaud et al. 2004).

1.1.3.3 AIEC in Crohn's disease

AIEC is regarded as a pathobiont – a typically commensal bacteria that can cause disease (Tyrer et al. 2014). There are currently seven identified strains of AIEC which have been sequenced, with type-strain LF82 the most widely utilised for laboratory research (Zhang et al. 2015). AIEC can adhere to and invade various human cells and cell lines, and CEACAM6 (mentioned previously) often appears to be intimately involved with this adherent phenotype (Bosca-Watts 2015).

The bacterium interacts with CEACAM6 using type I pili, and when inside its host cell, appears to actively upregulate CEACAM6 production, encouraging entry of yet more AIEC cells (Barnich et al. 2007). When transgenic mice harbouring a bacterial artificial chromosome containing complete human *CEACAM3*, *CEACAM5* (*CEA*), *CEACAM6*, and *CEACAM7* genes are challenged with AIEC, CEACAM6 has been shown to colocalise with LF82, highlighting its role in AIEC invasion (Carvalho et al. 2009). However, while the presence of CEACAM6 does appear to increase AIEC attachment and invasion efficiency, it should be noted that it is not strictly necessary for adherence. This is particularly evident when considering AIEC strain NRG857c, which readily colonises strains of mice lacking *CEACAM6* (Small et al. 2013).

AIEC has also been shown to have the ability to form biofilms on the gut mucosa, and to increase the permeability of epithelial cell monolayers (Wine et al. 2010; Martinez-Medina et al. 2009). These adherence-related phenotypes allow intimate contact of AIEC with the epithelial cells of the intestine and, in turn, allow bacterial invasion.

Table 1.1. **Countries and methods of study showing CD-*E. coli* association**. Table adapted from Agus et al. (2014).

Date	Country	Method	Sample
1978	UK	Antibody	Blood
1978	UK	Culture	Ileal and colonic biopsies
1997	Netherlands	DNA probe	Rectal biopsies and faeces
1998	France	Culture	Ileal biopsies
2001	France	Ribotyping	Ileal biopsies
2004	Ireland	Nested PCR	Microdissected granulomas
2004	France	Culture	Ileal biopsies
2005	UK	FISH	Rectal biopsies
2006	USA	OmpC Antibody	Serum
2007	UK	qPCR	Ileal biopsies
2007	USA	Culture	Colonic biopsies
2009	France	qPCR	Ileal biopsies
2009	Sweden	qPCR	Ileal biopsies
2009	Denmark	Culture	Faeces
2010	Germany	Cloning	Colonic biopsies
2010	Germany	qPCR	Faeces
2011	France	qPCR	Faeces
2011	Brazil	Culture	Rectal biopsies
2012	Brazil	Culture	Ileal, colonic, and rectal biopsies
2013	USA	Culture	Ileal biopsies
2013	UK	culture	Ileal, ileocolonic, and colonic biopsies
2014	Australia	qPCR	Ileal biopsies

1.1.3.4 AIEC adherence and invasion

Many bacterial species have developed the ability to invade and survive within human cells – particularly those of the intestine. *Salmonella enterica* utilises a type 3 secretion system (T3SS) located on *Salmonella* pathogenicity island 1 (SPI-1) to deliver effectors which manipulate the target cell cytoskeleton and permit bacterial invasion (Lorkowski et al. 2014). *Shigella flexneri* also utilises a T3SS to manipulate the host cell and gain entry, although exactly how invasion occurs is yet to be established (Weiner et al. 2016). *Yersinia pestis*, meanwhile, relies on internalisation by macrophage phagocytosis, at which point it utilises a plasmid-encoded T3SS to deliver effectors (such as LcrV and Yop effectors) to the host cell, circumventing phagocytic killing (see next section) (Pujol & Bliska 2005; Cornelis et al. 1998). Pathovars of *E. coli*, such as enteropathogenic *E. coli* (EPEC), manipulate host cells via the delivery of the translocated intimin receptor (Tir) using a T3SS. This effector inserts into the host cell membrane and serves as a receptor for EPEC adhesion.

In comparison to the above organisms, the invasive properties of AIEC are not well understood. Little is known about how the organism gains entry into host cells, which effectors are used, or which secretion systems (if any) are utilised, although entry has been observed to occur in a micropinocytosis-like manner (Glasser et al. 2001). What is known is that AIEC lacks any of the canonical invasion machinery of *Shigella* or *Salmonella* and lacks a T3SS, whilst harbouring two type six secretion systems (T6SS) (Miquel et al. 2010). Early research by Massier et al. (2015) suggests that these T6SSs may be involved in successful AIEC invasion, although further research is required.

Recent progress in the identification of AIEC virulence factors has shown a link between the IbeA invasin – also found in avian pathogenic *E. coli* (APEC) – and AIEC invasion, highlighting some of AIEC's close similarities to APEC and possibly elucidating one component involved in invasion (Cieza et al. 2015).

1.1.3.5 AIEC persistence

The chronic and life-long nature of CD suggests that the root cause of inflammation persists, or constantly re-appears, for the length of the disease, however, whether this cause is a persistent microbe or not is yet to be seen. In support of AIEC involvement with the persistence of inflammation in CD is its ability to not only invade intestinal epithelial cells, but to also survive and replicate within them while simultaneously stimulating prolonged localised inflammation (Eaves-Pyles et al. 2008). Furthermore, AIEC has been shown to be able to invade, survive and replicate within human macrophages, providing perhaps even more opportunity for sustained inflammatory stimulation (Elliott et al. 2015; Glasser et al. 2001). An example of AIEC's ability to attach, invade and replicate inside host cells has been shown through transmission electron micrographs produced by Darfeuille-Michaud et al. (2004) (Figure 1.2).

Whilst resident inside ileal epithelial cells, AIEC strains have been shown to stimulate a range of responses such as: the induction of chemokine ligand 20 (CCL20), IL-8, nuclear factor-κB (NF-κB); the reduction of levels of Atg5 and Atg16L1 through microRNA induction (Nguyen et al. 2014; Eaves-Pyles et al. 2008). These responses often appear to stimulate the inflammatory response, the downregulation of autophagic responses, or the recruitment of neutrophils and macrophages to the affected region, further implicating AIEC in CD (Nguyen et al. 2014).

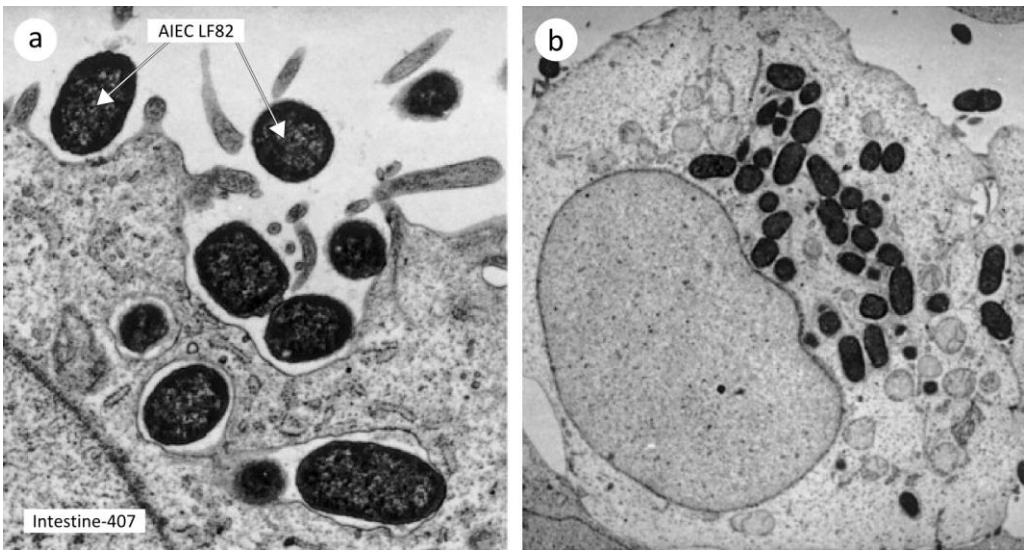


Figure 1.2. **AIEC LF82 attachment and invasion of epithelial cells**. A cross-section of intestine-407 cells infected with AIEC LF82. a) Intestinal cells showing membrane ruffling as bacteria adhere to the cell surface after a 3 h infection period. Bacteria are engulfed and enter the intestinal cell in an endocytic vacuole. b) Bacteria surviving intracellularly after escaping their endocytic vacuoles after a 5 h infection period. Figure adapted from Darfeuille-Michaud et al. (2004).

1.1.3.6 AIEC in macrophages

The recruitment of macrophages to an area witnessing a bacterial infection is an important function of the immune system (Shi & Pamer 2011). However, like several of other pathogenic bacteria, AIEC are able to evade phagocytosis when internalised by macrophages, and can survive and replicate extensively in phogolysosome-like compartments within cells (Bringer et al. 2012; Glasser et al. 2001). Other pathogenic bacteria that are capable of this include *Shigella flexneri*, *Salmonella* spp., and *Yersinia* spp. (including *Y. enterocolitica*, *Y. pestis*, and *Y. pseudotuberculosis*) (Navarre & Zychlinsky 2000).

Yersinia spp. appear to repress TNF-α expression during macrophage infection, resulting in the induction of host cell apoptosis (Navarre & Zychlinsky 2000). In contrast, LF82 have been shown to induce TNF-α

expression by up to ~2.7 fold in comparison to cells stimulated by lipopolysaccharide – a known inducer of TNF-α. The result of this is the recruitment of more macrophages to the site of infection (Bringer et al. 2012; Glasser et al. 2001). This has led to the suggestion that AIEC use TNF-α to manipulate the host immune system, stimulating the attraction of more macrophages to the infected area, and invading the newly recruited cells (Bringer et al. 2012).

Interestingly, however, Elliott et al. (2015) have observed a different TNF-α expression pattern when using AIEC and macrophages derived directly from CD patients. They showed that macrophage infection with any of the *E. coli* strains identified (two of which were suspected AIEC, and two non-AIEC) resulted in the reduction of macrophage TNF-α expression. This finding suggests that macrophages of CD patients may be unable to prevent *E. coli* invasion, and may be deficient in the signalling mechanisms necessary to recruit more macrophages to the site of infection. This supports the notion that immunodeficiency plays a significant role in CD, with cells of CD patients showing an irregular response to bacteria (Elliott et al. 2015; Tawfik et al. 2014). Despite this evidence, however, more research is needed to clarify the expression of TNF-α in response to AIEC infection *in vivo*.

To initially gain access to macrophages for invasion, AIEC must first traverse the intestinal barrier and access the lamina propria – a tissue region below the intestinal epithelial layer. A study by Chassaing et al. (2011) has shed light on how this is likely achieved, by showing that AIEC are able to target and invade M-cells above Peyer's patches in the gastrointestinal tract. This is done via the use of long polar fimbriae to interact with glycoprotein 2 (GP2) on M-cell surfaces. It is through this route that AIEC are initially thought to access the lamina propria, in which they are engulfed by macrophages (Strober 2011). After incomplete phagocytosis, however, AIEC stimulate CEACAM6 expression and the recruitment of more macrophages, as shown in Figure 1.1., above.

1.1.3.7 Summary of AIEC in CD

Inflammation of the gut in CD is likely due to a complex interplay of factors, involving: environmental factors such as diet, smoking and early gut colonisation; genetic factors affecting intestinal colonisation, bacterial handling and inflammatory responses; and microbial factors such as the presence of key bacterial species.

While it is clear that there are many risk factors associated with the persistent inflammatory response seen in CD, the stimulatory element of these responses does appear to be microbial in nature. In particular, the associated organism is seemingly able to tolerate an intracellular lifestyle, resulting in constant stimulation of the host inflammatory response. This intracellular presence is typically prevented via autophagy – a system in which CD sufferers have been shown to be at least partially deficient.

1.1.4 Autophagy

Macroautophagy, is one of three main types of autophagy: macroautophagy, microautophagy, and chaperone-mediated autophagy (See Figure 1.3). Macroautophagy involves the formation of a double membrane within the cytosol which engulfs random or selected cargo in what is known as an autophagosome. The autophagosome subsequently fuses with a lysosome to form an autolysosome and the cargo is degraded (Cuervo & Macian 2012). Microautophagy is a lesser-understood process, largely described in *Saccharomyces cerevisiae* (hereafter referred to as yeast), in which cytosolic content is directly engulfed via invaginations of the lysosome. Invaginations pinch off into the lysosome, surrounded by a single membrane, and are quickly degraded (van Zutphen et al. 2014; Cuervo & Macian 2012; Kiel 2010). Chaperone-mediated autophagy targets soluble proteins that contain a motif recognised by cytosolic hsc70. Proteins designated for degradation are chaperoned to a lysosome-bound receptor – lysosome-associated membrane

protein type 2A (LAMP-2A) – at which point the protein is internalised and degraded with the aid of lysosomal hsc70 (Cuervo & Macian 2012).

1.1.4.1 Selective autophagy

eukaryotic cells constitutively undergo macroautophagy (hereafter referred to as autophagy) at a basal level, until upregulated via stimulation by key events such as nutrient stress or bacterial invasion (Jin & Klionsky 2014). Autophagy has been implicated in a wide range of diseases, including cancer, neurodegenerative diseases, heart disease, liver disease, myodegenerative diseases and inflammatory diseases (White & DiPaola 2009; Levine & Kroemer 2008).

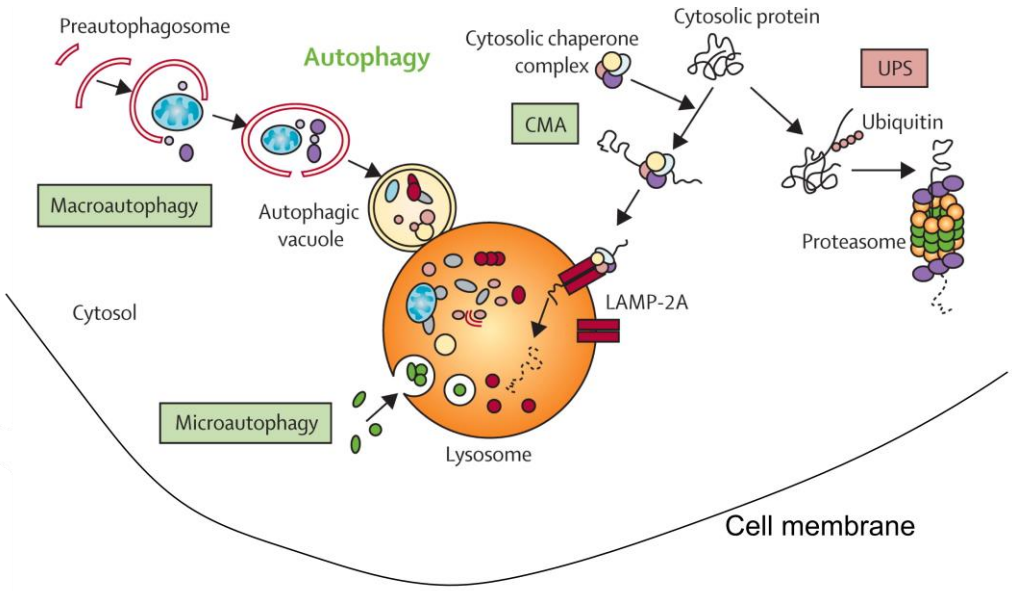


Figure 1.3**. Overview of the three main types of autophagy**. Macroautophagy: this process involves the formation of a double membrane within the cytosol which engulfs cargo in an autophagosome, an autolysosome is then formed upon fusion with a lysosome. Microautophagy: cytosolic content is directly engulfed by invaginations of the lysosome, and degraded. Chaperone-mediated autophagy (CMA): soluble cytosolic proteins containing a specific motif are chaperoned to the LAMP-2A receptor on the lysosome, where they are engulfed and degraded. Figure adapted from Martinez-Vicente & Cuervo (2007).

Autophagy was initially thought to be a non-selective process, where regions of the cytosol were randomly degraded “in bulk” as part of normal cell turnover, as well as in reaction to a stress response (Kaushik & Cuervo 2012). However, it was soon realised that, while upon first inspection autophagy seems to occur randomly, it is actually performed in a highly selective manner. As our understanding of autophagy has increased, it has become clear that autophagy may not occur randomly at all, particularly in mammalian cells when regarding cytosolic proteins (Abounit et al. 2012; Glick et al. 2010). In yeast, however, this selectivity is often achieved via the cytoplasm to vacuole targeting (CVT) pathway – although homologs of the molecular machinery involved in this pathway have not been identified in higher Eukaryotes (Umekawa & Klionsky 2012).

Autophagic selectivity in eukaryotic cells has been shown for numerous types of cargo, such as intracellular bacteria (xenophagy), mitochondria (mitophagy), and ribosomes (ribophagy) (Youle & Narendra 2011; Kraft et al. 2008; Levine 2005). The proposed survival and persistence of intracellular bacteria in CD – with AIEC in particular – therefore suggests that CD patients may have a defective or ineffective mechanism for the handling and degradation of bacteria within epithelial cells and/or macrophages. Autophagy and xenophagy have therefore been the subject of much work in relation to CD.

1.1.4.2 Autophagy machinery in different organisms

Over 40 autophagy-related genes (*Atg*) have been so-far identified in yeast, with homologues for many of these observed in other eukaryotes, including humans (Mizushima & Komatsu 2011; Weidberg et al. 2011). However, autophagy appears to be more complex in mammalian cells when compared to yeast, for example, Atg8 protein families contain six members in humans and only one in yeast (Behrends et al. 2010). In general, it appears that the non-selective autophagy Atg genes are conserved across eukaryotes,

whereas genes involved in more selective forms of autophagy are more likely to be organism-specific (Meijer et al. 2007).

1.1.4.3 Autophagy initiation

Under normal nutrient-rich conditions, the mechanistic target of rapamycin complex 1 (mTORC1) kinase represses non-selective autophagy. In yeast this is achieved via the hyperphosphorylation of Atg13, which inhibits interaction with Atg1 (Lin & Baehrecke 2015). The mammalian homologue of Atg1 is UNK1, and, under nutrient deprived conditions the UNK1 complex (containing Atg13, Atg10 and FIP200) is activated and localises to the pre-autophagosome structure (PAS) (also known as a phagophore) (C. H. Jung et al. 2010b). This process can be seen in Figure 1.4.

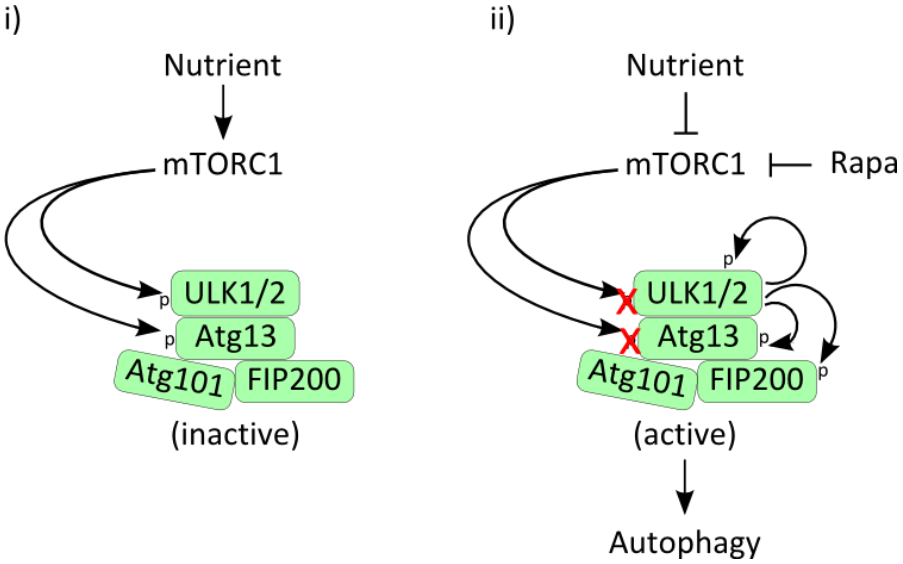


Figure 1.4. **Regulation of ULK1 and ULK2 complexes by mTOR**. i) In nutrient rich conditions, mTOR phosphorylates ULK1/2 and Atg13, preventing kinase activity. ii) Under nutrient deprived conditions, or when induced with rapamycin, mTOR is inactivated, preventing phosphorylation of ULK1/2. ULK1/2 is then activated, allowing self-phosphorylation, and phosphorylation of Atg13, Atg101 and FIP200. This complex then localises to the PAS, for autophagy. Figure adapted from C. H. Jung et al. (2010b).

The phosphatidylinositol (PtdIns) 3-kinase (PI(3)K) complex is then recruited to the PAS and, in turn, effector proteins such as WIPI1 and WIPI2 are recruited and allow autophagosome formation. For the elongation and closure of the double membrane structure, two complexes formed via a ubiquitin-like conjugation system are employed: the Atg12-Atg5-Atg16L1 complex, and the LC3 complex (described below) (Lin & Baehrecke 2015; Dooley et al. 2014; Huett & Xavier 2010; Mizushima et al. 1998). An overview of these processes can be seen in Figure 1.5.

The machinery involved the initiation of autophagy, production of the PAS and resulting production of the autophagosome is referred to as the core autophagy machinery. These components are generally regarded as essential for autophagosome initiation and formation. However, there is evidence to suggest that autophagosome production is still possible even when lacking some components, such as Atg5 or Atg7 in yeast (Nishida et al. 2009).

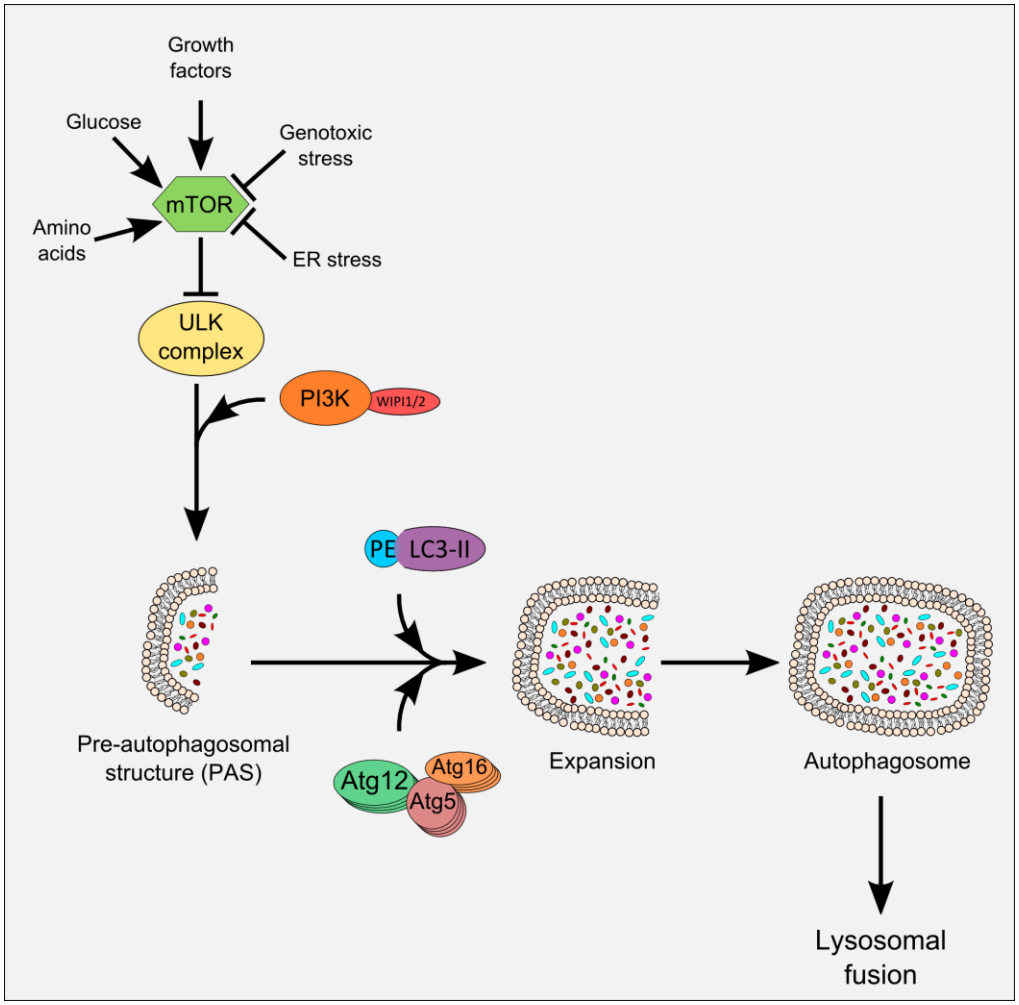


Figure 1.5. **Overview of mTOR-controlled autophagy in mammals**. mTOR repression via nutrient stress or other stresses results in ULK1/2 activation via hyperphosphorylation (see Figure 1.4., above). PI3K, along with WIPI1 and/or WIPI2, and the ULK1/2 complex then localise to the PAS. LC3-II and the Atg16L complex (see Figure 1.6., below) are also recruited to the expanding PAS, before PAS membranes fuse to form a complete autophagosome. Figure adapted from Maier & Britton (2012).

1.1.4.4 The Atg12-Atg5-Atg16L conjugation system

The Atg12-Atg5-Atg16L complex is essential for the formation of the PAS. This process is aided by ubiquitin-activating enzyme-like, and ubiquitin-conjugating enzyme-like enzymes, referred to as 'E1-like' and 'E2-like' enzymes, respectively. Atg12 is first activated by the E1-like Atg7, before being conjugated to Atg5 by the E2-like Atg10, to form an Atg12-Atg5 complex. An Atg16L homo-oligomer then interacts with Atg12-Atg5 to form

the Atg12-Atg5-Atg16L complex, allowing the expansion of the PAS. This process is summarised in Figure 1.6, alongside the LC3 conjugation system.

Expression of the Atg12-Atg5-Atg16L complex on plasma membranes encourages LC3-II localisation onto the plasma membrane, suggesting that the complex has a role in directing localisation, too (Fujita et al. 2008). NOD2 has been shown to recruit Atg16L to the membrane at sites of bacterial entry, and mutations in both *Nod2* and *Atg16L1* of CD patients (mentioned previously) may therefore produce an autophagy-deficiency phenotype (Travassos et al. 2010).

1.1.4.5 Microtubule-associated protein 1 light chain 3 (LC3)

In mammals, three protein families have shown homology to the yeast ubiquitin-like conjugation system, Atg8. Members of these families can be conjugated to phosphatidylethanolamine (PE) and localise to the autophagosome as outlined above (Kabeya et al. 2004). The archetypal example of the yeast Atg8 homologue is LC3, with much research performed to understand its function.

The C-terminal 22 amino acids of LC3 are cleaved by Atg4b after synthesis within the cell, resulting in cytosolic LC3-I. When conjugated to PE by the E2-like Atg3, LC3-II is formed and localises to the PAS (Kabeya et al. 2004). This process is shown in Figure 1.6.

Once the autophagosome is fully formed, many autophagosome-related complexes disassociate. However, LC3-II remains permanently lipidated to the autophagosome lumen, where it remains until the autophagosome fuses with a lysosome and it is subsequently degraded (Huett & Xavier 2010). This feature of LC3 has made it a useful biomarker for autophagosome identification and tracking via the observation of fused fluorescent molecules such as GFP (Pugsley 2016; Kabeya et al. 2004).

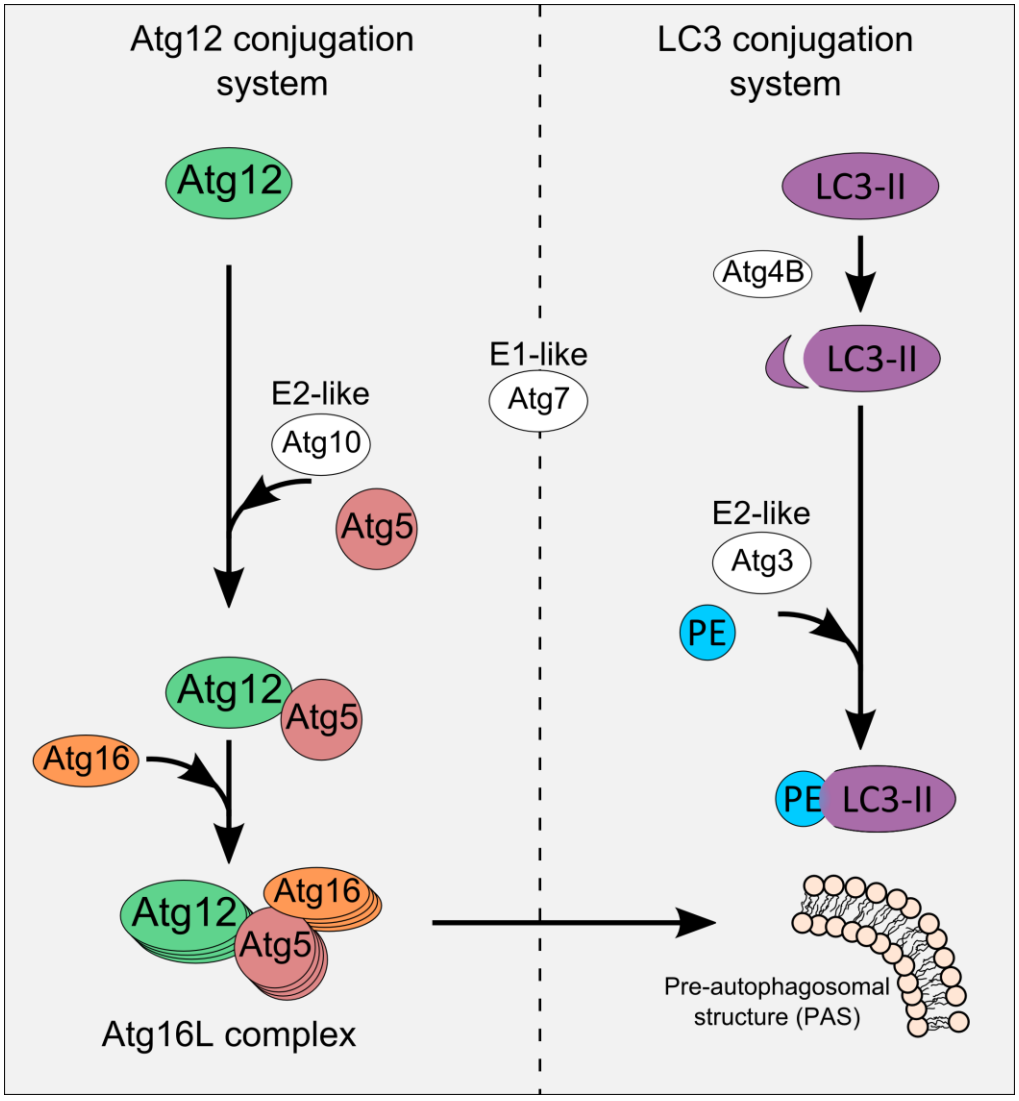


Figure 1.6. **The Atg12 and LC3 conjugation systems**. In the Atg12-Atg5-Atg16L conjugation system (left), activation and conjugation is performed using E1-like and E2-like enzymes Atg7, and Atg10, respectively. Conjugated Atg12-Atg5 then interact with Atg16 to form the Atg12-Atg5-Atg16L complex. In the LC3 conjugation system (right), LC3-I is formed by the cleavage of LC3 by Atg4b. LC3-I is then conjugated to PE by the E2-like enzyme, Atg3, resulting in the formation of LC3-II. Both the Atg12-Atg5-Atg16L complex and the LC3-II complex then localise to the PAS. Figure adapted from Pyo et al. (2012).

1.1.4.6 Formation of the autophagosomal membrane

It is still unclear whether the membrane structures used to form autophagosomes are assembled *de novo* or are recruited from other membrane-rich structures. A low abundance of protein in the autophagosomal membrane in comparison to other cellular membrane structures has been shown previously (Fengsrud et al. 2000). This suggests that the autophagosome is assembled *de novo* within the cytosol and does not incorporate other protein-rich membranes (Tooze & Yoshimori 2010). Furthermore, it has been suggested that cells display a preference for the incorporation of newly synthesised phospholipids into organelle membranes, supporting this theory (Girardi et al. 2011).

However, other research has detected both ER and Golgi markers in the autophagosomal membrane, leading to the suggestion that membrane is sourced from other cellular structures and excess protein then removed (Tooze 2013). These findings have led to the 'membrane recruitment' theory stimulating most interest in recent years, although *de novo* synthesis cannot be ruled out.

Exactly which structures – if any – are involved in donating membrane to the autophagosome is still unclear, although a variety of sources have been suggested (Carlsson & Simonsen 2015). The most commonly suggested membrane-structure appears to be the endoplasmic reticulum (ER), particularly when closely associated with mitochondria (Hamasaki et al. 2013; Hailey et al. 2010). However, many other membrane sources have also been suggested, including the Golgi, plasma membrane and endosomes (Knævelsrud et al. 2013; Geng et al. 2010; Ravikumar et al. 2010).

Difficulties in determining the origin of the autophagosomal membrane reside with differentiation between it and the cytosol in which it is formed. As autophagosomal contents inherently reflects the cytosol in which the autophagosome is formed, the isolation of the autophagosome from other

cellular components initially proved problematic. However, progress in this area has led to the identification of autophagosome-specific molecular markers, which have since aided research progress (Tooze & Yoshimori 2010). Despite this, research is still ongoing into the source of the autophagosomal membrane.

1.1.4.7 Autophagosome morphology

The size of an autophagosome typically varies from as little as 300 nm for protein aggregates in yeast, to several μm for bacteria or mitochondria in mammalian cells. What determines the size of the autophagosome is yet to be determined. One rather simplistic model is that protein-protein interactions between the cargo, intra-phagophore ligands and LC3-II directly determine phagophore size (Jin & Klionsky 2014). However, while LC3-II is thought to play a key role in extension and expansion of the autophagosome, this model does not explain the upper-limit of autophagosome size (Jin & Klionsky 2014; Xie et al. 2008). Interestingly, an increase in autophagosome number, but not size, is inducible via the deletion of *PHO23*. This gene encodes a transcriptional regulator in yeast which appears to influence *ATG9*; known to play a key role in the initial stages of PAS development (Jin et al. 2014).

Other models also exist, as seen with the suggestion that the curvature of the autophagosome is physically regulated by the shape of the Atg12-Atg5-Atg16L complex on the outer phagophore (Kaufmann et al. 2014). LC3-II has also been implicated, with levels of Atg8 in yeast shown to correlate with autophagosome size. Furthermore, PE, to which LC3-I is conjugated to form LC3-II, is a cone-shaped lipid, and has been suggested to help regulate curvature of the autophagosome via physical limitations in flexibility (Vance & Tasseva 2013; Xie et al. 2008).

1.1.4.8 Measurement of autophagy

The number of autophagosomes present in a cell is determined by frequency of production at the PAS stage, and degradation via fusion with a lysosome. This therefore provides two different aspects of measurement for autophagy: number of autophagosomes, and autophagic flux.

A high number of autophagosomes present at any one time may represent a high level of autophagosome production, or, alternatively it may represent a reduction in autophagosomal processing (a decrease in flux). Likewise, a lower number of autophagosomes present may represent a reduced level of autophagosome production, or an increase in autophagic flux. Considering one of these features alone (autophagosome number, for example), could lead to gross misinterpretation on the data. Strategies exist to differentiate autophagosome number and autophagic flux, and are described in the next section.

In yeast, the half-life of processing a phagophore through the Cvt pathway to completion is estimated to be ~45 min, whereas in nutrient deprived cells this can be as low as 10 min (Jin & Klionsky 2014). The time course over which autophagic flux occurs can therefore vary depending on cell nutrient state. While it is clear that an upregulation of autophagy leads to an increase in the number of active PAS sites, many aspects of autophagosome morphology and their related physiological mechanisms are still unclear.

1.1.4.9 Artificial manipulation of autophagy

The mechanistic target of rapamycin complex 1, as the name suggests, interacts with rapamycin – an antifungal derived from *Streptomyces hygroscopicus*. In mammals, rapamycin directly binds to FKBP12 which subsequently binds and inhibits the mTOR complex, thereby preventing the mTOR-driven negative feedback loop which usually represses autophagy (Huang et al. 2003). This process was discussed previously, and the effects of

rapamycin on autophagy can be seen in Figure 1.4 and Figure 1.5. Rapamycin is therefore commonly used in experiments to induce the cellular autophagy response that would typically result from nutrient and oxidative stress.

p62 is a ubiquitin and LC3-II binding protein that is expressed ubiquitously throughout mammalian tissues and in many of the higher Eukaryotes (but not yeast) (Komatsu et al. 2007; Joung et al. 1996). It has been shown to be degraded via autophagy after directly binding LC3-II, and to accumulate in autophagy-deficient cells (Komatsu et al. 2007). The interaction of p62 with ubiquitin and LC3-II has been shown to mediate the transfer of monoubiquitinated and polyubiquitinated protein aggregates to the autophagosome (Kim et al. 2008; Bjørkøy et al. 2005). Furthermore, p62 has been shown to play an intimate role in proteasome/autophagic interaction, with p62 expression upregulated during expression of aggregation-prone proteins, presumably to help deal with protein aggregates via both the proteasome and autophagy (Su & Wang 2011). p62 can therefore be followed to monitor autophagy, while its overexpression can also be used to induce autophagy (Mizushima et al. 2010).

While an increase in the number of autophagosomes present generally provides an indication of autophagy induction, the downstream handling of autophagosomes cannot be ignored. For example, the blockage of autophagosome-lysosome fusion may result in an increase in autophagosome numbers without an induction of autophagy. Ammonium chloride (NH_4Cl) is often utilised to discern between these two scenarios as it prevents lysosomal enzyme function, thereby allowing discrimination between an increase in autophagy, and a decrease in autophagosome handling throughput (Sun et al. 2015).

1.1.4.10 Xenophagy

The terms ‘xeno-’ and ‘-phagy’ refer to foreign content, and eating, respectively. Hence, ‘xenophagy’ refers to the degradation of content foreign to the cell (Bauckman et al. 2015). Most often, however, xenophagy is used in reference to the degradation of bacterial invaders – a process that is understudied in comparison to autophagy (Bauckman et al. 2015; Levine 2005).

Many bacteria are able to invade eukaryotic cells via phagocytosis, although most of these are destroyed by xenophagy, in which they remain compartmentalised in a phagosome before fusion with a lysosome. On the other hand, there are many bacteria which are able to both invade eukaryotic cells and reside there for long periods of time, such as UPEC (Bauckman et al. 2015). When a bacterium enters a mammalian cell, it resides – at least temporarily – within a phagosome, and may or may not be recognised by an autophagic adaptor protein such as p62 (Pareja & Colombo 2013). From here, one of three main xenophagy processing mechanisms can occur: LC3-associated phagocytosis (LAP), LC3-independent phagocytosis (LIP), or sequestration of the phagosome by a phagophore (Bauckman et al. 2015). An overview of LAP and sequestration of phagosomes by phagophores can be seen in Figure 1.7, and an overview of LIP can be seen in Figure 1.8.

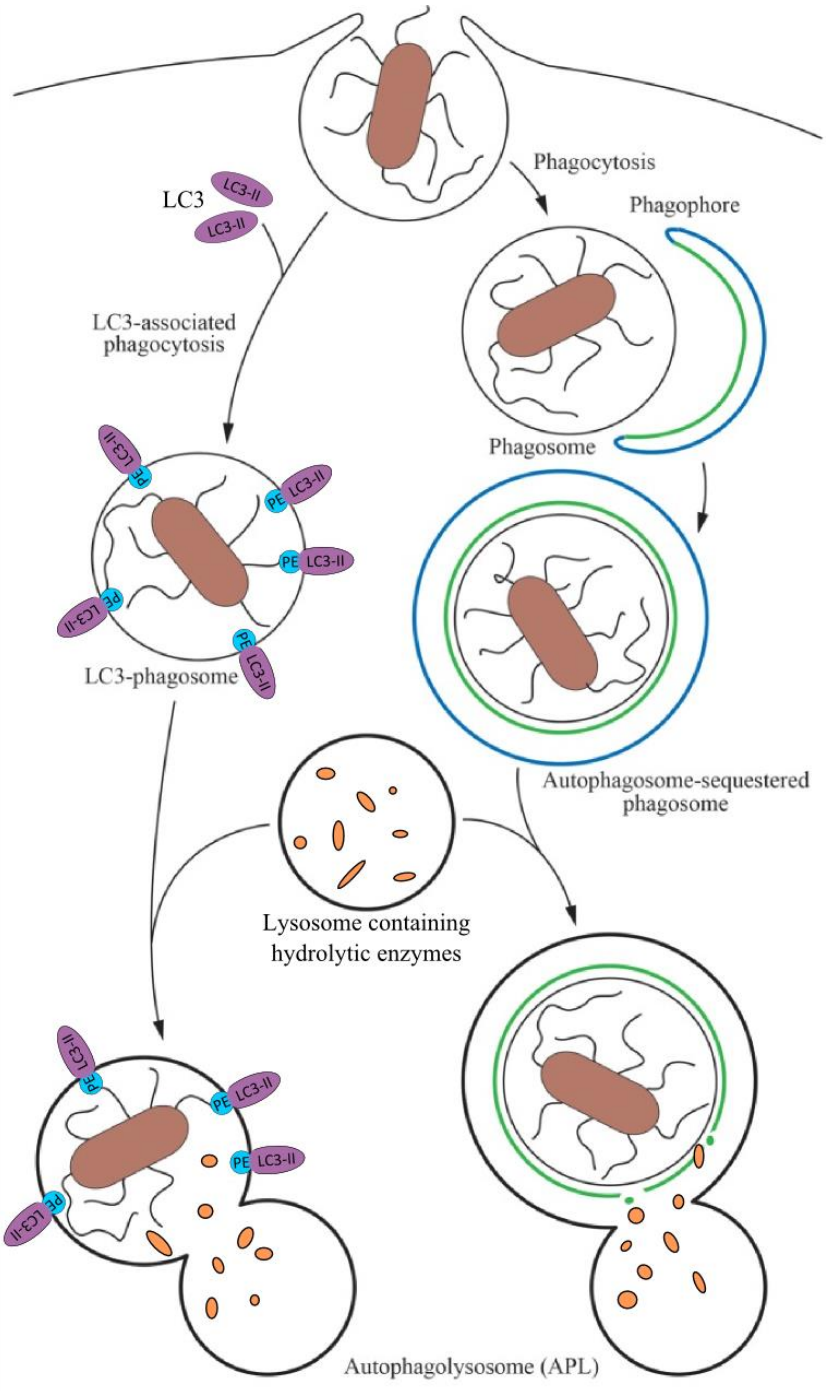


Figure 1.7. **An overview of two of the main pathways of xenophagy.** LC3-associated pathway (LAP) (left), and sequestration of a phagosome by a phagophore (right). On the left, bacteria enter the cell via phagosytosis. The phagosome is labelled with LC3, targeting the phagosome for lysosomal fusion. Upon fusion, an autophagolysosome is formed and the contents of the phagosome are degraded. On the right, a bacterium enters the cell by phagocytosis and is not labelled with LC3. Instead, the phagosome is surrounded by a phagophore, which is then fused with a lysosome to form an autophagolysosome. See below for an overview of the LC3-independent pathway (LIP). Figure adapted from Klionsky et al. (2014).

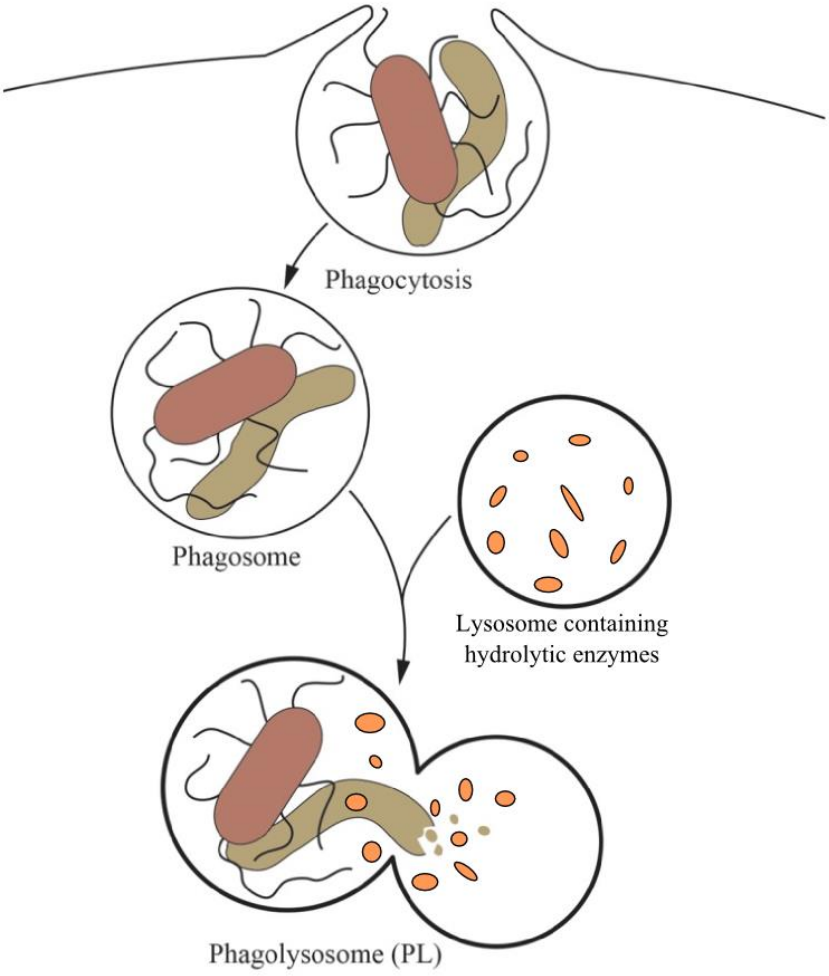


Figure 1.8. **An overview of the LC3-independent pathway (LAP) of xenophagy**. Bacteria enter the cell via phagocytosis and remain inside a phagophore. The phagophore is directly fused to a lysosome in an LC3-independent manner, resulting in the formation of a phagolysosme in which the contents of the phagosome are degraded. Figure adapted from Klionsky et al. (2014).

LAP is initiated by the tagging of a phagosome directly by LC3-II. Once tagged, the phagosome is targeted for lysosomal fusion, and forms an autophagolysosome as the contents of the phagosome are degraded. Alternatively, the LIP pathway can occur, where the phagosome is not decorated with LC3-II and is directly fused with a lysosome to form a phagolysosome. Finally, the phagosome can enter a cell and be engulfed by a phagophore (an autophagophore). At this point the usual autophagy pathway

occurs, with lysosomal fusion and creation of an autophagolysosome (Bauckman et al. 2015).

Differentiation between these structures is important when considering microbial escape from their respective compartments (Klionsky et al. 2014). The three processes mentioned deal with phagosome-derived bacteria. However, to deal with bacteria free in the cytoplasm, a more prototypical process is used in which cells are engulfed by a phagophore and fused to a lysosome to create an autolysosome (Bauckman et al. 2015).

1.1.4.11 Avoidance of xenophagy

Many organisms exist which can circumvent, halt or hinder xenophagy. Three main strategies are used by bacteria to avoid xenophagy: i) lysis and escape from the phagosome/phagophore, ii) modification of the phagosome/phagophore iii) tolerance of acidic or degradative conditions (Pareja & Colombo 2013).

Listeria monocytogenes uses method i), and lyses the phogosome using listeriolysin O (LLO) and one of two phospholipases, forming pores and disrupting the phagosomal membrane (Kayal & Charbit 2006; Goldfine & Wadsworth 2002). *Salmonella enterica* serovar Typhimurium uses method ii), and stimulates its own uptake into a modified phagosome called a *Salmonella* containing vacuole (SCV) using effector delivery via the SPI-1 T3SS. It then progressively modifies the SCV to prevent enrichment of lysosomal hydrolases, allowing bacterial survival. This is achieved using over 28 effectors secreted using a T3SS encoded on SPI-2 (Figueira & Holden 2012; Méresse et al. 2001; Haraga et al. 2008). Finally, *Coxiella burnetii* uses method iii) by forming a *Coxiella*-containing vacuole (CCV) which proceeds through the xenophagy pathway to fuse with a lysosome. However, the resulting hydrolytic enzymes present in the CCV, along with the low pH,

appear to stimulate *C. burnetii* metabolism through a yet to be determined mechanism (Newton et al. 2014).

While AIEC is able to adhere to and invade both epithelial cells and macrophages, it is not yet clear how it circumvents xenophagic degradation. Recent research has elucidated roles of pathways (such as the EIF2AK4-EIF2A/eIF2α-ATF4) and transcription factors (such as HIF1A) in xenophagic AIEC handling, although it is not clear if these are invoked by AIEC during invasion or not (Alexis et al. 2016; Mimouna et al. 2014; Chargui et al. 2012).

Therefore, while AIEC is expected to produce effectors for survival in its intracellular niche, these effectors and their method of delivery are yet to be identified. However, given the lack of any encoded T3SS and the presence of two T6SSs, delivery of effectors to target cells is suspected to be performed using one, or both, of the T6SSs present (Massier et al. 2015; Nash et al. 2010). Alternatively, effector delivery via outer membrane vesicles may circumvent the use of dedicated secretion systems (such as the T3SS and T6SS) altogether, as has been reported in *Salmonella* (see: Outer membrane vesicles) (Yoon et al. 2011).

1.2 Secretion systems

Bacteria have developed intricate and highly complex systems for the secretion of cargo into a range of target locations, including the extracellular milieu, other bacteria, and eukaryotic cells. Secretion systems are often utilised by pathogenic bacteria to deliver key proteins or toxins to target cells, and the loss of a particular secretion system often results in the loss of bacterial pathogenicity (Green & Mecsas 2016).

Due to the presence of two membrane structures in gram-negative bacteria – the inner and outer membranes (IM and OM, respectively) – secretion systems must handle the transfer of cargo across both if the cargo is destined for outside the OM. There are several known types of gram-negative secretion systems, mostly named numerically, these include: the Sec pathway, the Tat pathway, the type 1 secretion system (T1SS), T2SS, T3SS, T4SS, T5SS, and the recently identified T6SS. An overview of the dedicated secretion systems (T1SS-T6SS) can be seen in Figure 1.9, and the ability of the T3SS and T4SS to make direct contact with external membranes for protein delivery is shown. Meanwhile, the T6SS can be seen in its 'ready' state, and once fired, is also able to traverse the extracellular milieu, making contact with an external target cell membrane (not shown in Figure 1.9).

While the T1SS-T6SS are dedicated secretion systems, the general secretion (Sec) and the twin arginine translocation pathways (Tat) are more generalised, and are often responsible for the shuttling of proteins across the cytoplasmic membrane before they are secreted by a dedicated secretion system or by outer membrane vesicles (OMVs) (Green & Mecsas 2016; Schwechheimer & Kuehn 2015). The production of OMVs can therefore also be considered a method of secretion (see: Outer membrane vesicles).

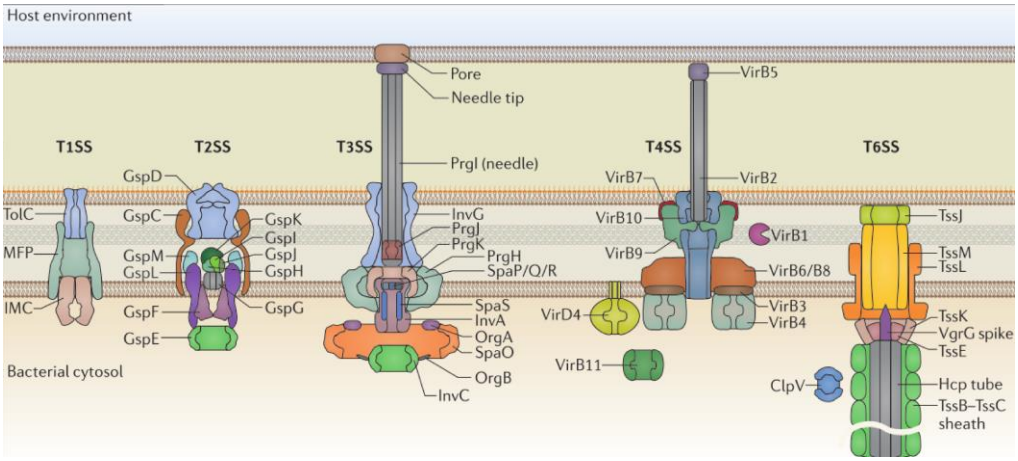


Figure 1.9. **An overview of the gram-negative dedicated secretion systems**. The T3SS can be seen extending across to the target cell membrane and forming a translocon pore through which effectors/proteins are secreted. The T6SS can be seen in its ready-state, with a needle structure formed inside the bacterial cell, ready to be fired across to the target cell membrane. For the T6SS, ClpV remains nearby, ready for recruitment post-T6SS-activation. Figure adapted from Costa et al. (2015) (license number: 3954421289494).

While *E. coli* are capable of using a range of these secretion systems, their presence is often strain dependent, with some strains lacking entire secretion systems. This is true, for example, with LF82, which lacks the T3SS but has two sets of T6SSs (Costa et al. 2015; Miquel et al. 2010; Nash et al. 2010).

With a lack of any T3SS, two other secretion methods have been suggested as primary routes through which LF82 may deliver effectors to target cells: the T6SS, and outer membrane vesicles (OMVs) (Agus et al. 2014; Tyrer et al. 2014; Miquel et al. 2010; Rolhion et al. 2005). Therefore, due to the association of these secretion mechanisms with LF82, the T3SS, T6SS, and OMV secretion mechanism are discussed below.

1.2.1 The type 3 secretion system

1.2.1.1 Overview

The presence of a type 3 secretion system (T3SS) was first suggested in 1993 by Salmond & Reeves (1993). Since then, evidence suggesting a role in pathogenicity has led to research in determining its structure, substrates and function (Büttner 2012).

T3SSs are found in a multitude of gram-negative bacteria, including highly pathogenic, and purely commensal bacteria (Green & Mecsas 2016). Once fully assembled, the secretion apparatus is often said to resemble a 'syringe' due to its shape and the way in which it delivers secreted proteins to its target. Cargo may be delivered immediately upon contact with the target cell membrane and/or be delivered later in the interaction; this suggests that a single T3SS structure may be utilised for the delivery of multiple effectors (Galán & Collmer 1999). The delivery mechanism is a Sec-independent method which is generally thought to occur via protein delivery directly to the target site, with no intermediate periplasmic step (Green & Mecsas 2016). However, evidence from *Yersinia* suggests that an intermediate periplasmic step may be possible, although more research into this is required (Edgren et al. 2012; Akopyan et al. 2011).

The expression of any given T3SS can depend upon a wide range of environmental stimuli, such as temperature, pH, or the presence of divalent cations (Francis et al. 2002). For example, the T3SS *Edwardsiella ictaluri* has been shown to be expressed in response to low intracellular pH and/or phosphate limitation, while the SPI-1 encoded T3SS of *S. typhimurium* is expressed in response to a low Mg^{+} concentration (Rogge & Thune 2011; Francis et al. 2002).

1.2.1.2 T3SS secreted proteins

Proteins destined for secretion via the T3SS typically have N-terminal secretion signals which remain uncleaved after secretion. Chaperones are often involved in the shuttling of these T3SS substrates to the base of the T3SS (see below), stabilising them until they are secreted into their destination cell (Munera et al. 2010).

Secreted proteins involved in pathogenicity typically fall into two categories: 'translocators' or 'effectors'. Translocators – such as those of the YopB and YopD family of *Yersinia enterocolitica* – form a pore in the target cell membrane, allowing translocation of effector proteins into the target host cell. In contrast, effector proteins – such as the shiga toxin of enterohaemorrhagic *E. coli* – are involved in the disruption of host cell functions (Munera et al. 2010).

Effectors may also function to prevent the creation of a hostile environment within the host cell, as is the case with YopE, YopH, YopO and YopT of *Y. enterocolitica*. These four effectors interfere with the host cytoskeleton and inhibit the production of reactive oxygen species (ROS), thereby increasing the likelihood of bacterial survival (Spinner et al. 2008).

1.2.1.3 T3SS structure

The core components of the T3SS consist of 9 proteins which are highly conserved across the seven predicted families of T3SSs, and a further 10-20 proteins which are important or essential for proper function. These secretion system proteins are often (but not exclusively) found in operons, resulting in pathogenicity islands (PAIs) which may be horizontally transferred between species, especially if present on plasmids (see: Bacterial virulence factors and pathogenicity islands) (Green & Mecsas 2016; Troisfontaines & Cornelis 2005).

The structure of the T3SS consists of three main components: the basal body, the needle component, and the translocon. The basal body spans both the IM and OM, with ring structures (connected by a central rod) anchoring the system in place on both membranes. The needle component protrudes from the OM into the extracellular milieu, and has a hollow core through which unfolded proteins can traverse. These two components are utilised for the secretion of protein into the extracellular space, however, a third and final component is required for entry into a target cell membrane: the translocon.

The needle component's tip complex initiates the formation of the translocon assembly, resulting in a pore in the target membrane, and delivery of effectors into the target host cell (Green & Mecsas 2016). A schematic of the T3SS can be seen in Figure 1.10, in which the system can be seen in a resting and operative state, along with an electron micrograph of the *Yersinia enterocolitica* T3SS needle component.

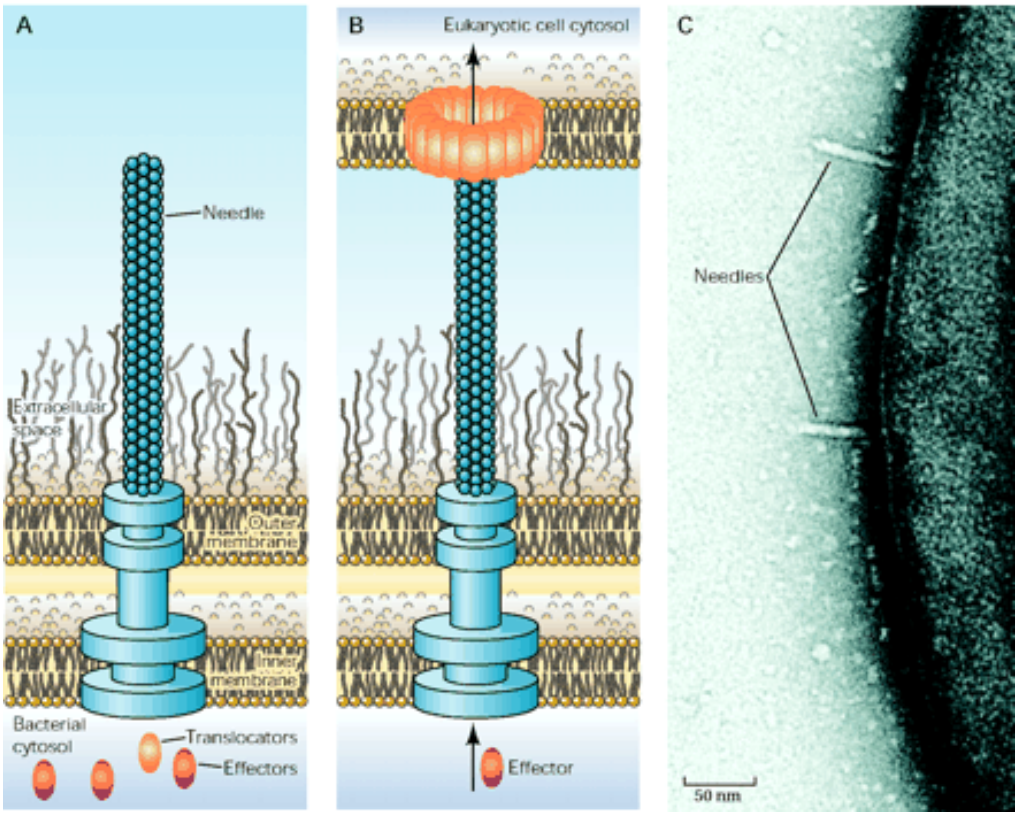


Figure 1.10. **Overview of the *Yersinia enterocolitica* T3SS**. a) The T3SS in its resting state. The apparatus spans both the IM and the OM, and the 'needle' of the apparatus protrudes into the extracellular milieu. B) The T3SS in its operative state. Translocators can be seen forming a pore in the target host cell membrane; effectors are loaded into the basal body machinery, they traverse the needle structure and are secreted directly into the host cell cytoplasm. c) Electron micrograph of the cell surface of *Yersinia enterocolitica* with the 'needle' structures of the T3SS visible protruding from the membrane. Figure adapted from Troisfontaines & Cornelis (2005).

1.2.1.4 Summary

The T3SS offers a convenient and well-studied route for the delivery of effectors into target cells. The lack of a T3SS in LF82 therefore leaves the potential for the use of another secretion system for effector delivery. The presence of two T6SSs may therefore be indicative of a T6SS-based method for effector delivery in LF82.

1.2.2 The type 6 secretion system

1.2.2.1 Overview

The T6SS is a recently discovered bacterial secretion system, and, as such, is not as well studied or understood as the T3SS. The system was near simultaneously identified in *Pseudomonas aeruginosa* – an organism of particular interest due to its major involvement in the morbidity and mortality of cystic fibrosis patients – and *Vibrio cholerae* – the organism responsible for cholera (Mougous et al. 2006; Pukatzki et al. 2006; Govan & Deretic 1996).

The *Ps. aeruginosa* genome encodes machinery for three separate T6SSs: Hcp secretion island-1 T6SS (H1-T6SS), H2-T6SS, and H3-T6SS (Hachani et al. 2016). Hood et al. (2010) have shown that the H1-T6SS is involved in the delivery of a toxin to a target bacterial cell, while Jiang et al. (2014) have shown that both the H2-T6SS and H3-T6SSs are utilised to target both prokaryotic and eukaryotic cells. This ability to affect both prokaryotic and eukaryotic cells, including HeLa, is also observed when concerning the single T6SS found in *V. cholerae*, where secretion of VgrG causes actin polymerisation and can lead to cell death (Altindis et al. 2015; Jiang et al. 2014; MacIntyre et al. 2010; Pukatzki et al. 2009).

1.2.2.2 Distribution and interactions

Since the discovery of the T6SS, the system has been identified in a wide range of gram-negative organisms (~25% of sequenced gram-negative bacterial species), including *Salmonella enterica*, *Serratia marcescens* and *E. coli* (Green & Mecsas 2016). Multiple T6SSs are often present in some organisms, such as in *Burkholderia pseudomallei*, which encodes six distinct T6SSs, accounting for an estimated 2% of its genome (Schwarz et al. 2010).

The T6SS was initially thought to only be involved in interactions between host bacteria and eukaryotic cells, as part of either a pathogenic or symbiotic

relationship (Filloux et al. 2008). However, we now know that the T6SS is also involved in bacteria-bacteria interaction, and has been linked to the density of certain cell populations and the hardiness of individual cells within said populations (Cianfanelli et al. 2015; Jani & Cotter 2010; Schwarz et al. 2010).

1.2.2.3 T6SS structure

The T6SS complex is energetically expensive in comparison to other secretion systems, due to the constant requirement to disassemble and reassemble multiple components of the machinery (Ho et al. 2014). The system contains 13 conserved core elements which are typically located in operons, and can be located on the genome or a plasmid. Interestingly, so-called 'orphan' copies of some core components are sometimes found outside the main T6SS cluster, often in close proximity to effectors secreted by the T6SS (Cianfanelli et al. 2015).

Although the T6SS underwent near-simultaneous discovery in *Ps. aeruginosa* and *V. cholerae*, the nomenclature surrounding the core proteins of the T6SS has generally been dominated by a naming strategy related to *Ps. aeruginosa*. However, since the T6SS is now known to be found in many different gram-negative organisms, a 'Tss' nomenclature has been suggested (Shalom et al. 2007). The 13 core components of the T6SS can therefore be seen in Table 1.2, alongside their 'common' names, along with the 'Tss' nomenclature-accepted name for the PAAR-motif-encoding needle spike (see below). The nomenclature used in reference to each respective component throughout this research can be seen highlighted in bold.

Table 1.2. **Unified 'Tss' nomenclature.** Nomenclature was suggested for all T6SS core components by Shalom et al. (2007). Those used in this research are highlighted in bold. Table adapted from Cianfanelli et al. (2015).

Gene No. in LF82	'Tss' name	'Common' alternative	Location/Role
LF82_014, LF82_016, LF82_439, LF82_446	TssA	-	Baseplate?
LF82_027, LF82_426	**TssB**	VipA	Needle sheath
LF82_026, LF82_427	**TssC**	VipB	Needle sheath
LF82_028, LF82_013, LF82_431	TssD	**Hcp**	Needle
LF82_025, LF82_445	**TssE**	HsoF	Baseplate
LF82_024, LF82_442	TssF	VasA	Baseplate?
LF82_023, LF82_443	TssG	-	Baseplate?
LF82_018, LF82_432	TssH	**ClpV**	Sheath recycling
LF82_029, LF82_033, LF82_433	TssI	**VgrG**	Needle spike
LF82_021, LF82_444	TssJ	Lip, SciN	Membrane complex
LF82_020, LF82_428	TssK	-	Baseplate?
LF82_019, LF82_429	TssL	Icmh/DotU, VasF	Membrane complex
LF82_015, LF82_438	TssM	IcmF, VasK	Membrane complex
N/A	TagD	**PAAR**	Needle spike tip

One of the most impressive features of the T6SS is its long needle-like projection. This consists of a haemolysin co-regulated protein (Hcp/TssD) 'needle' inside a TssBC 'sheath', with a VgrG/PAAR 'spike' (see below). This apparatus alone can reach ~1000 nm in length, and bears structural homology to the tail assembly of bacteriophage T4, as shown in Figure 1.11 (Mikhail M. Shneider et al. 2013; Ho et al. 2014).

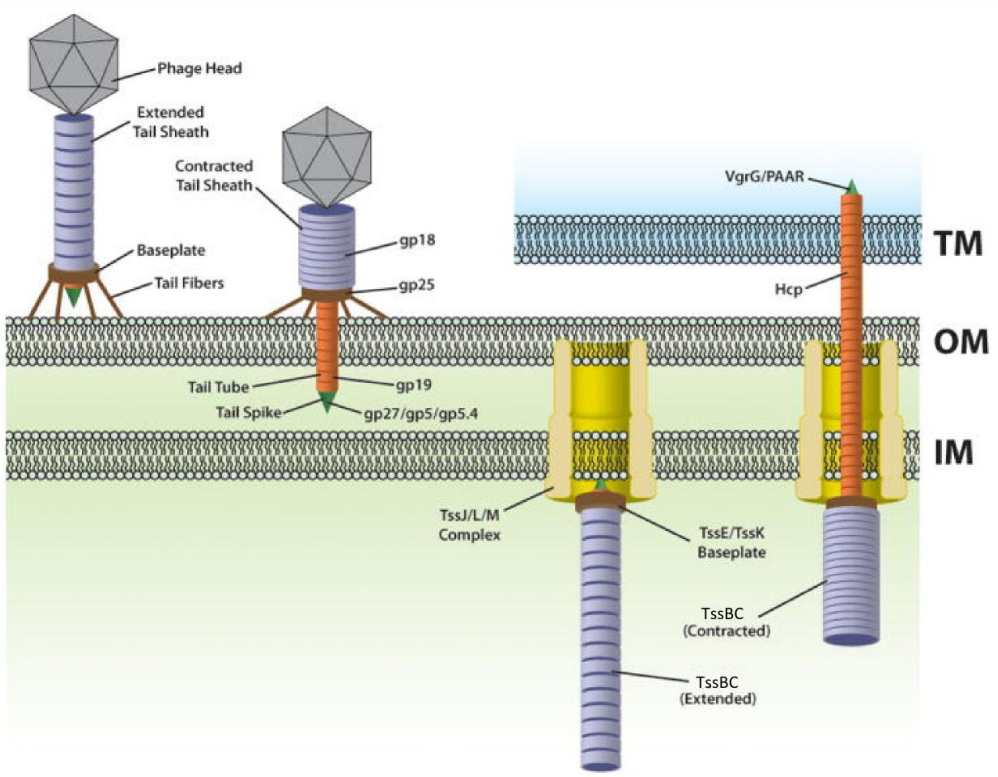


Figure 1.11. **Comparison of the T6SS machinery to that of the T4 bacteriophage**. While the phage utilises its machinery to attach to the OM of its target and inject content inwards, the T6SS can be considered a 'reversed system' where similar components are used to secrete content outwards. Homologous components are depicted in the same colour. Figure adapted from Ho et al. (2014).

Homotrimeric complexes of VgrG form a 'dull' spike, to which a PAAR-motif-containing protein binds, resulting in the addition of a 'sharpened' tip to the VgrG spike. The PAAR-motif can be found within numerous coding sequences – some of which potentially encode effectors – in any one genome, suggesting that the resulting proteins may be secreted via the T6SS (Ho et al. 2014; Mikhail M Shneider et al. 2013). Furthermore, it appears that effectors can be attached to the VgrG spike component, allowing the translocation of a variety of effectors/proteins via the needle tip alone (Mikhail M. Shneider et al. 2013).

The VgrG/PAAR spike is attached to an Hcp needle, consisting of stacked rings of Hcp hexamers. An outer sheath of TssB/TssC surrounds the Hcp needle, allowing the Hcp structure to be pushed outside the cell when the outer sheath undergoes a contraction event. The process of contraction and the needle 'firing' is estimated to occur in ~5 ms, after which a 30-60s disassembly phase occurs, in which the ClpV ATPase disassembles TssBC (Basler et al. 2012). Meanwhile, the Hcp/VgrG/PAAR needle complex punctures the target membrane, delivering any attached PAAR-containing effector/s (see later) (Ho et al. 2014). Figure 1.12 shows an overview of the T6SS complex in its extended, contracted (firing), and disassembly state.

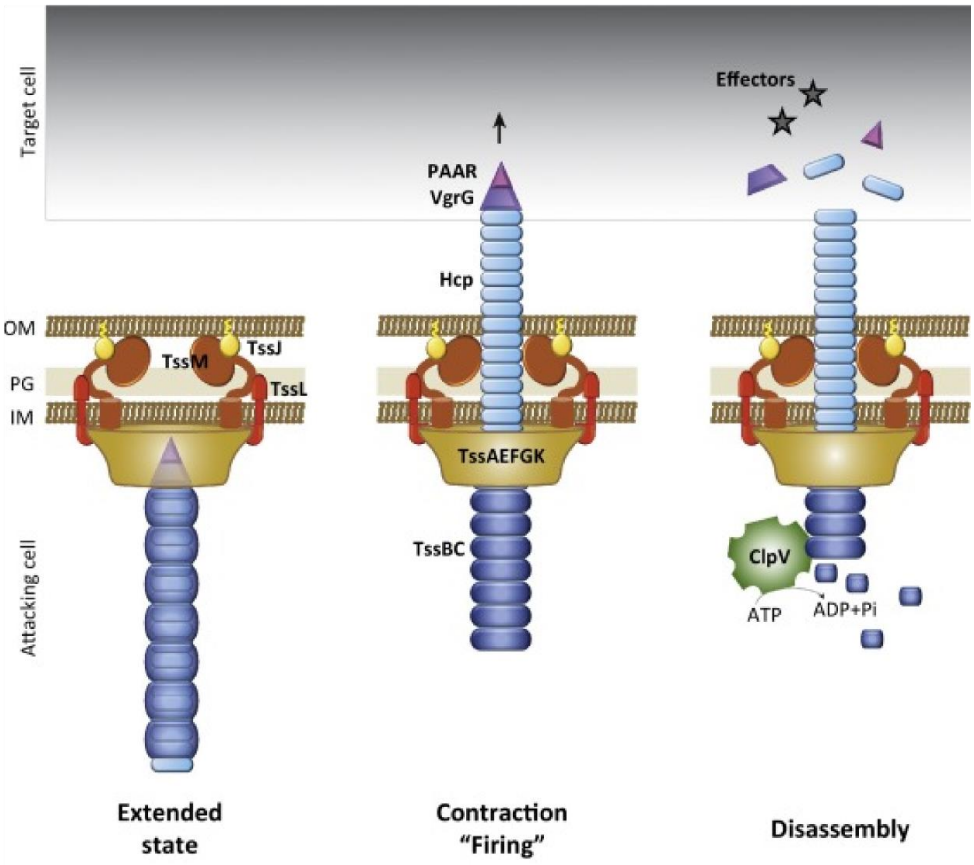


Figure 1.12. **Side-by-side comparison of the T6SS in its extended, contracted (firing), and disassembly states**. The baseplate structure can be seen expanding across both the IM and OM, to which the needle sheath connects. Left: In the extended state the sheath is in its expanded form and extends into the host bacterium's cell cytoplasm with a cylindrical Hcp needle complex inside. Middle: A conformational change leads to the contraction of the TssBC sheath, ejecting the Hcp needle out of the cell and into a target membrane in which effectors are delivered. Right: The disassembly phase then occurs, in which the TssBC sheath is disassembled by ClpV. The T6SS structure must be reassembled before secretion can occur once more. Figure adapted from Cianfanelli et al. (2015). IM: inner membrane; OM: outer membrane; TM: target membrane; PG: peptidoglycan.

1.2.2.4 T6SS effector delivery

Although the Hcp needle is hollow, the space inside appears too small for most effectors to traverse through in their folded states (Ho et al. 2014). The peptidoglycan-hydrolase effector Tse1, produced by *Ps. aeruginosa,* is one of the few examples of effectors small enough to fit inside this space fully assembled (Benz et al. 2012). However, prior to localising to the T6SS structure, some unfolded or semi-unfolded effectors are able to recruit Hcp, and are therefore incorporated into the Hcp needle along with their associated Hcp. This allows their delivery into target host cells where folding is completed, suggesting that Hcp is capable of displaying a chaperone role for some effectors (Ho et al. 2014).

The apparent array of delivery methods seen in the T6SS has led Ho et al. (2014) to suggest that there are five distinct mechanisms of effector/protein delivery: 1) Via modified forms of VgrG, which themselves display effector activity, 2) Via VgrG, to which effectors directly attach, 3) Through incorporation of PAAR-motif-containing effectors into the tip of the spike, 4) Through the attachment of effectors to already bound PAAR-motif-containing proteins, 5) By the binding of effectors to the luminal side of the Hcp needle complex. A schematic of these delivery mechanisms can be seen in Figure 1.13.

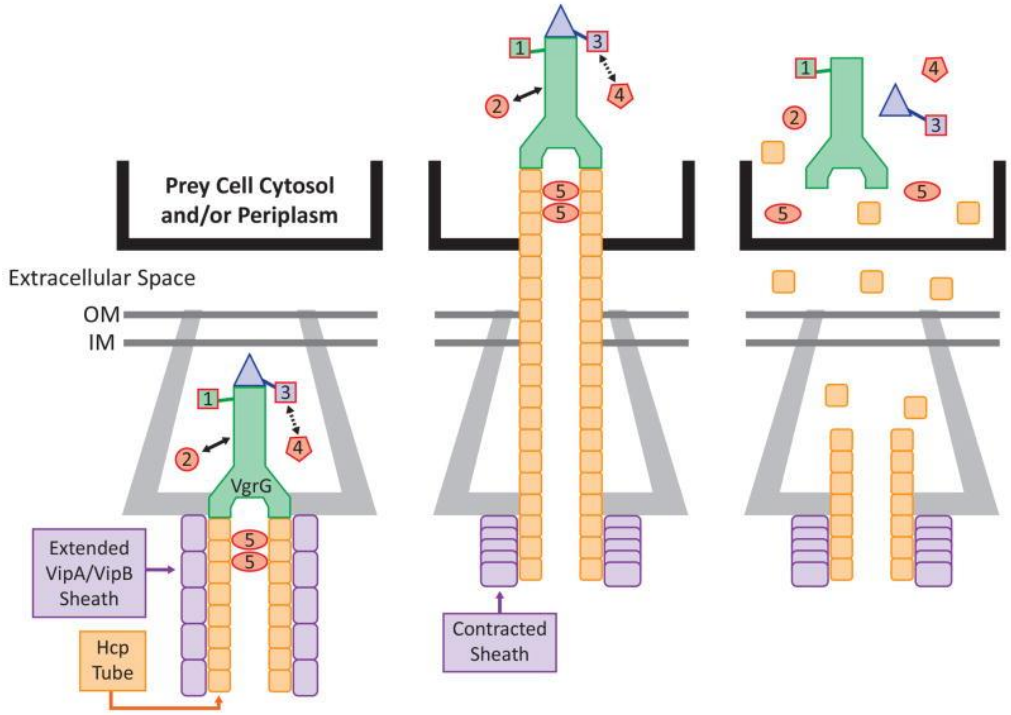


Figure 1.13. **A schematic of the delivery of effectors by the five known/predicted mechanisms of delivery used by the T6SS**. The schematic shows effector binding locations (1-5), and the T6SS in an expanded state (left), contracted/fired state (middle), and disassembly state (right). Delivery may occur via 1) Modified forms of VgrG which themselves display effector activity, 2) Effectors which attach directly to VgrG, 3) PAAR-motif-containing effectors which form the tip of the spike, 4) Effectors which attach to already bound PAAR-motif-containing proteins, 5) Effectors which bind the luminal side of the Hcp needle complex in an unfolded or semi-unfolded state. Figure modified from Ho et al. (2014).

1.2.2.5 Disassembly/recycling

Once the T6SS has successfully 'fired' and delivered its cargo to the target cell, disassembly occurs. At this stage, ClpV is recruited and begins disassembly of the TssBC outer sheath. It is not clear whether the Hcp needle is semi-retracted and disassembled intracellularly, or whether disassembly occurs outside the cell (Ho et al. 2014).

1.2.2.6 Summary

The wide-ranging distribution of the T6SS amongst gram-negative bacteria, and the apparent versatility with which effector delivery can occur using this system, make it an interesting candidate for effector delivery in the absence of the T3SS. While it is not yet known whether LF82 utilises either of its T6SSs for the purposes of effector delivery, preliminary evidence suggests that their disablement may affect the invasive ability of LF82, with T6SS knockouts showing a negative-invasion phenotype (Massier et al. 2015). This suggests that the T6SSs of LF82 may play a role in eukaryotic cell invasion – although to what extent is currently unclear.

1.2.3 Outer membrane vesicles

Outer membrane vesicles (OMVs) allow the delivery of a variety of bacterial proteins to other cells. Unlike dedicated secretion systems such as those described above, OMVs can deliver cargo containing a range of virulence factors directly to the cell cytoplasm of target cells over long distances, as well as over short distances. Bomberger et al. (2014) and Bomberger et al. (2009) have shown this with *Ps. aeruginosa*, which is able to deliver OMVs containing a cocktail of β-lactamase, haemolytic phospholipase c, and Cif – an effector which hinders ABC transporter function and expression – in conjunction with other virulence factors over long distances using OMVs.

The use of OMVs for transport of proteins over long distances has been compared to the use of exosomes by eukaryotic cells (Bomberger et al. 2009). Much like OMVs, the primary function of exosomes remains to be confirmed (see below), but evidence suggests their involvement in shuttling a wide range of material, such as tumour cell antigens, and even HIV particles (Bomberger et al. 2014; Schorey & Bhatnagar 2008; Gould et al. 2003). However, despite the comparison of OMVs to exosomes, it should be noted that all domains of life appear to produce some form of membrane bound vesicles (Schwechheimer & Kuehn 2015).

1.2.3.1 Structure and cargo

OMVs are released by both pathogenic and non-pathogenic bacteria. They are derived from the outer membrane (OM) of gram-negative bacteria and therefore consist primarily of phospholipids and lipopolysaccharides, along with an assortment of OM proteins (Roier et al. 2016; Ellis & Kuehn 2010; Bomberger et al. 2009). OMVs can vary in size between 10-300 nm in diameter, and cargo can consist of IM components, cytoplasmic or periplasmic proteins, DNA or RNA, and signalling molecules. These components may or may not be purposefully included in OMV cargo, and may inadvertently be packaged during OMV production (Roier et al. 2016). However, some proteins do appear to be selectively enriched in OMVs, as is the case with the heat labile toxin of enterotoxigenic *E. coli* (ETEC) (Bomberger et al. 2009).

1.2.3.2 Function and biogenesis

Although the biogenesis and primary function of OMVs remains ill-defined, it is known that OMVs originate via budding from the OM of bacteria, as shown in Figure 1.14. OMVs are involved in a wide range of functions, including the transport of cargo (such as toxins), DNA involved in horizontal gene transfer and biofilm production (Roier et al. 2016). OMVs have even been suggested to act as 'decoy targets' for antibiotics and phage, where they absorb some of the antibiotic and/or phage attach, although whether this is intentional or not remains to be seen (Schwechheimer & Kuehn 2015). Some examples of these functions can be seen in Figure 1.15.

There are 3 separate models proposed to explain the formation of OMVs. However, these models either require genetic modification of the studied organism, the introduction of stress upon the bacterium, or have only been demonstrated in one species (*Ps. aeruginosa*) (Roier et al. 2016). Roier et al. (2016) have recently described a new, more generalised model for OMV biogenesis, however, research into this area is still ongoing.

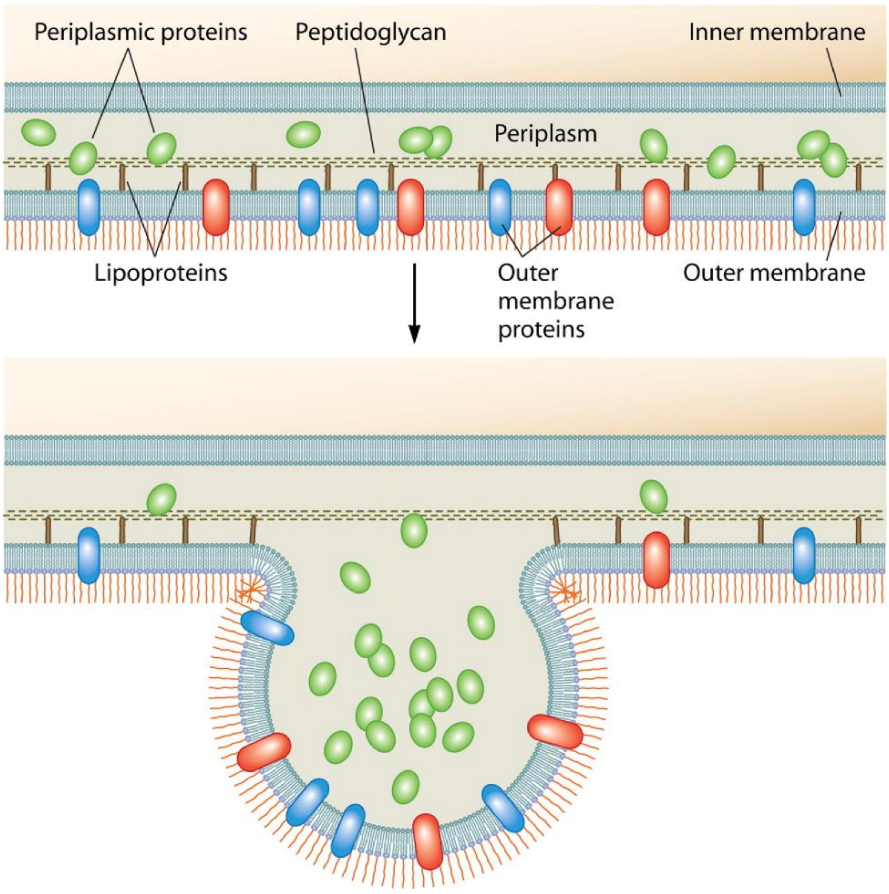


Figure 1.14. **Schematic diagram of the formation of an OMV from the bacterial OM**. OMVs are known to originate from the OM of gram-negative bacteria by budding. Their constituent components and contents therefore consist of components present in the OM and periplasmic space, as well as contents of the IM and cytoplasm. Figure adapted from Ellis & Kuehn (2010).

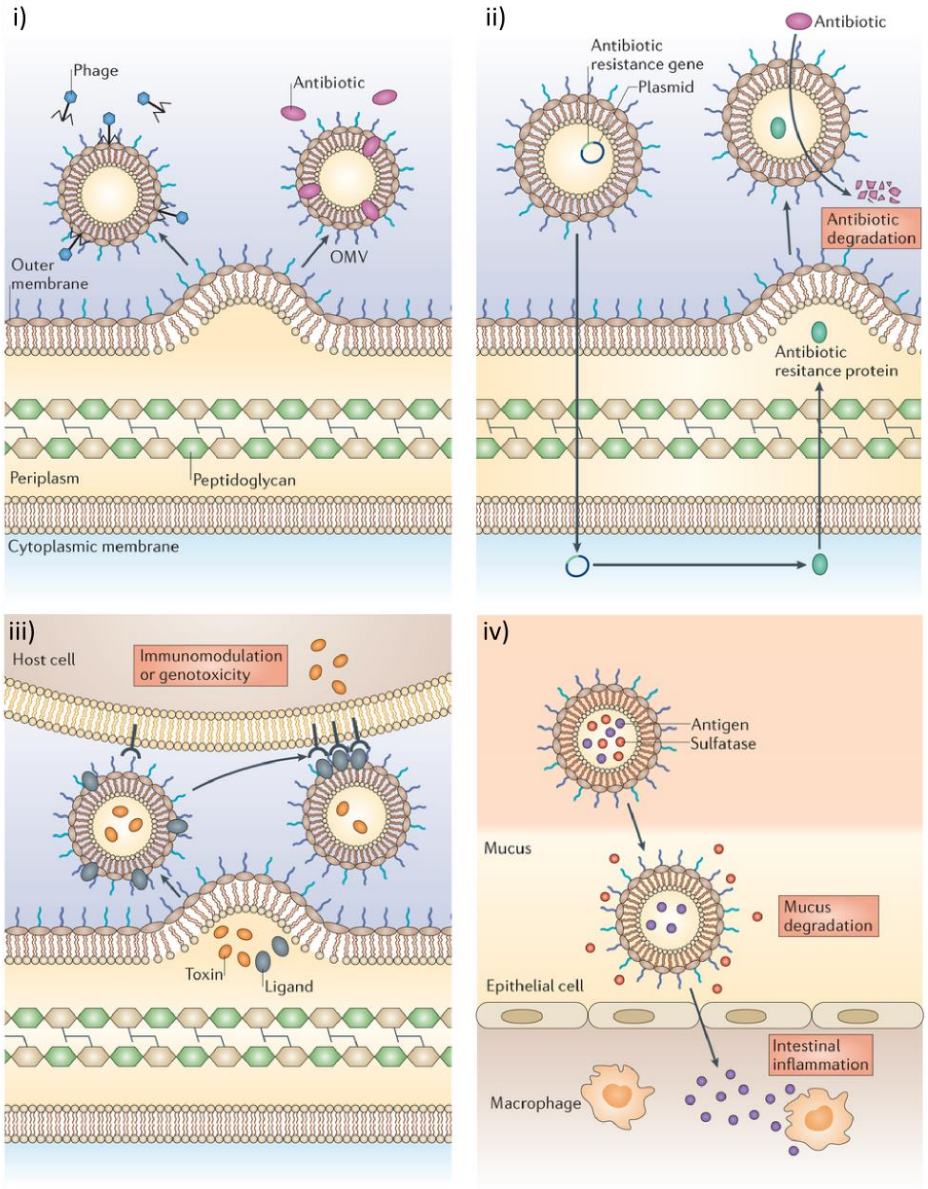


Figure 1.15. **Several known functions of OMVs**. Multiple roles have been suggested for OMVs, including i) OMVs acting as 'decoy targets' for phage and antibiotics; ii) OMVs containing DNA, leading to horizontal DNA transfer upon fusion with a target cell; iii) simultaneous delivery of multiple virulence factors to target cells; iv) OMVs penetrating the mucosal barrier, delivering virulence factors to cells with which the bacterium cannot make direct contact. Figure adapted from Schwechheimer & Kuehn (2015).

1.2.3.3 Comparison to T3SS and T6SS

The delivery of virulence factors via OMVs is thought to have several advantages over other delivery mechanisms. For example, secretion of free, soluble protein into the extracellular milieu may result in its degradation by digestive enzymes, while cargo within OMVs is largely protected against these external factors, allowing their undamaged delivery (Bomberger et al. 2009). Furthermore, the delivery of virulence factors via a dedicated secretion system may not allow simultaneous delivery of all necessary effectors, whereas fusion of an OMV to a target cell can allow direct and simultaneous delivery of multiple virulence factors into a host cell in a coordinated manner, as is the case with *Ps. aeruginosa* (Kunsmann et al. 2015; Bomberger et al. 2009).

1.2.3.4 Examples of OMVs for the delivery of virulence factors

The use of OMVs for the delivery of virulence factors has been documented for several pathogenic bacteria, such as *Salmonella* and *E. coli*.

Salmonella Typhimurium has been shown to utilise OMVs for the delivery of several virulence factors into mouse macrophages, without which bacterial fitness is reduced (Yoon et al. 2011). An outbreak strain of *E. coli* (O104:H4) has also been shown to secrete a blend of enteroaggregative *E. coli* and enterohaemorrhagic *E. coli*-associated virulence factors - including shiga toxin – using OMVs (Kunsmann et al. 2015).

While LF82 lacks any T3SSs, and has two T6SSs which have been shown to be associated with eukaryotic cell invasion, OMVs may also be utilised to deliver effectors to target cells (Massier et al. 2015). Rolhion et al. (2005) have shown that LF82 Δ*yfgL* shows a reduction in OMV production, and also a decrease in invasive ability of Intestine-407 cells. Furthermore, Tyrer et al. (2014) have shown that LF82 OMVs alone can result in DNA damage to Caco-2 cells. While it is not clear what role, if any, DNA damage plays in the

establishment of LF82 within a host cell, this shows that LF82 can utilise OMVs for the delivery of effectors to target cells.

1.3 Bacterial virulence factors and pathogenicity islands

1.3.1 Identification and purpose

Since the genome of *Haemophilus influenzae* was sequenced in 1995, sequencing technologies have seen an exponential increase in speed, and a decrease in cost, with the cost per megabase in 2015 reportedly as low as $0.014 (Wetterstrand 2015). This has led to easier access and an increased use of sequencing technologies for a range of organisms, including bacteria. While this technological advancement has led to the production of vast amounts of data, it has also allowed the comparison of multiple bacterial genomes and the identification of common features between them.

The so-called 'core-genome' is a set of the most basic and essential genes for bacterial survival under rudimentary conditions, and attempts are still underway to properly define it (van Tonder et al. 2014). Virulence factors are often found in addition to the core-genome, and encode components that contribute to an organism's disease-causing ability or their survival in a host (Ho Sui et al. 2009).

Interestingly, these virulence factors often differ from the core-genome in their percentage GC and/or codon usage (Hacker & Kaper 2000). This can be explained by their acquisition from an alternative source – such as another bacterium or virus which has a differing GC content and/or shows preference for a particular codon set – especially through horizontal gene transfer.

In the 1980s, it was noticed that certain regions of the uropathogenic *E. coli* (UPEC) genome tended to vary more than others when compared to other *E. coli* strains (Low et al. 1984). These regions often contained genes encoding

virulence factors – such as haemolysins, and fimbriae – and were named pathogenicity islands (PAIs) (Hacker et al. 1990).

1.3.1.1 Characteristics of PAIs

Virulence factors as well as CRISPR targeting sequences display a disproportionate association with PAIs (Ho Sui et al. 2009). Mobile genetic elements, such as transposons, are often responsible for the presence of PAIs, as they allow insertion or modification of genetic material into a particular genomic site with relative ease (Schmidt & Hensel 2004). In fact, the status of a bacterial genome at any one time can be considered to be in constant flux, with whole genome sequencing showing variations within species, and even within clinical isolates (Schmidt & Hensel 2004).

PAIs are seen in many different organisms, and are often absent in non-pathogenic relatives of any one particular strain. Despite this, PAIs can be maintained within some strains, even though they are not necessarily used for causing virulence. In these cases, PAIs are known as 'fitness islands' and are utilised for a fitness advantage rather than for virulence (Hacker & Kaper 2000).

Virulence factors often encode features such as components of secretion systems (T3SS and T6SS), invasins, toxins, and effectors (Hacker & Kaper 2000). These features may serve to facilitate pathogenicity, or may be entirely responsible for some aspects of disease. For example, *Salmonella* pathogenicity island 2 (SPI-2) (discussed below), which encodes a type 3 secretion system in *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium), is essential for the establishment of systemic infection and bacterial survival within mouse macrophages (Klein & Jones 2001). *Edwardsiella ictaluri* also utilises a PAI-based T3SS, which is essential for intracellular replication in catfish macrophages and catfish ovary cells (Dubytska et al. 2016).

Wang et al. (2010) have discussed the difficulties of identifying PAIs, with focus upon those which do not display a particularly variant GC content and/or codon bias compared to the rest of the genome. These PAIs may have been acquired long ago, and have since seen alterations to better suit the host CG content and/or codon usage. Lawrence & Ochman (1997) have discussed how these alterations work, and suggest that inherited genes, once resident in a new host, are subjected to the same mutations as the native genes, therefore resulting in progressive mutations which lead to alignment with the 'normal' GC content of the host. The work of Wang et al. (2010) therefore suggests that PAIs may exist which are not detectable via current methods of GC content analysis due to their amelioration to the new host % GC content– including in LF82.

1.3.2 Key examples of bacterial virulence factors and PAIs

There are numerous examples of occasions where the presence of virulence factors alone does not result in pathogenicity. This has been shown, for example, with *Neisseria*, where 67% of the virulence factors found in the pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae* were shown to also be present in the non-pathogenic relative, *Neisseria lactamica* (Snyder & Saunders 2006). In contrast, the expression of a single virulence factor alone – such as botulinum toxin from *Clostridium botulinum* – can be sufficient to cause disease (Ho Sui et al. 2009).

Several pathogenic bacteria have been the subject of much research to determine their method/s of pathogenicity, as well as their underlying virulence factors.

1.3.2.1 *Salmonella* Typhimurium

As mentioned previously, SPI-2 of *S.* Typhimurium is responsible for bacterial intracellular survival within mouse macrophages – a critical stage for systemic

infection in mice (Figueira & Holden 2012; Klein & Jones 2001). Both PAIs present in *S.* Typhimurium encode T3SSs, with SPI-1 becoming activated during the establishment of bacteria in the gut, and SPI-2 activated after intracellular uptake into a *Salmonella* containing vacuole (SCV) (Figueira & Holden 2012).

Loss of components of the SPI-2 encoded T3SS typically result in a virulence defect, usually presenting as a reduced ability for bacteria to survive and grow intracellularly (Figueira & Holden 2012; Salcedo et al. 2001). As mentioned previously, there are over 28 effectors associated with SPI-2 and involved in the modification of the SCV to prevent bacterial degradation (Figueira & Holden 2012; Haraga et al. 2008).

1.3.2.2 Enterohaemorrhagic *E. coli* O157:H7

Enterohaemorrhagic *E. coli* (EHEC) O157:H7 (hereby referred to as O157) is a food-borne pathogen capable of causing severe disease in humans (Paton & Paton 1998). It is likely that this organism originated from a non-pathogenic ancestor that obtained virulence genes via the acquisition of numerous PAIs through horizontal gene transfer (Wang et al. 2010).

All EHEC members contain a PAI which encodes a T3SS along with its associated effectors, as well as intimin – an adhesin which promotes adhesion of the bacterium to the target cell surface. This PAI enables EHEC strains to cause attaching and effacing lesions to the intestinal brush border, and it is therefore known as the locus of enterocyte effacement (LEE) (Morabito et al. 2003). Most notably, many EHEC variants of the shiga-toxin producing *E. coli* (STEC) – of which O157 is one – appear to maintain their pathogenicity while lacking intimin. However, O157 does not appear to be able to colonise intestinal epithelial cells without this virulence factor, severely reducing its disease-causing ability (Mohawk & O'Brien 2011).

1.3.2.3 Adherent invasive *E. coli* LF82

While often regarded as a pathogen in the literature, LF82 is a pathobiont, meaning it can remain as a commensal organism in some hosts, and cause disease in others (Tyrer et al. 2014). The organism has been associated with CD in particular, although this association is still unclear.

Miquel et al. (2010) report the LF82 genome to be of 4.88 Mbp in size and to contain 4376 CDSs. While the number of CDSs present in the LF82 genome is low compared to other pathogenic *E. coli*, this number is most similar to that of APEC_O1 (4428 CDSs), supporting the suggestion that APEC is the most similar *E. coli* pathovar to LF82 (see Table 3.2 later) (Miquel et al. 2010).

The genome of LF82 contains four predicted pathogenicity islands (see Figure 1.17), of which two – PAI-1 and PAI-3 – encode T6SSs. As mentioned previously (see: The type 6 secretion system: Summary), preliminary research by Massier et al. (2015) suggests that deletion of these T6SSs affects invasive ability. They have shown that deletion of either (or both) T6SSs results in a decreased ability for successful LF82 adhesion to and invasion of T84 cells. However, unlike a PAI-1-based T6SS deletion, or deletion of both T6SSs, removal of the PAI-3-based T6SS alone does not lead to a bacterial motility defect, and also does not lead to a reduced inflammatory response in a mouse model.

Meanwhile, LF82 PAI-2 shows significant homology to the 'high-pathogenicity island' of *Yersinia* – a PAI known to encode and regulate the expression of the yersiniabactin siderophore – and encodes 3 additional putative adhesins and invasins (see Figure 1.16) (Miquel et al. 2010). However, more research is necessary to determine the function of the PAIs identified in LF82, along with their associated effectors.

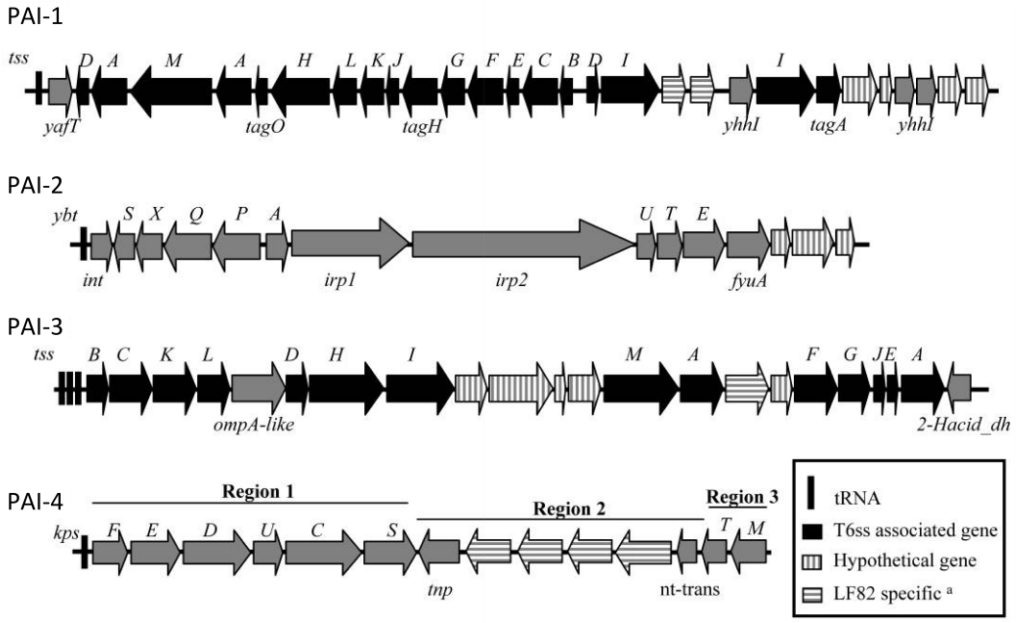


Figure 1.16. **The four predicated PAIs of LF82**. PAI-1 and PAI-3 encode T6SS-associated genes (black). PAI-2 shows similarity to the *Yersinia* high pathogenicity island, with three putative adhesins and invasins (vertical lines). PAI-4 encodes several transposase-flanked LF82 specific CDSs, and is similar to that found in UPEC 536. Figure adapted from Miquel et al. (2010).

With LF82 lacking any T3SS, it is likely that the bacterium has evolved new strategies for the delivery of virulence factors into target cells. The presence of two T6SSs may therefore suggest that LF82 uses one or both of these in lieu of a T3SS. For example, like *S. typhimurium,* members of the AIEC are known to remodel the host cell cytoskeleton during invasion. However, while *S. typhimurium* utilises its T3SSs for this process, LF82 has no such system and may therefore show evidence for the use of an individual or both of its T6SSs for actin remodelling and invasion (Ramos-Morales & Ramos-Morales 2012; Miquel et al. 2010). The use of a T6SS for this purpose has been shown in other organisms, such as *V. cholerae,* where VgrG (the spike component of the T6SS) is utilised for actin cross-linking, which contributes to host cell death (Ma & Mekalanos 2010).

1.3.2.4 Summary

As displayed here, the identification of bacterial virulence factors and their associated pathogenicity islands – if present – is essential for the

understanding of disease progression and causality. Research to reveal potential new virulence factors in LF82 is therefore necessary to support current knowledge, and to expand upon the methods via which LF82 circumvents the immune system and establishes itself as a known-associate of CD. However, with LF82 containing ~400 unstudied hypothetical CDSs alone, a screening system may be of use to help search for putative effectors.

1.4 High-content screens

High-content screening (HCS) was first introduced in 1996, and is an automated method for the quantitative analysis of fluorescence microscopy images on a cell-by-cell basis (Haney et al. 2006; Taylor 2007). Using HCS, hundreds, thousands or even millions of sample images can be evaluated in a quantitative manner, allowing the categorisation of cells into desired subclasses based upon one or multiple phenotype/s.

1.4.1.1 Fluorescence microscopy in high-content screening

Fluorescence microscopy is a convenient technique for the observation of the effects of genetic manipulation upon intracellular events (Zanella et al. 2010). While this method has been in use since the beginning of the 20th century, only recently (since the mid-1990s) has it seen combination with higher-content techniques (Zanella et al. 2010; Amos 1993).

Breakthroughs in the development of fluorescent probes and fluorophores have offered an ever-expanding range of fluorescent markers, with which subcellular effects can be visualised. However, not until auto-focusing, sample positioning, and downstream image analysis techniques were optimised, did true HCS become possible (Zanella et al. 2010; Taylor 2007).

1.4.1.2 Uses of high-content screening

The convenience of HCS was initially exploited by academia and later by the pharmaceutical industry (Taylor 2007). Developments in HCS were driven by the requirement to screen large amounts of data for academic research, with subsequent advancements resulting in the development of high-throughput screening (HTS), which was adopted readily by industry due to its potential for use in drug discovery (Taylor 2007; Haney et al. 2006).

Since then, HCS has been used for a range of research applications, including: the identification of genes involved in *Mycobacterium tuberculosis* phagosome remodelling in macrophages via fluorescently staining acidic intracellular compartments (Brodin et al. 2010); the screening of chemical compounds which disrupt the ability of *M. tuberculosis* to replicate in macrophages (Christophe et al. 2009); and the production of a HCS for the analysis of the effects of T3SS and T6SS knockouts on the ability of *Burkholderia pseudomallei* to produce multinucleated giant cells via cell-to-cell fusion (Pegoraro et al. 2014).

Other groups have used HCSs for the quantification of a range of phenotypic changes in eukaryotic cells when infected with bacteria. For example, Martinez et al. (2015) have described a method for the infection of vero cells with a transposon mutant library of *Coxiella burnetii*, in which they use automated image acquisition coupled with image analysis software to quantify host cell morphology. Studies such as these have laid the groundwork for similar phenotypic screens, in which individual genes may be expressed in a host organism, and the subsequent effects upon phenotype analysed.

1.4.1.3 Advantages of HCSs

The ability of HCSs to process, and analyse thousands, or millions of subsequent images is something which humans are also capable of achieving.

However, where humans tire, have a limited capacity to discriminate small differences, and may struggle to record large volumes of information, computers excel. Computers do not require sleep or coffee breaks, can discriminate small differences based upon exact measurements (within their limits of detection), and can record vast amounts of data with ease (Wollman & Stuurman 2007).

A significant amount of time and resources may be required to train, test and optimise a computer-based analysis system. It may therefore be argued that a human-based analysis approach may be suited to smaller screens in an effort to save time and resources. However, not only do human analysers require training – and therefore have their own associated time and cost deficit – but it has also been shown that using a computational approach can result in the identification of phenotypes otherwise overlooked by human image analysts. For example, Huang & Murphy (2004) have demonstrated this by comparing the results of multiple automated image analysis techniques, and displaying their ability to differentiate the localisation of two Golgi proteins (giantin and Gpp130) ~82-96% of the time (depending upon imaging technique). Meanwhile, humans have been shown to have great difficulty in distinguishing the localisation of these proteins from one another (Murphy et al. 2003).

In addition to these features, computational-based approaches are more reproducible (Jones et al. 2006; Wollman & Stuurman 2007). This is not only useful for reproduction and validation within and between sets, but also for the ability to share a particular image analysis pipeline between different experimental set ups, and to share pipelines with other research groups.

A further benefit of computational analysis of screen data is their unbiased nature. In practice this means that the same rules are applied to each image in any given image set, resulting in the equal analysis of all images. This also allows the results obtained from an automated screen to have a high level of

reproducibility compared to those analysed by humans. In contrast, humans are susceptible to bias when analysing data – particularly data in large volumes (Wollman & Stuurman 2007).

Finally, Macarron et al. (2011) offer an insight into the often-quoted 'drawbacks' of using HCSs. These include the claims that HCS data is of poor quality, HCSs are expensive and time consuming, and that HCSs fail to identify many true hits. However, these drawback are largely due to a misunderstanding of the methods involved. HCSs generally see a higher level of control and quality assurance methods applied compared to low-content methods. The initial outlay in terms of cost and time may be significant, but, when considering the number of targets screened, it is often better value per sample than low-content screens. And finally, while the rate of successful identification of hits using HCSs is typically taken to be between 50-60%, this does not differ from more traditional techniques, which may also be successful or unsuccessful (Macarron et al. 2011).

1.4.1.4 Disadvantages of HCSs

The analysis of screen image data using an automated computer-based approach in HCS offers many advantages. However, unlike humans, computers are often unable to detect patterns which humans find extremely easy to identify, and while computer-based algorithms and machine learning are improving, this issue has still not been resolved entirely (Murphy et al. 2003).

Other disadvantages of using these systems – such as the high costs of initial setup of high capacity storage, computational power, computer training, human interaction and networking speeds for the transfer of images – have largely been solved over the last two decades (Taylor 2007; Wollman & Stuurman 2007). However, Kang et al. (2016) note that these limitations still exist, depending upon the desired number of items to be screened. For

example, the screening of hundreds of thousands of compounds is still likely an undertaking too expensive for an individual lab to perform. Despite this, developments in HCS are ongoing and are progressively allowing for larger libraries to be created and screened on lower budgets (Michiels et al. 2002).

While HCSs require a large initial investment, the number of samples screened often results in a lower cost per hit than when using other techniques (Macarron et al. 2011). High-content screening therefore offers a cost effective method for the assessment of multiple candidate drugs/proteins/samples, despite its potential disadvantages.

1.5 Aims

The overall aim of this research is to enhance the current understanding of the Crohn's disease-associated bacterium, adherent invasive *E. coli* (AIEC). This will be achieved by the development of a high-content screen for the identification and subsequent evaluation of putative effectors from the AIEC type-strain, LF82. For example, CDSs seen to affect autophagic flux, alter host cell survival or organelle/cell morphology, and to colocalise with key subcellular or cytoskeletal components of host cells (such as actin, nuclei or autophagosomes) will be of particular interest due the known involvement of these phenotypes in cellular adherence and invasion.

Firstly, genes unique to LF82 will be identified from a small pool of both pathogenic and commensal *E. coli* genomes, thus removing known effectors and the *E. coli* 'core-genome' from the analysis set. This approach aims to isolate putative novel virulence factors from LF82, while maintaining a large enough pool of coding sequences (CDSs) to aid and evaluate the development of a high-content screen (HCS). A reporter cell line will be created, expressing an mCherry-LC3 fusion for the evaluation and detection of autophagy. Individual CDSs will be fused to an mEmerald GFP fluorescent marker, before their expression in the HeLa-mCherry-LC3 reporter cell line. An automated image analysis pipeline will then be designed, allowing the screening of cell images for the effects of individual CDS expression. The use of such a pipeline will allow quantitative data to be obtained on a cell-by-cell basis. This data will be analysed via the creation of multiple screen-analysis graphs, allowing the comparison of multiple CDSs and the identification of significant hits from the screen data.

Selected hits from the screen will be taken forward to begin phenotype confirmation and assessment of effector status. This will be performed via the deletion of each corresponding hit CDS from the LF82 genome, and subsequent evaluation of the ability of each deletion mutant to invade

eukaryotic cells. Further assays will be developed for the preliminary assessment of CDS co-localisation, and also possibly secretion.

While efforts have been made to ensure that the screen is optimised and will be successful, it is possible that no phenotypes of interest will be detected. If it appears that this is due to low levels of transfection, the screen may have to be performed using midiprep DNA with a reduced number of LF82 unique genes. If transfection is successful but there are no phenotypes of interest, then focus will shift to the effects of T6SS knockouts on LF82 invasion. Even with a successful screen, it is likely that some effectors may remain undetected due to low transfection rates or cell death. This is a known limitation of this strategy and is difficult to control for without the use of more sensitive detection techniques which are not possible in a high-content setting.

Chapter 2: Materials and methods

2.1 Materials

2.1.1 Tables

2.1.1.1 Bacterial strains

Table 2.1. ***E. coli*** **strains**

Strain	Description	Purpose	Source or reference
AIEC LF82	Adherent Invasive *E. coli* NC_011993. Type strain, genome sequenced.	Type strain used for gene deletions and invasion assays	Miquel et al. 2010
LF82 Δ*p314*	LF82 strain with 428 bp of the LF82_314 (469 bp) CDS removed	Invasion assays to investigate role of p314 upon invasion	This study
LF82 Δ*TssE1*	LF82 strain with 341 bp of the LF82_025 (440 bp) CDS removed	Invasion assays to investigate role of TssE1 upon invasion	This study
LF82 Δ*TssE3*	LF82 strain with 396 bp of the LF82_445 (429 bp) CDS removed	Invasion assays to investigate role of TssE3 upon invasion	This study
LF82 Δ*TssE1*-Δ*TssE3*	LF82 strain with 341 bp of the LF82_025 (440 bp) CDS and LF82_445 (429 bp) CDS removed	Invasion assays to investigate role of TssE1 and TssE3 upon invasion	This study
PPY	DH10B derivative: F⁻ *end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*LΔ*lac*X74 *Φ80lac*ZΔM15*ara*D139Δ(*ara,leu*)7697 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *cyn*X*::* [*ara*C pBAD- *redα* EM7- *redβ* Tn5-*gam*] *λ*⁻	Production of PPY extract for SLiCE	Zhang et al. 2012
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIq ZΔM15 Tn10 (Tet^r)]*	General workhorse for cloning. Negative control for invasion assays.	Lab strain

2.1.1.2 Mammalian cell lines

Table 2.2. **Mammalian cell lines**

Strain	Description	Purpose	Notes
HeLa CCL2	Homo sapiens epithelial adenocarcinoma cells. Adherent.	Workhorse for transfections and invasion assays.	American type culture collection (ATCC)
HeLa-mCherry-LC3	HeLa CCL2 transduced cell line which has mCherry-LC3 integrated into the genome.	Transfections during screen	This study
Hek 293T	*Homo sapiens* epithelial kidney cells. Adherent. Constitutively express the simian virus 40 (SV40) large T antigen.	Transfected with lentiviral vectors	ATCC
Hek 293T + psPAX + pMD2.G + pWPXLD-mCherry-LC3	Hek 293T transfected cell line expressing lentiviral particles for the integration of mCherry-LC3 into target cells via lentiviral transduction.	Lentiviral transduction	This study
Caco-2	*Homo sapiens* epithelial colorectal adenocarcinoma cells. Adherent.	Invasion assays	Received as a kind gift from Dr. Chris Hale, Wellcome Trust Sanger Institute, Cambridge, UK

2.1.1.3 Plasmids

Table 2.3. **Plasmids.**

Plasmid	Description
pDoc-K	Used for GeneDoctoring for amplification of the Kan cassette with homologous flanking regions, in conjunction with pDoc-C
pDoc-C	Used for GeneDoctoring to carry Kan cassette and regions of homology to desired site deletion into host strain. Carries *SacB* (levansucrase) to produce levan from sucrose, causing cell toxicity.
pACBSCE	Used for GeneDoctoring. Carries I-SceI site. Induction of encoded I-SceI and λ Red genes via L-arabinose results in 'self-cure' of vector.
pBAD18-Kan	Arabinose inducible bacterial cloning vector, Kan^R.
pBAD18-Kan-p314	pBAD18-Kan overexpressing LF82_314 under arabinose control
pBAD18-Chl	Modified version of pBAD18-Kan, with chloramphenicol resistance
pBAD18-Chl-p314	Modified pBAD18-Kan, for the overexpression of LF82_314 under arabinose control
pBAD18-p152-GSK	PBAD18 vector with p152 CDS-GSK tag fusion under arabinose inducible control.

Plasmid	Description
pCMV-mEm-4GS	Vector used for SLiCE reactions. EcoRV-HF digested and used in SLiCE reactions to make library DNA-GFP fusions.
pCMV-3xFLAG-4GS	Vector used for the expression of the FLAG tag in negative controls for initial autophagosome detection testing.
pCMV-3xFLAG-4GS-p152	Vector used for the expression of FLAG-tagged p152 in initial autophagosome detection testing.
pCMV-GFP-p62	Vector used for the express of p62 controls in the screen
pCMV-3xHA-SseD	Vector used for the express of SseD controls in the screen
pWPXLd	Lentiviral toolkit vector: transfer plasmid.
pKD-sgRNA-p15	Temperature sensitive, arabinose inducible λ red, anhydrotetracycline inducible sgRNA expression that targets pCas9-cr4 plasmid. (Reisch & Prather 2015).
pKD-sgRNA-ack	Temperature sensitive, arabinose inducible λ red, anhydrotetracycline inducible sgRNA expression that targets *ack*. (Reisch & Prather 2015).
pKD-sgRNA-p314	Derived from pKD-sgRNA-ack – targets *p314.* Made using CPEC. Temperature sensitive.

Plasmid	Description
pKD-sgRNA-TssE-1	Derived from pKD-sgRNA-ack – targets *tssE* on pathogenicity island I. Made using CPEC. Temperature sensitive
pKD-sgRNA-TssE-3	Derived from pKD-sgRNA-ack – targets *tssE* on pathogenicity island III. Made using CPEC. Temperature sensitive
pCas9-cr4	Anhydrotetracycline inducible Cas9 nuclease with ssrA tag and constitutive tetR. (Reisch & Prather 2015).
psPAX2	Lentiviral toolkit vector: expresses lentiviral packaging genes. psPAX2 was a gift from Didier Trono (Addgene plasmid #12260)
pWPXLD-mCherry-LC3	Modified pWPXLd transfer vector. Encodes a mCherry-LC3 fusion to be integrated into the target genome.
pMD2.G	Lentiviral toolkit vector: expresses lentiviral envelope gene. pMD2.G was a gift from Didier Trono (Addgene plasmid #12259)
pBABE-mCherry-EGFP-LC3	A vector used for the attempted creation of mCherry-LC3 for use in lentiviral transduction of HeLa cells.

2.1.1.4 Oligonucleotides

Table 2.4. **Oligonucleotides.**

Oligo name	Sequence	Description
ack_deltest_UpF	CATGCGCTACGCTCTATGG	Forward primer amplifying across genomic *ack*
ack_deltest_DownR	CGCCTTTGCGTTCCATTGC	Reverse primer amplifying across genomic *ack*
TssE1_deltestF	GCGTGAGCTGAACACCTGG	Forward primer amplifying across genomic *tssE* on PAI 1
TssE1_deltestR	CCGGGCGTAAATAGTTTGCC	Reverse primer amplifying across genomic *tssE* on PAI 1
TssE3_deltestF	CCGCTGGATAAAGAGACGC	Forward primer amplifying across genomic *tssE* on PAI 3
TssE3_deltestR	GTGATATTCTGGAATGTACCTG	Reverse primer amplifying across genomic *tssE* on PAI 3

Oligo name	Sequence	Description
p314_deltestF	GAGGAACATCAGCGATAGC	Forward primer amplifying across genomic *p314*
p314_deltestR	GCGCGCTTAGCTACACC	Reverse primer amplifying across genomic *p314*
p314_RT_F	GATCCCCACTCAGACACACC	Forward primer for reverse transcriptase PCR amplification of p314
p314_RT_R	CACAGTGGTCAGCCAGTACG	Reverse primer for reverse transcriptase PCR amplification of p314
rpoB_RT_F	CGTCCACAAGTTCTGGATGTACC	Forward primer for reverse transcriptase PCR amplification of *rpoB* control
rpoB_RT_R	CCGCTGTAGCTCTGAATCGG	Reverse primer for reverse transcriptase PCR amplification of *rpoB* control

Oligo name	Sequence	Description
pCMV_sF	ACTTCTGCTCTAAAAGCTGCGG	Forward primer for colony PCR of library pCMV-mEm-4GS vectors
pCMV_sR	CACTGCATTCTAGTTGTGGTTTGTC	Reverse primer for colony PCR of library pCMV-mEm-4GS vectors
pCMV_s_small_F	GATCACATGGTCCTGCTGGAGTT	Forward primer for colony PCR of library pCMV-mEm-4GS vectors containing small genes
p152_GSK_EcoRI_F	TGGAATTCAGGAGGATATTCATATGAAATCGAACATCATCCAC	Forward primer for amplification of LF82_152 from the LF82 genome, adding an EcoRI site
p152_GSK_NotI_R	GAGCGGCCGCCTTTTAATCTATCAAGGAGTGACT	Reverse primer for amplification of LF82_152 from the LF82 genome, adding a NotI site. Omits native stop codon.

Oligo name	Sequence	Description
GSK_tag_F	AATTCTTAATTAAGATCTGCGGCCGCATGAGTGGTCGCCCTCGCACTACTAGTTTCGCTGAAAGTTGAGGTACCT	GSK tag with EcoRI overhang. For annealing with GSK_tag_R. Omits native stop codon.
GSK_tag_R	CTAGAGGTACCTCAACTTTCAGCGAAACTAGTAGTGCGAGGGCGACCACTCATGCGGCCGCAGATCTTAATTAAG	GSK tag with XbaI overhang. For annealing with GSK_tag_F. Contains its own stop codon.
Common_sgRNA_R	TTTATAACCTCCTTAGAGCTCGA	Common primer for the amplification of pKD-sgRNA in two parts
Common_pTet_F	CCAATTGTCCATATTGCATCA	Common primer for the amplification of pKD-sgRNA in two parts

2.1.1.5 Bacterial media

Media and buffers were sterilised by autoclaving and made using ddH_2O unless otherwise stated.

Lysogeny Broth (LB) (1 L): 10 g Bacto/tryptone, 5 g yeast extract, 0.5 g NaCl. Add 1% Bacto agar for LB agar.

2x YT Broth (1 L): 16 g Bacto/tryptone, 10 g yeast extract, 5 g NaCl, adjusted to pH 7.2.

Super Optimal Broth (SOB) (1 L): 20 g Bacto/tryptone, 5 g yeast extract, 0.584 g NaCl, 0.186 g KCl, 2.4 g $MgSO_4$.

2.1.1.6 Bacterial media antibiotics and supplements

Table 2.5. **Media antibiotics**.

Antibiotic	Stock concentration	Working concentration	Solvent
Ampicillin (Amp)	100 mg mL^{-1}	100 µg mL^{-1}	70% ethanol
Kanamycin (Kan)	50 mg mL^{-1}	50 µg mL^{-1}	ddH_2O
Chloramphenicol (Chl)	34 mg mL^{-1}	34 µg mL^{-1}	100% ethanol
Spectinomycin (Spct)	50 mg mL^{-1}	50 µg mL^{-1}	ddH_2O
Streptomycin (Str)	50 mg mL^{-1}	50 µg mL^{-1}	ddH_2O
Tetracycline (Tet)	12.5 mg mL^{-1}	50 µg mL^{-1}	70% ethanol

Table 2.6. **Media supplements**.

Additive	Stock concentration	Working concentration	Solvent
L-Arabinose	20%	0.05%	ddH_2O
Glucose	20%	5%	ddH_2O
Anhydrotetracycline (aTc)	100 µg mL^{-1}	200 ng mL^{-1}	50% ethanol

2.1.1.7 Mammalian cell media and buffers

BSA: Bovine serum albumin (Sigma).

DMEM: Dulbecco's modified eagle medium, high glucose, GlutaMAX Supplement, pyruvate (Gibco).

NCS: Newborn calf serum, supplemented (HyClone), used at 10%.

Opti-MEM: Opti-MEM reduced serum media (Gibco).

PBS: Dulbecco's phosphate buffered saline without $MgCl_2$ and $CaCl_2$.

Typical cell culture medium: Medium consisted of DMEM + 10% NCS + Gentamicin[20] unless stated otherwise.

High BSA medium: DMEM + 10% NCS + gentamicin[20] + 1% BSA, used for lentiviral harvest.

2.1.1.8 Mammalian cell media antibiotics and supplements

Table 2.7. **Mammalian cell media and supplements**.

Antibiotic/Additive	Stock concentration	Working concentration
Gentamicin (gent)	50 mg mL^{-1}	20 µg mL^{-1}
DMSO	100%	10%

2.1.1.9 Other buffers and supplements

Antibody diluting solutions (Abdil): 0.1% Triton X-100, 2% BSA, in PBS.

20% arabinose: 20% L-arabinose in ddH_2O. Filter sterilised.

Formaldehyde fixing solution: 4% formaldehyde in PBS. Prepare fresh as needed.

15% glycerol: 15% glycerol in ddH_2O. Filter sterilised. Stored at 4°C.

Glycerol-mannitol: 20% glycerol, 1.5% mannitol in ddH_2O. Filter sterilised. Stored at 4°C.

Laemmli buffer (4x): 40% glycerol, 240 mM Tris-HCl (pH 6.8), 8% SDS, 0.04% bromophenol blue, 10% beta-mercaptoethanol.

Lysis buffer: 40 mM tris-acetate (pH 7.8), 20 mM sodium-acetate, 1 mM EDTA, 1% SDS. Filter sterilised.

6x Orange G loading dye: 10 mM Tris-HCl (pH 7.6), 60 mM EDTA (pH 8.0), 60% glycerol, 0.1% Orange G. Stored at 4°C short-term, -20°C long-term.

PBS-Tween (PBST): PBS with 0.1% v/v Tween 20

Phenol/ethanol for RNA preparations: 95% ethanol, 5% phenol.

Ponseau S stain: Ponceau S (0.1% w/v), acetic acid (5% v/v) in ddH_2O.

SLiCE buffer (10x): 500 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$, 10 mM ATP, 10 mM dithiothreitol. Stored at -20°C.

TB buffer (TB) (1x): 250 mM KCl, 15 mM $CaCl_2$, 10 mM PIPES-K (pH 6.7), 55 mM $MnCl_2$.

TBST: 0.15 M NaCl, 0.02 M Tris-Cl (pH 7.4), 0.1% Triton X-100.

TAE buffer (1 L) (50x): 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA.

Tris-EDTA (TE) buffer (1x): 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0).

Tris-glycine SDS running buffer: 25 mM Tris base, 192 mM glycine, 0.1% SDS, final pH 8.3.

Tris-glycine transfer buffer: 12 mM Tris base, 96 mM glycines, final pH 8.3.

Triton X-100: Diluted using PBS. Supplier: BDH.

Trypsin-EDTA (10x): 0.5% trypsin, 0.2% EDTA in PBS. Used at 10x. Supplier: PAA.

2.2 Methods

2.2.1 Bacterial growth and storage conditions

E. coli strains for use in cloning and transformation were grown in LB broth and on LB agar with relevant antibiotic selection, unless stated otherwise. Cultures were grown to stationary phase, with typical incubations performed at 37°C for ~16 h in a New Brunswick floor standing shaking incubator at 300 rpm for liquid cultures. LB plates were incubated at 37°C for at least 16 h,

unless otherwise stated. AIEC strain LF82 was also cultured under these conditions during use for cloning and transformation, with ampicillin used for selection.

E. coli XL1-blue and AIEC LF82 used for invasion assays were grown to stationary phase at 37°C (~18 h). Bacteria were then diluted 1:33 and grown at log phase before their use in assays (see: Invasion assays using HeLa CCL2).

Bacterial frozen glycerol stocks were picked using pipette tips and grown in LB broth to stationary phase (16 h) before streaking for single colonies on LB plates (both with relevant antibiotics). Long term storage of bacteria was achieved using a final concentration of 20% (v/v) sterile glycerol in a 1 mL culture, and samples frozen at -80°C.

2.2.2 Bacterial manipulation

2.2.2.1 Chemically competent cell preparation

Bacteria were grown overnight on LB agar with appropriate antibiotics and ~10 colonies used to inoculate 10 mL SOB. Broths were incubated for a maximum of 16 h at 37°C, 300 rpm. 2 mL and 8 mL of overnight culture were used to inoculate 2x 1 L flasks containing SOB, and incubated at 18°C, 300 rpm until one broth reached an OD_{600} of 0.5-0.6. This culture was cooled on ice for 10 min and then centrifuged at 3220 *g* for 10 min, 4°C. Pellets were resuspended in 50 mL TB and incubated on ice for 10 min, and then centrifuged again. Pellets were resuspended in 40 mL TB and 3 mL DMSO (Fisher) added dropwise. Cells were incubated for 10 min on ice, and 500 µL aliquots snap frozen with liquid nitrogen before storing at -80°C.

2.2.2.2 Chemically competent cell preparation for use in high-content screening

XL1-blue were prepared as described above until the final aliquot step. After resuspending cells and adding DMSO, 40 µL aliquots were dispensed into each well of a 96-well PCR plate. Cells were then snap frozen by placing each plate into liquid nitrogen, before storing at -80°C until used.

2.2.2.3 Chemically competent cell transformation

XL1-blue *E. coli* were routinely used for chemical transformation. Cells were defrosted and kept on ice at all times. 60 µL of cells for each reaction was transferred to pre-chilled 1.5 mL centrifuge tubes and left for 5 min before adding 200 ng of plasmid DNA. After 10 min, tubes were placed in a 42°C water bath for 40 sec, before returning directly to ice for 5 min. For plasmids containing Amp resistance cassettes, cells were directly transferred to LB-Amp plates and incubated overnight at an appropriate temperature. For all other resistance cassettes, cells were resuspended in 1 mL ice-cold SOB and incubated at a relevant temperature, shaking at 300 rpm for 1 h. Cells were then pelleted at 500 *g* for 2 min using a benchtop centrifuge and supernatant removed. Pellets were resuspended using 60 µL SOB and transferred to LB plates with antibiotics for overnight incubation.

2.2.2.4 High-throughput chemical transformation

From a library of post-SLiCE reaction DNA – consisting of identified LF82 CDSs recombined using SLiCE into pCMV-mEm-4GS vectors (see: Seamless Ligation in Cell Extract (SLiCE)) – 3 µL of SLiCE reaction was added to XL1-blue cells in 96-well PCR plates, and cells subjected to a heat-shock transformation as described above. Samples were then diluted via the addition of 100 µL SOB, and 20 µL used for the plating process outlined below.

A specialised plating system was used for the high-throughput plating of transformation samples. For this, a 3D bead-guide was designed, based upon

a multichannel plating unit developed by Hamilton et al. (2002). TinkerCad software (www.tinkercad.com) was utilised for the design of this bead-guide, and it was manufactured by Medical Manufacturing (University of Nottingham, UK) from polytetrafluoroethylene (PTFE). For more information regarding bead-guide design and Figures displaying the bead-guide, please refer to: 3D bead-guide design, and Figure 3.6.

The bead-guide was designed for placement onto pre-set agar in rectangular OmniTray plates (Nunc). Single sterile 5 mm glass beads (Sigma) were then dispensed into each lane created by bead-guide placement via the use of a 5 mm single-bead dispenser (TissueLyser). Transformation samples (20 µL) were then dispensed onto each bead using a multichannel pipette, and beads rolled along their respective lanes, resulting in the spreading of samples across the agar. Bead-guides were cleaned using 70% ethanol between uses.

2.2.2.5 Traditional electrocompetent cell preparation

Bacteria grown overnight on an LB plate were used to inoculate 3 mL LB broth (with appropriate antibiotics) and incubated at 37°C for 18 h. A 1:100 dilution was made using LB with antibiotics, and allowed to grow to an OD_{600} of 0.6-1.0. Samples were kept on ice at all times, and collected at 3000 *g* at 4°C for 15 min in an Eppendorf 5804-R swing-bucket centrifuge. Pelleted cells were resuspended and washed three times in 2x volumes of filter sterilised ice-cold 15% glycerol, centrifuging each time. Supernatant was removed and final cell pellets resuspended in 1:100 of the original growth volume using 15% glycerol. Single use aliquots (100 µL) were stored at -80°C until used.
This method was used from 2009-2015, at which point it was superseded by the Glycerol-mannitol method (below).

2.2.2.6 Glycerol-mannitol electrocompetent cell preparation

The glycerol-mannitol method for preparation of electrocompetent *E. coli* was derived from Warren (2011). Bacteria grown overnight on an LB plate

were used to inoculate 3 mL LB broth (with appropriate antibiotics) and incubated at 37°C for 18 h. A 1:100 dilution was made using 300 mL LB with antibiotics, and allowed to grow to an OD_{600} of 0.5-0.6. Cultures were then chilled on ice for 10 min, before centrifugation at 2000 *g* for 15 min at 4°C. Supernatant was removed, and the remaining pellet resuspended in 40 mL ice-cold ddH_2O.

The suspension was then divided into two 50 mL Falcon tubes, and 12 mL filter sterile glycerol-mannitol drawn up into a sterile pipette. 10 mL of this glycerol-mannitol solution was dispensed very slowly into the bottom of each tube to create a dense cushion of glycerol-mannitol below the aqueous portion containing bacteria. Samples were centrifuged at 2000 *g* for 15 min at 4°C with slow acceleration and deceleration. Supernatants were then removed using a 10 mL pipette, and the remaining pellet resuspended 1:20 of the original growth culture volume using glycerol-mannitol. Single use aliquots (100 µL) were stored at -80°C until used.

2.2.2.7 Electroporation

Plasmid DNA and PCR products were aliquoted into chilled electroporation cuvettes (1 mm or 2 mm, short) (Geneflow) to give final amounts of 200 ng, and 100 ng DNA, respectively. 40 µL electrocompetent cells were combined with DNA, and left on ice for 5 min. Samples were then electroporated using an Eppendorf Eporator at 2500 V or 1800 V for 2 mm and 1 mm cuvettes, respectively. Samples were returned to ice, and 1 mL chilled SOC added to each, before incubation at a relevant temperature, at 300 rpm for 1 h. Samples were centrifuged at 500 *g* to collect cells. Pellets were resuspended in 100 µL SOC and 50 µL transferred to LB plates with relevant antibiotics for growth overnight.

2.2.3 Mammalian cell culture

2.2.3.1 HeLa CCL2 cell growth conditions

HeLa CCL2 cells and their derivatives were routinely cultured in DMEM with 10% newborn calf serum (NCS) and 20 µg mL^{-1} gentamicin. Typical growth containers were 100 mm standard TC dishes (Sarstedt). HeLa were checked daily and subcultured when they neared/reached confluence.

To subculture, cell media was first warmed to 37°C in a water bath. Cells were washed with 10 mL PBS, and treated with 500 µL trypsin-EDTA (PAA). Plates were incubated at 37°C + 5% CO_2 for 5-10 min before resuspending in cell culture media. For routine subculture, a 1:4 dilution of cells was made using fresh cell culture media in a new plate. Where more specific seeding densities were required, HeLa were enumerated using a haemocytometer and diluted appropriately.

2.2.3.2 Caco-2 cell growth conditions

Caco-2 cells were treated similarly to HeLa cells (above), but were incubated for twice as long as stated for HeLa, due to their longer generation time.

2.2.3.3 Hek 293T cell culture

Hek 293T cells were cultured using DMEM + 10% NCS + gentamicin[20]. During media changes, care was taken to ensure that cells were not removed via shearing force due to their less adhesive nature. Trypsin-EDTA incubation was performed for 5 min at room temperature, and typical subculture was performed with a 1:8 dilution.

2.2.3.4 Long-term storage of mammalian cells

A confluent culture of low-passage number cells was washed with PBS and trypsin-EDTA treated (as above). Cells were resuspended in 9.5 mL cell culture media and transferred to a 15 mL Falcon tube. Cells were then

pelleted by centrifugation at 400 *g* for 2-3 min and supernatant removed. Pellets were then resuspended in 2 mL filter sterile cell culture media + 10% DMSO, and 1 mL samples aliquoted into cryotubes. Tubes were frozen at -80°C before transferring them to liquid nitrogen.

2.2.3.5 Thawing frozen mammalian cells

Tubes containing frozen cells were removed from liquid nitrogen and rapidly defrosted in a 37°C water bath. The 1 mL aliquot was then transferred into a 100 mm tissue culture dish with 10 mL cell culture media. After 24 h, cell media was changed, with cells ready to use ~3 days after initial thawing.

2.2.4 Mammalian cell manipulation

2.2.4.1 Mammalian cell transfection

Lipofectamine 2000 was used for HeLa CCL2 transfections in 12-well plates, and was supplied by Thermo Fisher Scientific.

To achieve 60-70% confluence for 12-well plate transfections, HeLa were typsin-EDTA treated, enumerated and diluted to achieve a seeding density of $1x10^5$ cells mL^{-1}. 1 mL of cells was dispensed into each well of a 12-well plate and incubated overnight at 37°C, 5% CO_2. Media was replaced with gent-free media. For each well, 100 µL Opti-MEM (Gibco) was combined with 1 µL Lipofectamine 2000 reagent and incubated at room temperature for 5 min. Meanwhile, 100 µL of Opti-MEM was mixed with 500 ng of vector. Lipofectamine 2000 and vector solutions were then combined, mixed gently, and incubated at room temperature for 20 min. Transfection mixtures were added to HeLa cell wells dropwise and plates incubated for 4 h. Transfection media was then changed for DMEM + NCS + gent[20] and plates incubated at 37°C + CO_2 overnight, with cells ready for analysis/further use the next day.

2.2.4.2 Mammalian cell reverse transfection

Lipofectamine 3000 (Thermo Fisher Scientific) was used for HeLa mCherry-LC3 transfections in either black 96-well clear bottom polystyrene microplates (Corning), or black 96-well round chimney CELLSTAR Bio-One tissue culture plates (Greiner).

HeLa mCherry-LC3 were trypsin-EDTA treated, enumerated and diluted to achieve a seeding density of $2.5x10^5$ cells mL^{-1}. 125 ng of each vector was dispensed into each well before the addition of cells. 25 µL of Opti-MEM and 0.8 µL of supplied P3000 reagent were mixed, added to each well and incubated for 3 min at 37°C + 5% CO_2. 30 µL of Opti-MEM and 0.3 µL of Lipofectamine 3000 reagent were then added to each well, and returned to the incubator for 10 min. 100 µL of cells were then dispensed into each well, and plates incubated for 4 h at 37°C, 5% CO_2, after which media was replaced with gent-free media for overnight incubation.

2.2.4.3 Transfection of Hek 293T cells for lentiviral production

Hek 293T cells were trypsin-EDTA treated, enumerated and diluted to achieve a seeding density of $8x10^4$ cells mL^{-1} in a 100 mm plate. Cells were incubated overnight to achieve a confluence of 60-70% next day. For preparation of transfection solution, 9 µg of psPAX2, 900 ng pMD2.G and 9 µg pWPXLd modified vector were combined with Opti-MEM to a final volume of 310 µL. Separately, 36 µL of Opti-MEM containing 54 µL of FuGene 6 (Promega) was prepared. Both solutions were gently mixed and incubated at room temperature for 10 min. Solutions were combined and incubated for a further 30 min. Hek cells were carefully washed once using PBS, media replaced with gent-free DMEM + NCS, and the transfection solution added to the plate dropwise before incubating overnight.

Media was exchanged for 15 mL high BSA medium (DMEM + 10% NCS + gentamicin[20] + 1% BSA) and cells incubated overnight. Cell media was

collected, stored at 4°C, and replaced again with 15 mL high BSA medium for another overnight incubation. Cell media was then collected and combined with the previously collected medium, and stored at 4°C for up to 3 days.

Lentiviral harvest media was centrifuged at 500 *g*, 4°C for 5 min and supernatant stored in 1 mL aliquots at -80°C.

2.2.4.4 Lentiviral transduction

HeLa CCL2 cells were trypsin-EDTA treated, enumerated and diluted to achieve a seeding density of $1x10^5$ cells mL^{-1} in a 12-well plate. Once 60-70% confluence was achieved, cell media in each well was replaced with 400 µL DMEM + NCS + $gent^{20}$ containing 8 µg mL^{-1} Polybrene (Sigma) and 600 µL lentiviral harvest. Plates were centrifuged at 1250 *g* for 30 min and incubated overnight. Individual wells were then trypsin-EDTA treated, cells enumerated and a dilution series performed across a 96-well tissue culture plate to result in an aliquot containing 1 cell per 100 µL in the final row.

Plates were incubated for over a week with regular media changes before checking for successful transduction using fluorescence microscopy. Wells that contained a homogeneous population of transduced cells were trypsin-EDTA treated and cells used to seed 1 well of a 12-well plate. Cells were incubated until confluent, used to seed a 100 mm plate until confluence, and subsequently trypsin-EDTA treated and frozen for long-term storage.

2.2.4.5 Co-localisation assays and immunofluorescence

HeLa cells were seeded onto glass slides in 12-well plates overnight. Reverse transfection (see above) was then used for the co-transfection of HeLa cells, with 500 ng of each pCMV-mEm-4GS vector of interest, and 200 ng of co-localisation vector. Transfection reagent volumes were scaled to accommodate 12-well plates, rather than the 96-well plates used in the high-content screen (see: Mammalian cell reverse transfection, above).

Post-transfection, cells were fixed using 4% formaldehyde in PBS, and stained using Hoechst 33342 and phalloidin 647, as described below.

2.2.4.6 Immunofluorescence

For immunofluorescence, HeLa cells were reverse transfected, as described above: Co-localisation assays and immunofluorescence. Next day, media was removed from each well and cells treated with 1 mL of 100% ice cold methanol for 2 min to fix cells. Cells were then rehydrated by washing with TBST 3x. Wells were blocked using Abdil for 10 min, before applying primary antibody in Abdil for 25 min, and subsequently washing each well three times using TBST. Secondary antibodies in Abdil were then applied for 25 min, and cells washed again three times with TBST. Final washes contained Hoechst 33342 and were left for 5 min, before aspirating well contents and mounting slides, as described in the next section.

Two anti-β tubulin primary antibodies were used in separate attempts at this experiment, and consisted of: Anti-β tubulin Mouse monoclonal sc58886 (Santa Cruz Biotech), and Anti-β tubulin Rabbit monoclonal 9F3 (Cell Signalling Technologies) received as a kind gift from Dr. Claire Friel (University of Nottingham, UK).

2.2.4.7 Fixing mammalian cells for imaging

For microscope slides, HeLa were seeded onto circular glass cover slips placed into 12-well plates prior to transfection. Post-transfection, slides were washed in wells once with PBS, and 1 mL 4% formaldehyde in PBS dispensed into each well. Plates were then incubated for 20 min at room temperature. Slides were washed again using PBS, and 1 mL 0.1% triton X-100 (BDH) was added for 5 min to permeabilise cells (if using phalloidin). Slides were washed with PBS, and 1 mL PBS containing 0.1% Hoechst 33342 (Molecular probes) and 2% phalloidin 647 (Molecular probes) applied for 15 min. PBS was used

for a final wash, before inverting cover slips onto microscopy slides and mounting using polyvinyl alcohol mounting medium with DABCO (Sigma).

2.2.4.8 Fixing mammalian cells for automated-imaging

HeLa transfected in 96-well plates were fixed using 100 µL 4% formaldehyde per well. Wells were then washed with PBS, and 0.1% Triton X-100 applied to each for 5 min to permeabilise cells. Cells were washed twice with PBS, and PBS containing 0.1% Hoechst 33342 and 2% phalloidin 647 applied. Plates were incubated for 10 min, and staining solution removed. Finally, cells were left to rest in 100 µL PBS at 4°C until imaged.

2.2.5 Nucleic acid manipulation

2.2.5.1 Restriction digestion

Restriction enzymes were supplied by New England Biolabs (NEB), and reactions performed according to the supplier's instructions; final reaction volumes were 50 µL. Where possible, restriction enzymes were heat inactivated at their recommended temperature for 20 min.

2.2.5.2 Dephosphorylation

5' and 3' DNA dephosphorylation was performed to prevent DNA self-ligation using Alkaline Phosphatase, Calf Intestinal (CIP) supplied by NEB. For each dephosphorylation reaction, 10 U of CIP was added directly to the restriction digestion reaction and incubated at 37°C for 1 h.

2.2.5.3 PCR amplification

DNA polymerases Phusion and One*Taq* were supplied by New England Biolabs (NEB), and DreamTaq was supplied by Thermo Fisher Scientific. Phusion was used for DNA amplification in cloning experiments and those which required high fidelity. One*Taq* was used for DNA amplification in

diagnostic testing, and DreamTaq was used during RT-PCR for detecting DNA contamination.

For each 20 µL Phusion PCR reaction 4 µL 5x Phusion HF buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 0.2 µL Phusion DNA polymerase and template DNA were combined with nuclease-free water. Template DNA amounts were 50 ng to 250 ng for genomic DNA and 1 pg to 10 ng for plasmid DNA.

For each 25 µL One*Taq* PCR reaction 5µL 5x One*Taq* standard buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 0.125 µL One*Taq* DNA polymerase and template DNA were combined with nuclease-free water. Template DNA amounts were 1 ng to 1 µg for genomic DNA, and 1 pg to 1 ng for plasmid DNA.

For colony PCR diagnostic testing, 1 bacterial colony or 0.5 µL overnight bacterial broth was resuspended in 100 µL ddH_2O, and incubated at 95°C for 10 min. 1 µL of boilate was used as template DNA in colony PCR reactions.
DreamTaq 25 µL PCR reactions contained 2.5 µL 10x DreamTaq buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 0.15 µL DreamTaq DNA polymerase and template DNA, with nuclease-free water. Template DNA/RNA amounts were 100 to 1000 ng for genomic DNA or RNA.
PCR cycling conditions used are described in Table 2.8. Annealing conditions were calculated using the NEB Tm Calculator, which utilises the Breslauer et al. (1986) method for Phusion, and the SantaLucia (1998) method for One*Taq* Tm calculations.

PCR primers were routinely designed using a combination of software: SerialCloner (http://serialbasics.free.fr/Serial_Cloner) and Tm Calculator (NEB). SerialCloner was used to manipulate DNA *in silico,* whilst Tm Calculator was used to optimise oligonucleotides Tm values for respective PCR reactions. See Table 2.4 (Oligonucleotides) for a list of primers used.

All oligonucleotides were synthesised by Integrated DNA Technologies (IDT).

Table 2.8. **PCR cycling conditions**.

Step	Phusion	One*Taq*	DreamTaq
Initial denaturation	98°C	94°C	94°C
	40 s	30 s	15 min
Denaturation	98°C	94°C	94°C
	20 s	20 s	1 min
Annealing	45-72°C	45-68°C	50°C
	20 s	30 s	1 min
Extension	72°C	68°C	72°C
	20 s kb^{-1}	1 min kb^{-1}	1 min kb^{-1}
Final extension	72°C	68°C	72°C
	5 min	5 min	10 min
Hold	4°C	4°C	4°C
Cycles:	30	25	30

2.2.5.4 PCR R1 and PCR R2 for SLiCE cloning

Primer design for the library of LF82 CDSs identified with InParanoid (see: InParanoid genome analysis) was performed using a perl script, Primal.pl (courtesy of Dr. Bill Wickstead, University of Nottingham, UK). This script allowed the addition of 'common' extensions to each amplified product of each member of the LF82 unique gene set, in what was termed 'PCR R1'. 'Common' extensions were added in PCR R1 for utilisation in PCR R2 (below). Designing primers in this way allowed global control over the desired melting temperature (Tm) of all primers designed, allowing the use of a single Tm during PCR R1 for the majority of amplifications.

Two desired Tm settings were applied using Primal.pl: a 'stringent' Tm of 50-55°C, and a 'less-stringent' Tm of 50-60°C. The latter was applied to those

CDSs for which Primal.pl could not find a suitable primer within the 'stringent' parameters set. See Appendix I for a list of primers.

PCR R1 was performed according to the parameters outlined in Table 2.9 using Phusion polymerase. CDSs which did not result in successful amplification after their first attempt at PCR R1 were re-attempted using the same PCR R1 parameters. If these CDSs failed their second attempt, a third and final attempt was made using a modified Tm of 52°C.

PCR R2 primers were designed using Serial Cloner. Primers contained regions of homology for the extensions added during PCR R1, and were designated as 'PCR_R2_Common_F' and 'PCR_R2_Common_R'. These primers were designed to add regions of homology for the destination vector, pCMV-mEm-4GS, when restriction digested with EcoRV-HF for the SLiCE reaction (see: Seamless Ligation in Cell Extract (SLiCE)). PCR R2 parameters used can be seen in Table 2.9.

Table 2.9. **PCRs used for SLiCE cloning**. PCR parameters for amplifying LF82 unique genes (PCR R1), and for adding extensions to each gene product (PCR R2).

Step	PCR R1	PCR R2
Initial denaturation	98°C	98°C
	30 s	30 s
Denaturation	98°C	98°C
	20 s	20 s
Annealing	55°C	47°C
	30 s	30 s
Extension	72°C	72°C
	1 min 30 sec	1 min 30 s
Final extension	72°C	72°C
	5 min	5 min
Hold	4°C	4°C
Cycles	30	20

2.2.5.5 Circular polymerase extension cloning (CPEC)

CPEC was previously described by Quan & Tian (2009) and was used here for creating pKD-sgRNA variants. This technique was used in combination with the no-SCAR system (see: Scarless Cas9 assisted recombineering (no-SCAR)) for LF82 recombineering. Two oligonucleotides were designed for each recombineering event. Each featured homology to the pKD-sgRNA vector via the addition of the sequences indicated in Table 2.10 (right hand column), and homology to the desired target gene. Gene homologous sequences preceded a protospacer adjacent motif (PAM) of nGG in the genome. A list of oligonucleotides used can be seen in Table 2.10.

Table 2.10. **Oligonucleotides for CPEC targeting the *ack* gene.** Final oligonucleotides purchased contained both the gene target homology and vector homology regions in one oligonucleotide.

Oligo name	Recombineering gene target homology	pKD-sgRNA vector homology
ackF_sgRNA	AATAAACAGGAAGCGGCTTT	GTTTTAGAGCTAGAAATAGCAAG
ackR_pTet	AAAGCCGCTTCCTGTTTATT	GTGCTCAGTATCTCTATCACTGA
TssE1_sgRNA	TTCTCGTCTGAAATCGCTGG	GTTTTAGAGCTAGAAATAGCAAG
TssE1_ptet	CCAGCGATTTCAGACGAGAA	GTGCTCAGTATCTCTATCACTGA
TssE3_sgRNA_2	ACGCCTGAGTCAGATACAGG	GTTTTAGAGCTAGAAATAGCAAG
TssE3_pTet_2	CCTGTATCTGACTCAGGCGT	GTGCTCAGTATCTCTATCACTGA
p314_sgRNA	GGGTTGCAGCAGAGTCAGAC	GTTTTAGAGCTAGAAATAGCAAG
p314_ptet	GTCTGACTCTGCTGCAACCC	GTGCTCAGTATCTCTATCACTGA

'Common' oligonucleotides were also designed, which were homologous to the pKD-sgRNA vector, 4.4 Kb and 2.8 Kb upstream of the target gene oligonucleotides. These can be found in Table 2.4 (Oligonucleotides). A Phusion PCR reaction was performed using the sgRNA and pTet amplified primers with corresponding 'common' primer (see Figure 2.1). PCR reactions were DpnI treated to remove methylated plasmid DNA, using 20 U of enzyme for each reaction. DpnI digestions were incubated at 37°C for 15 min, and samples subsequently run on a 1% agarose gel. Relevant bands were gel extracted and purified, before using 100 ng of each vector fragment in a CPEC Phusion PCR. The PCR cycling conditions used can be seen in Table 2.11.

XL1-Blue cells were chemically transformed using 15 µL of CPEC PCR product. Cells were grown for at least 24 h on LB-Spct plates at 30°C. Once grown, individual clones were selected for long-term storage and for plasmid extraction.

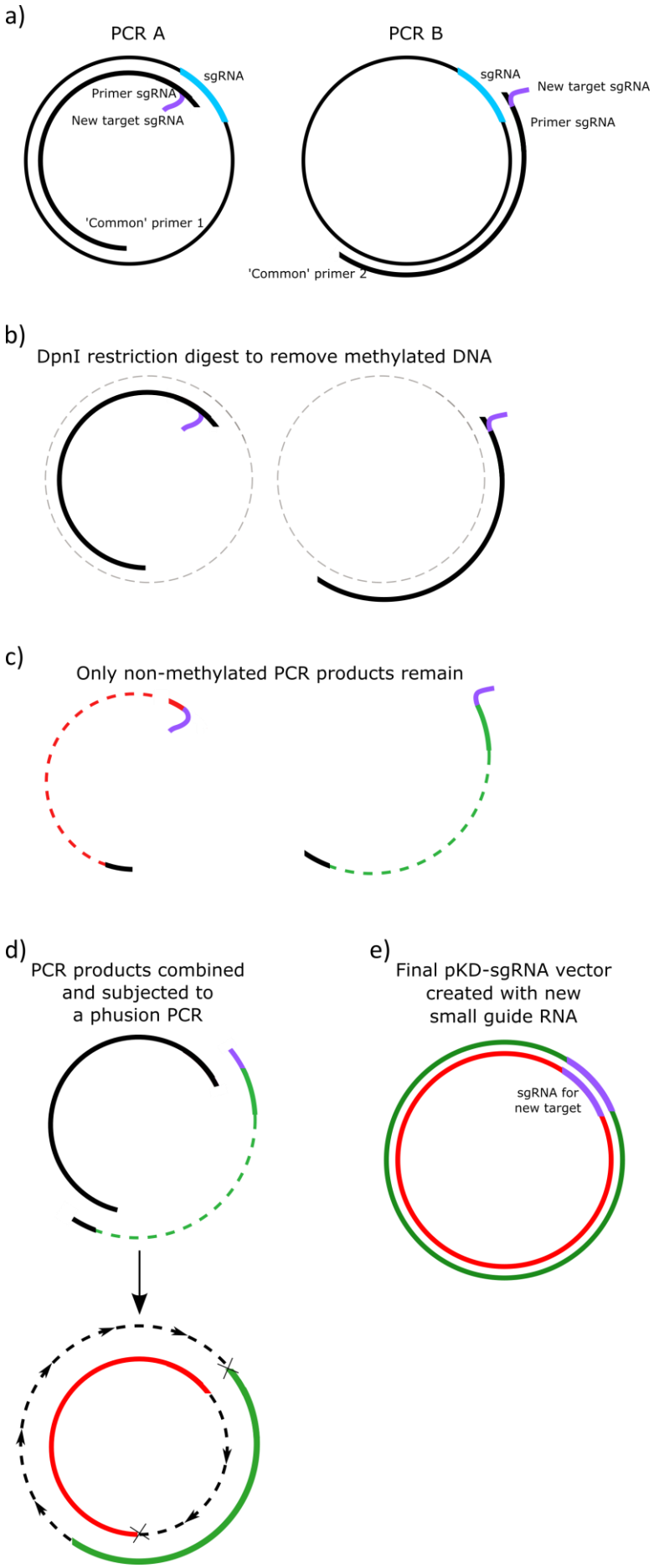

a)
PCR A
sgRNA
Primer sgRNA
New target sgRNA
'Common' primer 1
PCR B
sgRNA
New target sgRNA
Primer sgRNA
'Common' primer 2
b)
DpnI restriction digest to remove methylated DNA
c)
Only non-methylated PCR products remain
d)
PCR products combined and subjected to a phusion PCR
e)
Final pKD-sgRNA vector created with new small guide RNA
sgRNA for new target

Figure 2.1. **Circular polymerase extension cloning (CPEC).** a) Two separate PCR reactions amplify overlapping halves of the pKD-sgRNA vector, with extensions encoding a new sgRNA targeting sequence (purple). The undesired sgRNA region is omitted from the amplification. b) Both PCR reactions are DpnI digested to remove parent, methylated vector. c) Only PCR amplified fragments remain after DpnI digestion. d) PCR products are combined and act as primers for one another in a subsequent PCR reaction. This creates a semi-complete vector with a nick on each strand (X on diagram) where extended PCR products meet their respective vector fragments. e) The semi-complete vector is transformed into *E. coli* and nicks are sealed, resulting in a complete pKD-sgRNA vector with a modified sgRNA region for the targeting of a new genomic location.

Table 2.11. **CPEC cycling conditions.**

Step	**Phusion**
Initial denaturation	98°C 3 min
Denaturation	98°C 30 s
Annealing	58°C 30 s
Extension	72°C 3 min
Final extension	72°C 10 min
Hold	4°C
Cycles:	15

2.2.5.6 Reverse-transcriptase PCR

RT-PCR reactions were performed using the OneStep RT-PCR kit (Qiagen), which utilises HotStarTaq DNA polymerase, with Omniscript and Sensiscript reverse transcriptases. For each 25 µL reaction 2.5 µL 5x Qiagen OneStep RT-PCR buffer, 2.5 µL 5x Q-solution, 1 mM dNTPs, 0.6 µM forward primer, 0.6 µM reverse primer, 1 µL OneStep RT-PCR enzyme mix and template RNA were combined with RNase-free water. Template RNA amounts were 10-150 ng, DNA control amounts were 50-150 ng. PCR cycling conditions for the OneStep RT-PCR can be seen in Table 2.12.

Table 2.12. **RT-PCR cycling conditions**.

Step	OneStep RT-PCR
Reverse transcription	50°C 30 min
Initial denaturation	94°C 15 min
Denaturation	94°C 1 min
Annealing	50°C 1 min
Extension	72°C 1 min
Final extension	72°C 10 min
Hold	4°C
Cycles:	30

2.2.5.7 DNA ligation

DNA ligations were performed using either T7 or T4 DNA ligases (NEB). For each 30 µL reaction, 50 ng vector DNA and insert DNA at a molar ratio of 1:3 were combined with 400 U of DNA ligase, nuclease-free water and relevant DNA ligase buffer. Reactions were incubated at 16°C overnight, and T7 ligase reactions used directly for transformation. T4 ligase reactions were subjected to heat inactivation at 65°C for 15 min prior to transformation.

2.2.5.8 Oligonucleotide annealing

Oligonucleotides were resuspended to 100 µM, and 1 µL of each oligonucleotide combined in a 1.5 mL Eppendorf tube with 1 µL T4 buffer, 6.5 µL ddH_2O, and 0.5 µL T4 polynucleotide kinase (PNK). Tubes were first incubated at 37 °C for 30 min, before moving to a 95 °C hot block. Metal inserts were removed from the hot block, and allowed to cool slowly to room temperature, thus allowing oligonucleotides to anneal.

2.2.5.9 DNA sequencing

DNA sequencing was performed using Sanger 4 colour sequencing technology by Source Bioscience, Nottingham, using sequencing oligonucleotides designed as described above.

2.2.6 Nucleic acid purification

2.2.6.1 Agarose gel extraction and purification

PCR products, restriction digests and ligation reactions subjected to agarose gel electrophoresis were purified using the Qiagen QIAquick Gel Extraction Kit. Gel band extraction was performed using a clean razor blade, and the gel fragment subjected to the manufacturer's recommended protocol. Final elution was performed using 25 µL of the supplied elution buffer.

2.2.6.2 PCR purification

PCR reactions were purified using the QIAquick PCR purification kit (Qiagen) and the supplied protocol. Final elution was performed using 25 µL of the supplied elution buffer.

2.2.6.3 Genomic DNA preparation

Genomic DNA was extracted and purified from *E. coli* using the chloroform extraction technique, as described by Chen & Kuo (1993). 1.5 mL of saturated culture was centrifuged at 9700 *g* for 3 min using a benchtop centrifuge. The

resulting pellet was lysed using 200 µL lysis buffer (40 mM tris-acetate (pH 7.8), 20 mM sodium-acetate, 1 mM EDTA, 1% SDS.), and resuspended in 66 µL 5M NaCl. Lysate was centrifuged at 9700 *g*, 4°C for 10 min, and the supernatant collected and mixed thoroughly 50% v/v with chloroform. The sample was then centrifuged for a further 3 min, and the upper aqueous phase mixed with 500 µL 100% ethanol in a new tube. Samples were centrifuged for 3 min and pellets washed twice using 500 µL 70% ethanol. Supernatant was removed and pellets allowed to dry at 50°C before resuspending the final genomic DNA precipitate in 50 µL TE and storing at -20°C.

2.2.6.4 Plasmid DNA extraction and purification

High-purity endotoxin-free DNA extraction was performed using the NucleoBond Xtra Midi EF Midiprep Kit (Macherey-Nagel). Cultures were prepared by inoculating 3 mL of LB (plus relevant antibiotics) with bacteria containing the plasmid of interest, and incubating at 37°C for 6 h. Large, 500 mL baffled glass flasks containing 300 mL LB (with relevant antibiotics) were then inoculated with the entire 3 mL of bacterial culture. These were incubated overnight at 37°C with shaking at 300 rpm before harvesting. The manufacturer's instructions were then followed to extract and purify plasmid DNA using the centrifugation method, with isopropanol/ethanol precipitation. Final plasmid DNA was eluted in 500 µL TE buffer.

Lower purity plasmid extraction was performed using a QIAprep Spin Miniprep Kit (Qiagen). Bacteria were inoculated into 3 mL LB (with relevant antibiotics) and incubated overnight at 37°C before harvesting. Final plasmid DNA was eluted in 25 µL of supplied elution buffer.

High-content plasmid extraction was performed using the NucleoBond 96 Xtra HF (Macherey-Nagel). Final plasmid DNA was eluted using 50 µL of supplied elution buffer.

2.2.6.5 RNA extraction and purification from bacteria

E. coli were cultured under relevant test conditions and 3 mL of broth collected. For agar cultures, samples were collected via the addition of 4 mL LB broth to agar surfaces to resuspend cells. 3 mL of the resuspension broth was then collected.

Samples were mixed by inversion with 1 mL phenol/ethanol, and incubated on ice for 45 min. They were then centrifuged using a benchtop centrifuge at 3000 *g*, 4°C for 10 min, and supernatant removed. Samples were centrifuged again for 2 min, and remaining supernatant removed before storing tubes at -80°C.

Bacterial pellets were resuspended in 100 µL TE containing 50 mg mL^{-1} lysozyme and incubated at room temperature for 5 min. 75 µL lysis reagent (Promega SV total RNA purification kit) was added, samples mixed by inversion, and 350 µL RNA dilution buffer added. Samples were incubated at 70°C for 3 min and centrifuged at max speed for 10 min. 400 µL supernatant was transferred into a clean tube containing 200 µL 100% ethanol and mixed. Samples were then dispensed into the supplied spin columns and subjected to centrifugation, washing and DNase treatment as per the manufacturer's instructions. After final washes, column-bound-RNA was incubated for 1 min in 50 µL RNase-free water, and centrifuged for 1 min. Flow-through was transferred to a clean 0.5 mL microcentrifuge tube and 1 µL DNase and 5 µL DNase buffer added from the Ambion DNA-free kit for 25 min at 37°C. Manufacturer's instructions were then followed until the final centrifugation step, where samples were centrifuged at 14000 *g* for 2 min, and final supernatant containing RNA carefully collected.

RNA integrity was checked using the Agilent bioanalyser.

2.2.6.6 DNA/RNA quantification and purity

DNA/RNA was quantified using a Nanodrop 2000 (ThermoFisher), and purity measured using the 260:280 nm absorbance ratio.

2.2.7 Gel electrophoresis

2.2.7.1 Agarose gel electrophoresis

Agarose gels were cast using 1x TAE and agarose (Sigma) for electrophoresis. DNA was stained using SYBR Safe DNA gel stain (Thermo Fisher Scientific), used at a ratio of 1:20000. DNA samples were loaded onto the gel with 1x Orange G loading dye, and run at 10 V cm^{-1}. Gels were visualised using a Syngene Ultraslim LED transilluminator and ImageJ using the ImageJ Video capture macro. DNA ladders used were the 1 Kb DNA ladder and 100 bp DNA ladder, supplied by NEB.

2.2.7.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

For mammalian cells, PBS with 0.5 mM EDTA was used to lift cells, and centrifugation then used to pellet samples. Supernatant was removed and 60 µL 2x laemmli buffer in ddH_2O added to each sample, before heating to 95°C for 5 min. Samples were run on a Criterion TGX precast tris-glycine gel (18%) at 150 V for ~1 h 30 min along with Precision Plus dual colour protein ladder (10-250 kD) (Biorad).

2.2.7.3 Western blot

SDS-PAGE gel protein was transferred to a PVDF membrane (Hybond-P membrane (Amersham)) at 30 V for 1 h. Ponceau S stain was used to detect membrane protein bands, and removed using multiple ddH_2O washes. 5% milk was used to block membranes for 1 h at room temperature before the addition of antibodies. Washes between antibody applications were performed using PBST. Final visualisation was performed using ECL Prime

western blotting substrate (Amersham), Hyperfilm ECL film (Amersham) and the Konica Minolta SRX-201 automated developer.

2.2.8 Gene deletion strategy and assays

2.2.8.1 Scarless Cas9 assisted recombineering (no-SCAR)

The no-SCAR method described by Reisch & Prather (2015) was used for LF82 recombineering and production of deletion mutants. An overview of this method can be seen in Figure 2.2.

Deletion oligonucleotides were designed as ~80mers, with ~40 bp homology to the upstream and ~40 bp homology to the downstream region of the desired deletion gene lagging strand. Oligonucleotides were modified with phosphorothioate bonds for 4 bp at both 5' and 3' ends, in accordance with previously described parameters to improve oligonucleotide stability (Wang & Church 2011).

pCas9-cr4 and relevant pKD-sgRNA vectors were sequentially transformed into LF82 via electroporation, with pCas9-cr4 transformation first, followed by pKD-sgRNA. pkD-sgRNA-containing clones were cultured at 30°C to maintain the temperature-sensitive vector. Clones were made electrocompetent using the glycerol-mannitol method and tested for anhydrotetracycline (aTc) sensitivity. aTc sensitive LF82 pCas9-cr4 pKD-sgRNA were prepared for recombineering by growth to an OD_{600} of 0.4-0.5, and the addition of arabinose for 30 min. Cells were then made electrocompetent using the glycerol-mannitol method and transformed using 10 µM of relevant deletion oligonucleotide. Recovery was performed for 1.5 h at 30°C in 1 mL SOB-arabinose broth. A 1:100 dilution of recovery broth was made using LB, and 50 µL transferred to LB-Chl-Spct plates +/- aTc for overnight incubation at 30°C. Large colonies from the aTc^+ plates were selected and tested for successful deletion mutations via colony PCR and confirmed via sequencing.

Removal of the pKD-sgRNA vector was induced by growth on LB-Chl at 45°C for 15 h, and loss confirmed via spectinomycin sensitivity testing. Removal of pCas9-cr4 was induced by preparing electrocompetent cells from the pKD-sgRNA-loss clones, and transforming with pKD-sgRNA-p15 – a vector containing a guide sequence directed towards the pCas9-cr4 vector. Cells were recovered at 30°C in SOB for 1.5 h, and an additional 2 h with the addition of 2x normal working concentration aTc. 100 µL of the samples was transferred to LB-Spct-aTc plates and incubated overnight at 30°C. Colonies were subcultured onto LB plates and incubated for 15 h at 45°C, with resulting colonies tested for spectinomycin sensitivity and chloramphenicol sensitivity. Final deletion strains were frozen for long-term storage.

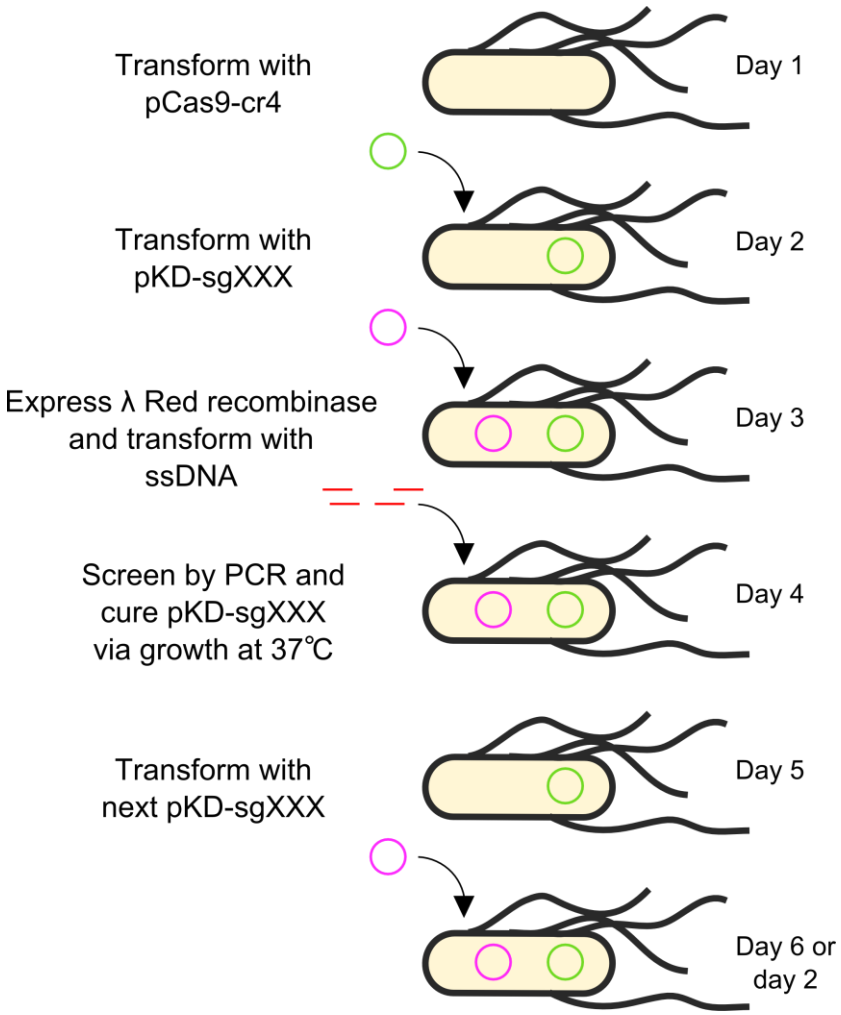


Figure 2.2. **Workflow of the no-SCAR system for gene deletion**. Days 1 & 2: Bacteria were transformed with the Cas9 and λ Red recombinase-expressing pCas9-cr4 vector, followed by transformation with the newly created pKD-sgRNA vector. Day 3: λ Red recombinase expression was induced via arabinose induction. Day 4: Linear DNA was introduced via transformation for recombination with target deletion sequence. Colony PCR was used to indicate potential success of deletion, and growth at 37°C used to cure pKD-sgRNA vector. Day 5: The pKD-sgRNA-p15 vector was introduced to force loss of pCas9-cr4. Alternatively, at this point, the introduction of pKD-sgRNA can be used to restart the process for the deletion of another target sequence. Final deletion mutants were confirmed via sequencing. Figure adapted from Reisch & Prather (2015).

2.2.8.2 Invasion assays using HeLa CCL2

HeLa CCL2 cells were trypsin treated and enumerated as described previously (see: Mammalian cell culture). HeLa were subsequently diluted to $1x10^5$ cells mL^{-1} and 1 mL dispensed into each well of a Corning Costar 12-well plate. Plates were incubated at 37°C + 5% CO_2 for 18 h before trypsin-EDTA treating one well to enumerate cells using a haemocytometer, this value was recorded for later use in calculating multiplicity of infection (MOI). Cells were then infected as described below.

Bacterial strains to be assayed for invasive ability were inoculated into 3 mL LB (with relevant antibiotics) and incubated for 18 h at 37°C shaking at 300 rpm. Broths were diluted 1:33 with a final volume of 1 mL LB and returned to the shaking incubator for 1 h 30 min. OD_{600} measurements were then determined using an Eppendorf BioPhotometer Plus; samples were diluted to give a final MOI of 10 cells $HeLa^{-1}$ (assuming an OD_{600} of 1.0 = $8x10^8$ bacteria mL^{-1}). This pre-infection media was diluted 1:1000 using PBS and 6 x 20 µL spots were dispensed onto LB agar (no antibiotics) before incubating at 37°C overnight.

HeLa cells were washed using 1 mL DMEM + NCS (no gentamicin), and 1 mL bacterial culture added to each respective well. Plates were centrifuged at 700 *g* for 15 min using an Eppendorf 5804-R swing-bucket centrifuge, before returning to the incubator for 1 h 45 min. Each well was then washed twice using 1 mL cell culture media, and 1 mL DMEM + NCS + $Gent^{100}$ added to each for 40 min. High gent media was then replaced with normal cell culture media.

After a total of 1 h of gentamicin treatment, wells were washed 3x with PBS and 500 µL 1% Triton X-100 in PBS added for 5 min to permeabilise cells. Triton X-100 solution was pipetted up and down 5x before dispensing into a 1.5 mL centrifuge tube. A 1:10 dilution was made using PBS, and 6 x 20 µL spots were dispensed onto LB agar (no antibiotics) for both undiluted and

1:10 diluted samples. Plates were incubated overnight at 37°C for CFU enumeration the next day.

2.2.8.3 Invasion assay using Caco-2 cells

Caco-2 cells were treated similarly to HeLa cells, as described above. However, as Caco-2 grow at roughly half the rate of HeLa, incubation periods were doubled from those stated for HeLa.

12-well plates were therefore seeded with $1x10^5$ cells mL^{-1} and incubated at 37°C for 36-48 h before enumeration. Bacterial strains were grown in 3 mL LB broth at 37°C for 24 h, without shaking. OD_{600} measurements were then determined using an Eppendorf BioPhotometer Plus, and samples diluted to give a final MOI of 100 bacterial cells Caco-2^{-1}. Bacteria were allowed to invade for 2 h, before application of gentamicin-containing media, Triton X-100 treatment, and plating for enumeration, as described above.

2.2.8.4 GSK-tag vector creation

The pBAD18-Kan vector was restriction digested using EcoRI-HF and XbaI, and then CIP treated. Digests were subjected to gel electrophoresis, from which successfully digested vector was selected and get purified.

Two oligonucleotides were designed to result in a GSK tag with 3' and 5' overhangs which, once annealed, corresponded to overhangs present in successfully digested pBAD18-Kan vector. These oligonucleotides can be seen in Table 2.4 (Oligonucleotides).

Oligonucleotides were annealed, as described above (see: Oligonucleotide annealing), and were subsequently ligated into the digested pBAD18-Kan vector using T4 ligase. Ligation reactions were then used to transform XL1-

blue cells via heat shock, before selecting for successfully ligated and transformed vector via growth on LB-Kan agar. Next day, colonies were selected and grown in LB-broth for a midiprep. Successfully midiprepped DNA was subjected to a restriction digest using SpeI – a restriction site of which is found within the GSK tag sequence – and subsequent gel electrophoresis to visualise digested vector. Final purified vector was sequenced to confirm GSK tag insertion.

2.2.8.5 GSK-tag secretion assay

The p152 ORF was amplified from LF82 genomic DNA using oligonucleotides which introduced EcoRI and NotI restriction sites to 3' and 5' ends, respectively (p152_GSK_EcoRI_F, and p152_GSK_NotI_R in Table 2.4 (Oligonucleotides)). Amplified product was subjected to gel electrophoresis and gel purification, before being restriction digested using EcoRI-HF and NotI-HF, and introduced to the corresponding digested pBAD18-Kan-GSK vector via T4 ligase. Ligation reactions were heat inactivated and used to transform XL1-blue cells via heat shock, before being cultured on LB-Kan agar overnight. Colonies were selected from the test plate for growth in LB-Kan broth to be midiprepped and subsequently sequenced, confirming p152 insertion. The pBAD18-p152-GSK vector was then transformed via electroporation into LF82. Successfully transformed samples were frozen for glycerol stocks.

LF82 containing the pBAD18-p152-GSK vector (referred to as LF82-GSK) was grown overnight in LB-Kan at 37°C, along with wild-type (wt) LF82. Next day, LF82-GSK was divided into two 15 mL falcon tubes, where 1% arabinose was added to one broth – the 'induced' sample – and one sample was left without arabinose – the 'non-induced' sample. These samples were used to perform a HeLa cell invasion assay (see: Invasion assays using HeLa CCL2).

Invasions were performed for 3 different invasion time periods: 2 h (no gentamicin), 3 h (with 1 h gentamicin), and 5 h (with 3 h gentamicin). Once complete, cells were recovered using PBS with 0.5 mM EDTA treatment and were subjected to SDS-PAGE (see: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)), followed by Ponceau S staining, and Western blotting. A rabbit polyclonal anti-GSK-3β-Tag antibody (Cell Signalling Technologies, #9325) was used for the detection of both phosphorylated and unphosphorylated GSK tags, with a goat anti-rabbit HRP conjugate (Sigma, A0545) used as a secondary antibody.

2.2.9 AIEC genomic library creation and screening

2.2.9.1 Identification of AIEC CDS for analysis

The AIEC reference genome (NC_011993.1) was obtained from ftp://ftp.ncbi.nlm.nih.gov/. Six other genomes were also obtained from this location (see: Table 2.13).

Table 2.13. **List of genomes used for comparison using InParanoid**.

Strain	Accession number
EHEC O111:H	NC_013364.1
EPEC O127:H6	NC_011601.1
UPEC 536	NC_008253.1
K12 DH10B	NC_010473.1
K12 MG1655	NC_000913.3
K12 W3110	NC_007779.1

InParanoid 4.1 was downloaded and used to determine likely inparalogs from pairwise comparisons of LF82 vs each of the strains listed in Table 2.13 (O'Brien 2004). From these results, the list of inparalogs identified were removed from the LF82 base genome, thus resulting in a list of LF82 CDSs that were not determined to be imparologs to another strain. This process

was then repeated for each subsequent strain, progressively lessening the repertoire of CDS present in the LF82 list of non-inparologs. The use of InParanoid is further discussed in: InParanoid genome analysis

To confirm that the final non-inparalog LF82 CDS list was indeed a reasonable representation of 'unique' genes, the resulting list was subjected to a nucleotide BLAST against all comparison strains. Furthermore, this list was also compared to a list of CDSs present/not present in various *E. coli* strains provided by Miquel et al. (2010) (supplementary material: Table S2).

2.2.9.2 Imaging of screen plates

Transfections were performed in 96-well plates and cells fixed and stained using Hoechst 33342 and phalloidin 347, as described above (see: Fixing mammalian cells for automated-imaging). The ImageXpress Ultra (School of Life Sciences Imaging (SLIM), University of Nottingham, UK) was used to image plates in an automated fashion. Images were acquired at 40x magnification for 6 sites per transfection well. These sites were arranged as seen in Figure 2.3 to avoid any edge-effects, and were spaced in an attempt to obtain a representative overview of the entire well. The ImageXpress Ultra is known to show some inconsistency toward the edges of 96-well plates, hence the lack of use of outer wells.

Excitation wavelengths of 405, 488, 561 and 635 nm were utilised for image collection, and underwent initial testing to set laser power and gain; these were kept constant for all samples.

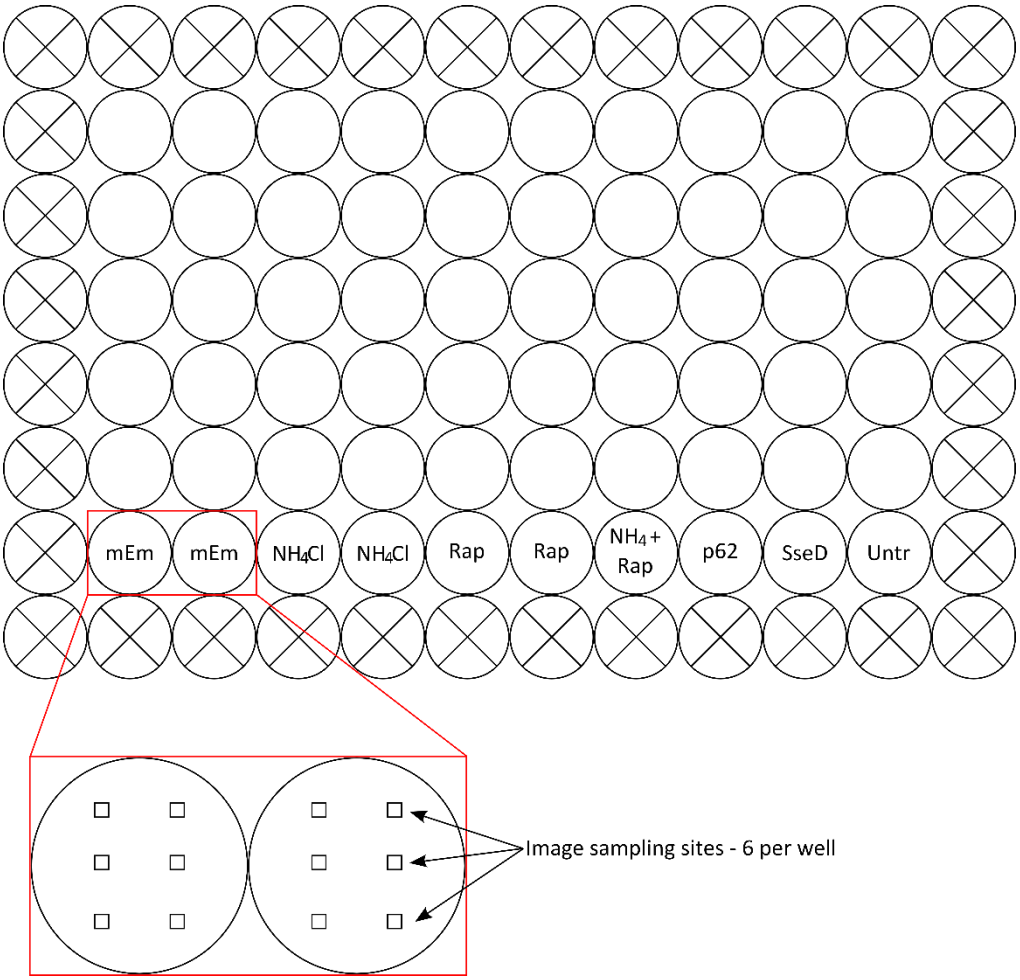


Figure 2.3. **Layout of a typical 96-well transfection plate used for cell transfection and imaging**. Crosses indicate wells not used due to limitations of the IX Ultra microscope. The magnified region highlights the arrangement of the 6-imaging sites used for each well. Wells G02-G11 were control wells. mEm: cells transfected with empty pCMV-mEm-4GS vector. NH_4Cl: cells treated with ammonium chloride. Rap: cells treated with rapamycin. NH_4Cl + Rap: cells treated with ammonium chloride and rapamycin. P62: cells transfected with a vector containing a p62-GFP fusion. SseD: cells transfected with a vector containing unlabelled *Salmonella typhimurium* secreted effector protein SseD. Untr: untransfected cells not treated with ammonium chloride or rapamycin.

2.2.9.3 Analysis of screen data

Images obtained from the ImageXpress Ultra automated microscope were analysed using CellProfiler (Carpenter et al. 2006). Subsequent data obtained from CellProfiler was analysed using CellProfiler Analyst and R (R Development Core Team 2011; Jones et al. 2008). R packages used for data manipulation and analysis were RSQLite, ggplot2 and Rcmdr (Wickham et al. 2015; Wickham 2009; Fox 2005).

2.2.10 Seamless Ligation in Cell Extract (SLiCE)

2.2.10.1 PPY extract preparation

The *E. coli* PPY strain (received as a kind gift from Dr. Yongwei Zhang, Albert Einstein College of Medicine, New York) was cultured under standard conditions on LB-Strep-Chl agar. One colony was used to inoculate 25 mL 2xYT-Strep broth in a Falcon 125 mL baffled Erlenmeyer flask and incubated at 37°C for 16 h at 300 rpm. The OD_{600} of the overnight culture was then determined, and the culture diluted to OD_{600} of 0.03 in 50 mL 2xYT-Strep broth + 0.1% glycerol. This culture was returned to the incubator until the OD_{600} reached 5-5.5. L-arabinose was then added to give a final concentration of 0.1%, and incubation continued for 2 h.

Broth was centrifuged in 50 mL Falcon tubes at 3220 *g,* 4°C for 20 min, and resulting pellets washed with 50 mL ddH_2O before centrifugation under the same conditions. For each 0.23 g of wet weight, pellets were resuspended in 300 µl CellLytic™ B Cell Lysis Reagent (Sigma, B7435). The suspension was then aliquoted into 1.5 mL protein lobind tubes and incubated at room temperature for 10 min. Tubes were centrifuged at 15000 *g* using a benchtop microcentrifuge for 2 min, and supernatant aliquoted into clean 1.5 mL protein lobind tubes with equal volumes of 100% glycerol. 40 µL aliquots were stored in protein lobind tubes at -80°C indefinitely.

2.2.10.2 SLiCE PCR

Non-inparalog LF82 CDS were PCR amplified ready for SLiCE using two Phusion PCR reactions (PCR R1, and PCR R2). PCR R1 resulted in the amplification of LF82 unique genes, and added short extensions homologous to a 'common primer' (used later).

PCR purification was then performed using the NucleoFast 96 PCR kit (Macherey-Nagel) and the resulting product amplified using the 'common primers' in a second (PCR R2) PCR. This PCR used a common primer pair to add long (~40 bp) 5' and 3' extensions to each amplified product, resulting in extensions homologous to the desired insert region on the pCMV-mEM-4GS vector. For more information of PCR R1 and PCR R2, see: PCR R1 and PCR R2 for SLiCE cloning, and Figure 2.4.

2.2.10.3 SLiCE reaction

Non-inparalog LF82 CDS were each cloned into the pCMV-mEm-4GS vector using the Seamless Ligation in Cell Extract method (Zhang et al. 2012). pCMV-mEm-4GS vector was digested using EcoRV-HF (NEB) and gel purified. A 10 µL SLiCE reaction was prepared by combining 100 ng of linearised pCMV-mEm-4GS vector, 1 µL PPY extract, 1.2 µL 10x SLiCE buffer and 2 µL PCR R2 product with ddH_2O. Components were mixed using a pipette and tubes incubated in a 37°C water bath for 1 h. SLiCE reactions were used directly for transformation of chemically competent XL1-Blue, using 3 µL of SLiCE reaction and 40 µL of competent cells (see: High-throughput chemical transformation). An overview of the SLiCE reaction can be seen in Figure 2.4.

The presence of successful SLiCE vectors in XL1-blue cells was determined via the use of colony PCR, using the pCMV_sR and pCMV_sF primers. Small fragments of <150 bp were amplified using pCMV_sR and pCMV_s_small_F if particular difficulty was seen in observing PCR products on gels after amplification with the above primers.

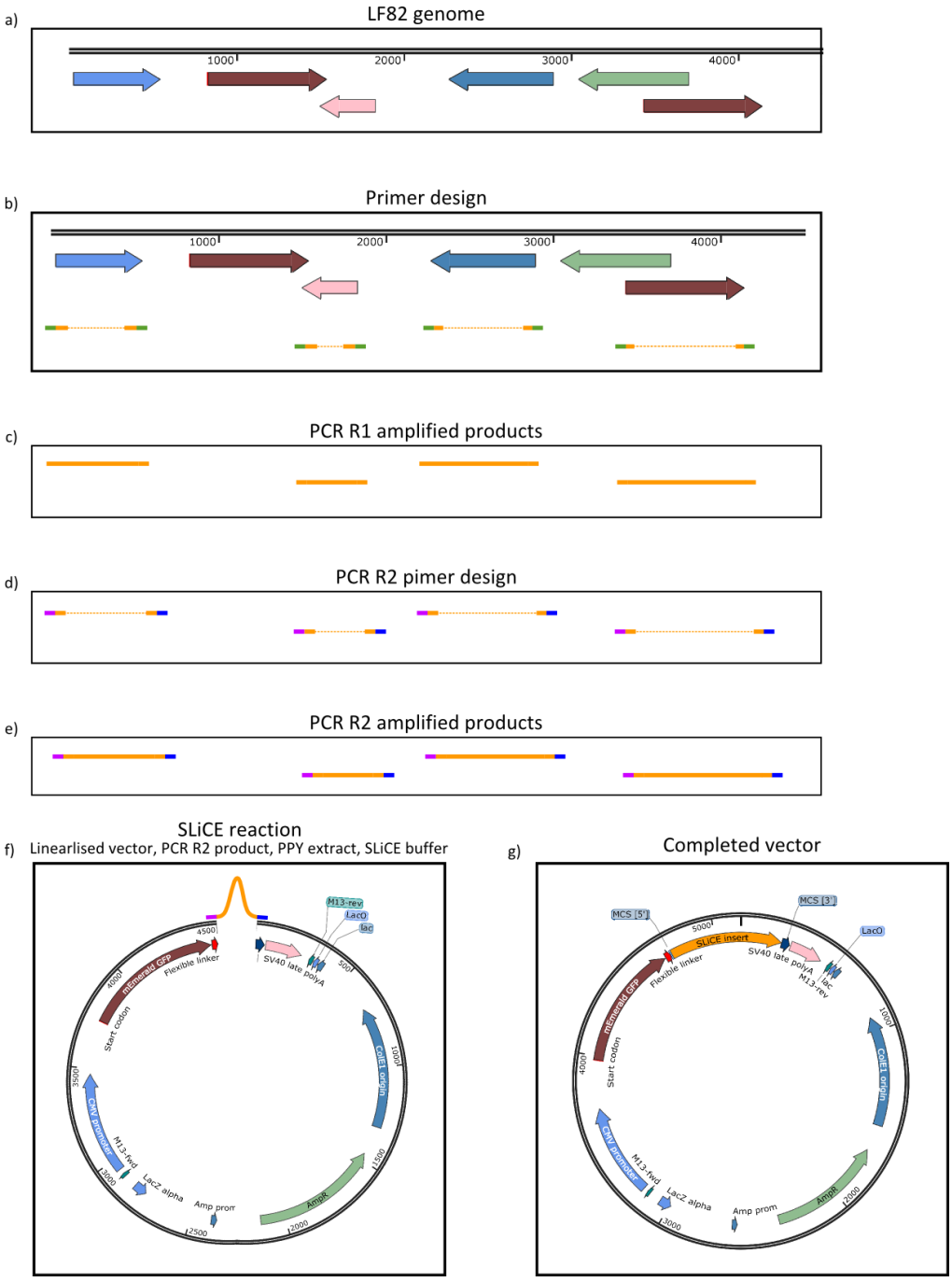

a)
LF82 genome
1000
2000
3000
4000
b)
Primer design
1000
2000
3000
4000
c)
PCR R1 amplified products
d)
PCR R2 pimer design
e)
PCR R2 amplified products
f)
SLiCE reaction
Linearised vector, PCR R2 product, PPY extract, SLiCE buffer
M13-rev
LacO
lac
SV40 late polyA
mEmerald GFP
Flexible linker
Start codon
ColE1 origin
CMV promoter
M13 fwd
LacZ alpha
Amp prom
AmpR
g)
Completed vector
MCS [5']
MCS [3']
SLiCE insert
SV40 late polyA
M13-rev
lac
LacO
mEmerald GFP
Flexible linker
Start codon
ColE1 origin
CMV promoter
M13 fwd
LacZ alpha
Amp prom
AmpR

Figure 2.4. **Overview of the SLiCE method for the insertion of multiple CDSs into a destination vector**. a) The LF82 genome containing several CDSs of interest. b) PCR primers were designed for each of these CDSs of interest using Primal.pl. c) Individual CDSs were amplified in a PCR R1 reaction. d) PCR primers were designed for each PCR R1 CDS using 'common' regions, added to each PCR product during PCR R1. e) PCR products from PCR R2 containing regions of homology to the desired vector insert region. f) The SLiCE reaction, in which each PCR R2 product was individually subjected to SLiCE via combination of: restriction digested vector, PCR R2 product for insertion, PPY extract, and SLiCE buffer. g) The completed SLiCE reaction, resulting in the insertion of each desired CDS into a specific site of a mammalian expression vector. When expressed, each CDS will be fused to mEmerald GFP via a flexible linker.

Chapter 3: The identification, cloning and transfection of a subset of LF82 CDSs

3.1 Adherent invasive *E. coli* (AIEC)

3.1.1 Background

Adherent invasive *E. coli* (AIEC) strain LF82 (hereafter referred to as LF82) is an AIEC type-strain initially isolated from a CD patient. It was originally identified by its enhanced ability to adhere to and invade epithelial cells in comparison to other *E. coli* strains (Boudeau et al. 1999). Although other strains of *E. coli* – such as Enteroinvasive *E. coli* (EIEC) – also display adherence and invasion phenotypes, these organisms typically have canonical virulence genes also found in *Shigella* and *Salmonella.* However, unlike these strains, LF82 appears to be devoid of any established *Shigella* or *Salmonella* virulence genes directly relating to bacterial invasion of mammalian cells (Boudeau et al. 1999; Miquel et al. 2010).

3.1.1.1 AIEC pathogenicity

Bacterial pathogenicity can be caused by a number of virulence factors. These factors can be secreted into bodily fluids, presented on the bacterial cell surface, or delivered directly into target cells. A common method for the delivery of a virulence factor into a target cell is via one of the bacterial secretion systems. A bacterial protein that is delivered into a target cell – bacterial, human or otherwise – using a secretion system is termed an 'effector' (Qiu et al. 2016). The most extensively studied effector delivery secretion system is the T3SS, however, the recently described T6SS has also been shown to deliver effectors (Büttner 2012; Russell et al. 2012; Russell et al. 2011).

Currently, little is known about LF82 effectors – particularly with regard to the strong adherence and invasive phenotypes displayed. However, the lack of any LF82-encoded T3SS and presence of two T6SSs, has been proposed to be linked to AIEC intracellular survival, although this is yet to be confirmed (Nash et al. 2010). This theory is supported by work by Massier et al. (2015) who have shown that LF82 T6SS deletions can result in defects in adhesion

and invasion with epithelial cells, but interestingly not in macrophages. Therefore, while evidence suggests a role of T6SS in AIEC adhesion and invasion, the effectors involved in this process or delivered via this mechanism are not yet known.

3.1.1.2 Previous AIEC screening techniques

In an attempt to better characterise AIEC and its virulence factors and/or effectors, previous attempts have been made to compare LF82 with other *E. coli* serotypes and related members of the *Enterobacteriaceae*. For example, Miquel et al. (2010) showed that LF82 is most closely related to avian pathogenic *E. coli* (APEC) due to the similarity of regions of genomic plasticity, and single nucleotide polymorphisms (SNPs) in virulence factor encoding genes. The technique used for this comparison utilised a TBLASTN-based homology score strategy for analysis, and successfully identified 115 LF82 CDSs that were not present in comparison strains.

The usefulness of a BLAST-based approach such as that used by Miquel et al. (2010) can be limited in certain situations. For example, given species A and B, if protein A1 reported high similarity to protein B1, it does not necessarily mean that protein B1 is most similar to A1; B1 may in actuality be most similar to protein A2. This limitation is largely overcome by the implementation of a reciprocal best BLAST hit (RBBH), where protein B1 would also undergo a BLAST to determine its similarity partner; only those proteins that hit against one another would then be considered as bona fide matches. However, an RBBH lacks the ability to identify inparalogs – coding sequences (CDSs) which were duplicated after the speciation event and therefore occur one or more times throughout a given genome. Thus the use of InParanoid – a software package which utilises the BLAST algorithm as well as its own inparalog identification algorithm – is commonplace for the identification of true orthologs and inparalogs (Ostlund et al. 2010; Remm et al. 2001).

3.1.1.3 InParanoid

The NCBI BLAST platform is a fast and indispensable tool for seeking sequence similarity between nucleotide and/or amino acid sequences, and can be used to successfully detect similarity between DNA or amino acid sequences. However, as described above, it is limited in its ability to differentiate orthologs, paralogs, out-paralogs and inparalogs for any genome or proteome not previously analysed using phylogenetic trees. Genes are said to be orthologs when they appear in divergent species but are predicted to have arisen from a common ancestor. Genes are said to be paralogs when they are duplicated within the genome of a species as a result of a gene duplication event. When paralogs occur pre-speciation, they are referred to as out-paralogs, and when they occur post-speciation, they are referred to as inparalogs (see Figure 3.1).

The ORFs of the proteomes of two strains from which orthologs/homologs are desired to be identified are submitted to InParanoid. From these, InParanoid then performs an all-against-all BLAST to detect similar ORFs between and within proteomes. For example, given strain A and strain B, a BLAST would be performed between A+B, B+A, A+A, and B+B. By using this method, high similarity ORFs can be identified between and within the submitted proteomes. These primary matches are then subjected to an RBBH to identify likely orthologs/paralogs in the opposing proteome. This method has the advantage of identifying ORFs closely-related to any RBBH hit that may otherwise be overlooked by an RBBH-only approach, thus including likely orthologs/homologs that would otherwise be omitted as a result.

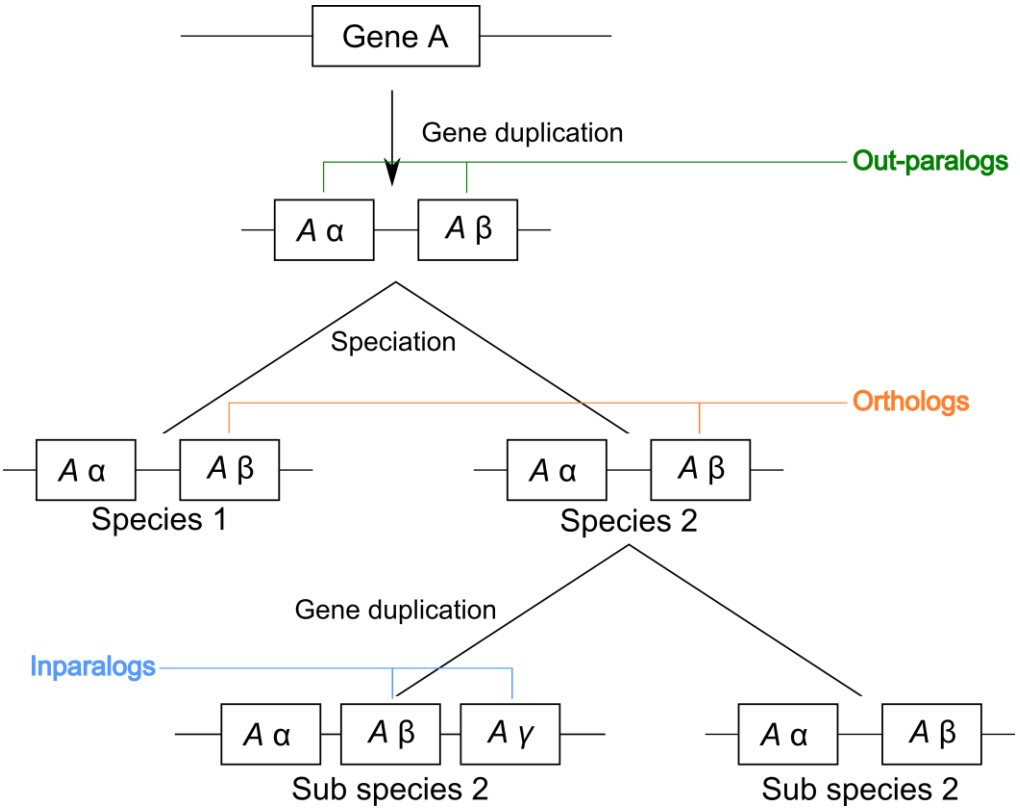


Figure 3.1. **Schematic showing differences between orthologs, inparalogs and out-paralogs**. Orthologs are homologs that are present in different organisms and result from a speciation event. Paralogs are homologs that occur as a result of gene duplication within a species. With relation to a particular species, paralogs can arise within (inparalogs) the current species of interest, or outside (out-paralog) the species of interest (in the parental species). An example of this can be seen here: If Species 2 is the organism of interest, paralogs occurring within sub species 2 are therefore inparalogs as they occur within the species of interest, whereas paralogs occurring outside the species of interest are therefore considered out-paralogs.

The InParanoid software was therefore utilised to analyse the ORFs of the LF82 proteome, comparing them with ORFs of the proteomes of other *E. coli* strains (see below). A subset of LF82 ORFs was produced and subjected to a high-content screen to search for phenotypes associated with pathogenicity. This is the first time that LF82 ORFs have been analysed this way, and have been subjected to a high-content screen like this.

3.2 Aims

The comparison of LF82 and other *E. coli* strain genomes will allow the identification of genes unique to LF82 (within this subset). Fusion of these 'unique genes' to a GFP reporter will allow their co-localisation to be monitored after expression in a HeLa reporter cell line. Automated microscopy will be used to acquire multiple images from each sample site, allowing the downstream phenotypic analysis of each expressed CDS. The aims of this chapter were to:

- Compare the genome of AIEC type-strain LF82 to that of several commensal and pathogenic *E. coli* strains, allowing the identification of CDSs unique to LF82.
- Create a vector library for the entire 'LF82 unique gene' set, to allow expression of each CDS as a GFP-CDS fusion.
- Create a HeLa cell reporter cell line expressing mCherry-LC3 for the downstream analysis of autophagy effectors.
- Transfect and express library vectors in the HeLa mCherry-LC3 reporter cell line.
- Perform automated microscopy for image collection, from which downstream phenotypic analysis will be performed.

3.3 Workflow of LF82 CDS selection and cloning for library creation

Table 3.1 shows an overview of the workflow used to identify and clone LF82 unique genes, along with the success rates of these steps and their subsequent addition to a GFP-tagged vector library.

Table 3.1. **Workflow of Chapter 3**. The identification, PCR amplification and cloning of LF82 unique genes into the pCMV-mEm-4GS vector.

Workflow of screen preparation		
Step	**Result**	**Notes**
Identification of unique LF82 genes using InParanoid	225 CDSs identified to include in library	
Comparison of gene set to that of Miquel et al. (2010)	10 extra CDSs added to library. 235 LF82 unique genes total	
Round 1 PCR	170 appeared successful (72%)	
Round 2 PCR	180 appeared successful (77%)	
Colony PCR after SLiCE reaction	157 identified as highly likely to have undergone SLiCE successfully (67%)	
Round 1, Round 2 and Colony PCRs reattempted with currently unsuccessful CDSs	59 more CDSs identified as likely to have undergone SLiCE successfully. Total: 216 CDSs (92%)	
Sequencing used to confirm successful SLiCE CDSs	170 CDSs confirmed as having undergone SLiCE successfully (72%)	
Final attempt at Round 1, Round 2 and Colony PCR for SLiCE (Set 4)	42 CDSs appeared successful	Included as a separate 'set': Set 4. (See: Revisiting SLiCE failures for the creation of Set 4)
Sequencing used to confirm Set 4 SLiCE	39 CDSs confirmed as hacing undergone SLiCE successfully	
Total number of CDSs cloned into library:		**209/235 (89%)**

3.4 *E. coli* genome comparisons

3.4.1 *E. coli* serotype selection

Several *E. coli* serotypes were selected for comparison with AIEC LF82; these consisted of: EHEC O111:H$^-$, EPEC O127:H6, UPEC 536, K12 DH10B, K12 MG1655 and K12 W3110. These strains were chosen based upon sequence availability, strain pathogenicity and/or commensal status, strain similarity to AIEC LF82, and diversity of non-core gene content.

The pathogenic strain genomes used – EHEC O111:H$^-$, EPEC O127:H6 and UPEC 536 – were selected to ensure that the final LF82 unique gene set (see below) did not contain any orthologs or inparalogs of EHEC, EPEC or UPEC virulence genes. This was performed in order to help identify virulence factors likely to be unique to LF82. The more closely related APEC_O1 was not utilised for this comparison step due to fears that it would result in too large a reduction in putative LF82 effectors identified. APEC_O1 was, however, utilised in comparisons made by Miquel et al. (2010) suggesting that these fears were perhaps misguided. The non-pathogenic *E. coli* K12 strains – K12 DH10B, K12 MG1655 and K12 W3110 – were utilised primarily to identify and remove any orthologs of core *E. coli* CDSs from the final CDS set.

Data for each *E. coli* serotype was downloaded as a GenBank file from the NCBI FTP server (ftp://ftp.ncbi.nlm.nih.gov/). Accession numbers can be found in the section: Identification of AIEC CDS for analysis.

3.4.2 InParanoid genome analysis

Proteomes for all *E. coli* serotypes were extracted as FASTA files from respective GenBank files using the Artemis genome browser (Carver et al. 2012). The offline version of InParanoid 4.1 software developed by O'Brien et al. (2005) was then employed to interrogate the *E. coli* serotype proteomes. The purpose of this comparison was to obtain a subset of LF82

coding sequences (CDSs) which did not appear to be orthologs or inparalogs to the CDSs of the other comparison serotypes, thus identifying CDSs likely specific to LF82. The LF82 subset identified was therefore considered as CDS set 'unique' to LF82 when compared to the analysed proteomes, and is therefore routinely referred to as a "LF82 unique gene" subset.

InParanoid 4.1 was designed to compare two proteomes at once, and, as such, it was necessary to use it in a manually iterative manner. To do this InParanoid was supplied with the LF82 proteome and the proteome of one of the comparison strains to obtain a list of orthologous CDSs. The CDSs contained on this list were then removed from the LF82 proteome data file, and the adjusted LF82 proteome resubmitted with the next comparison serotype (Figure 3.2). Once this had completed all iterations, considering each comparison strain, 225 CDSs remained that were only observed in LF82. A list of these CDSs can be seen in Appendix II: LF82 unique genes.

3.4.2.1 Confirmation of LF82 unique gene subset

To ensure accuracy of these CDSs, they were subjected to a BLASTP search against the comparison strain proteomes, which showed no significant results. This suggests that members of the LF82 unique gene list are not present in the other strains tested, indicating that the use of InParanoid for removal of orthologous genes was successful.

Although identifying the source of any individual gene unique to LF82 was beyond the scope of the current study, it is likely that LF82 previously acquired these CDSs via either gene duplication and diversification or through horizontal gene transfer from other organisms outside of the comparison set used here. The former involves the initial duplication of a gene, which may then undergo a wide range of mutations, altering its function. Marques et al. (2008) give the example of a gene duplication event which gives rise to a redundant copy of a gene, providing the host with twice the amount of the associated protein, potentially conferring a fitness advantage. Horizontal

gene transfer, on the other hand, is the acquisition of DNA from another bacterium which is often maintained if it confers a fitness advantage for the new host.

It may be difficult to determine the source of any one gene if it has been present in the organism of interest for some time as, over time, the percentage GC content (which varies between organisms) can be shifted towards that of the host organism (Lawrence & Ochman 1997). It is likely that the unique genes identified from LF82 here are a mixture of duplication-derived and horizontally transferred genes.

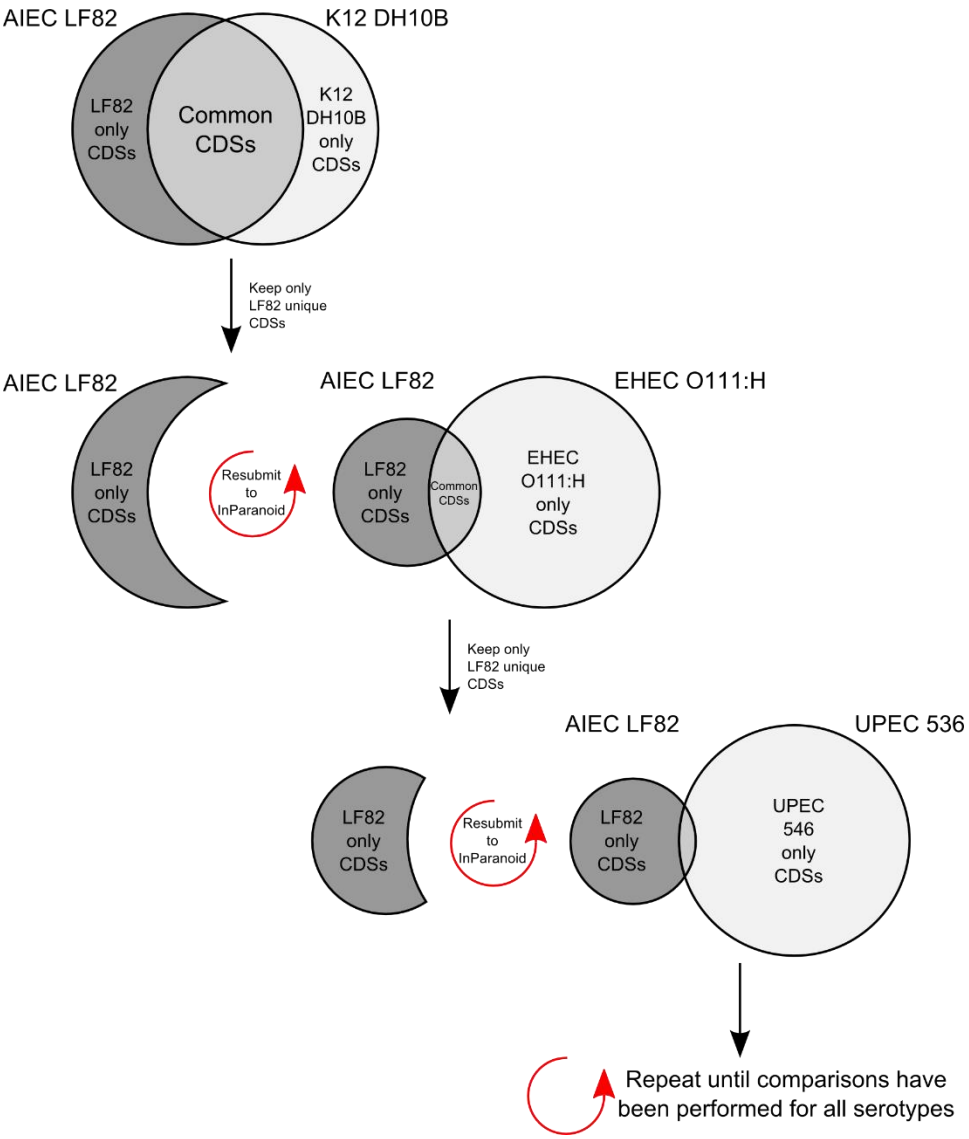


Figure 3.2. **Schematic diagram illustrating the manual iterations performed for proteomic comparisons between LF82 and the comparison strains**. 'Common CDSs' refer to CDSs determined as orthologs and inparalogs by InParanoid, which were removed from the LF82 proteome after each iteration. Remaining LF82 CDSs were then combined into a new 'LF82 only CDSs' subset before resubmitting this to InParanoid for comparison to the next *E. coli* strain.

The LF82 unique gene list was also compared to a list compiled by Miquel et al. (2010) (supplementary table S2 of the referenced article) which contained genes identified as unique to LF82, compiled via an alternative strategy. This comparison utilised a TBLASTN-based approach (discussed earlier) that is less robust but easier to perform. The number of *E. coli* strains used for comparison was therefore higher than the that used here – a full list of strains used for comparison by Miquel et al. (2010) can be seen in Table 3.2.

Any genes present in the LF82 unique gene list obtained from Miquel et al. (2010) that were not present in the current list, were added. This resulted in the addition of 10 CDSs to the current LF82 unique gene list, resulting in a total of 235 genes identified as unique to LF82. This entire gene set was subjected to a nucleotide BLAST against AIEC strain NRG857C, showing that 204 of the 235 CDSs were also present in NRG857C. This suggests that the LF82 unique gene list is representative of the AIEC pathovar as whole, although without comparison to the other AIEC strains (whose genomes are not available), this cannot be confirmed.

A full list of these CDSs can be seen in Appendix II: LF82 unique genes, along with those identified both in this study and by Miquel et al. (2010). This list comprised the entire CDS subset used for this research.

Table 3.2. ***E. coli* strains used by Miquel et al. (2010)**. These strains were used for comparison to LF82 genomes using a BLAST-based approach. Strains also used in the current research are highlighted in bold.

Strain	Pathovar	Genome size (bp) and site of infection
536	**Uropathogenic *E. coli* (UPEC)**	**4938920 Urinary tract/Urethra**
UTI89	UPEC	5065741 Urinary tract/Urethra
CFT073	UPEC	5231428 Urinary tract/Urethra
APEC_O1	Avian Pathogenic *E. coli* (APEC)	5082025 Respiratory tract
E2348/69 (O127:H6)	**Enteropathogenic *E. coli* (EPEC)**	**4965553 Intestinal mucosa**
S88	Meningitis-associated *E. coli* (MNEC)	5032268 Digestive tract → Meninges
UMN026	Extra-intestinal pathogenic *E.* coli (ExPEC)	5202090 Urinary tract/Urethra
IAI39	ExPEC	5132068 Urinary tract/Urethra
SMS_3_5	Environmental	5068389 NA
EDL933	Enterohemorrhagic *E. coli* (EHEC)	5528445 Intestinal mucosa
157_H7	EHEC	5498450 Intestinal mucosa
EC4115	EHEC	5572075 Intestinal mucosa
E24377A	Enterotoxigenic *E. coli* (ETEC)	4979619 Ilial mucosa
55989	Enteroaggregative *E.* coli (EAEC)	5154862 Intestinal mucosa
IAI1	Commensal	4700560 NA
ED1a	Commensal	5209548 NA
SE11	Commensal	4887515 NA

Strain	Pathovar	Genome size (bp) and site of infection
ATCC 8739	Commensal	4746218 NA
HS	Commensal	4643538 NA
DH10B	**K12**	**4686137** **NA**
MG1655	**K12**	**4639675** **NA**
W3110	**K12**	**4646332** **NA**

3.4.3 Analysis of LF82 unique genes

3.4.3.1 Prediction of T6SS secretion signals

The LF82 unique gene set was searched to identify any putative signal sequences for T6SS secretion. PAAR (proline-alanine-alanine-arginine) motifs were considered along with hRxGhhYhh motifs (where h represents a hydrophobic residue) – the so-called MIX motif (marker for type 6 effectors) (Salomon et al. 2014). However, no motifs were found to suggest that any of the LF82 unique genes identified were secreted via the T6SS. However, type 6 secretion of any LF82 unique genes cannot be ruled out, as the current understanding of type 6 secretion signals is poor, and other secretion signals almost certainly remain to be discovered.

SecretomeP, an online tool for the identification of secreted proteins through 'non-classical' methods (including the T6SS), described by Bendtsen et al. (2005), was also employed to analyse the LF82 unique gene set. However, again no hits were found, suggesting that either none of the genes identified are secreted via the non-classical methods which SecretomeP considers, or that secretion signal data is not complete enough to identify them.

3.4.3.2 LF82 unique genes compared to regions of known genomic plasticity

The core bacterial genome consists of the genes necessary for basic survival and growth. In order to adapt to different niches, however, flexibility within the genome exists, allowing some genetic elements to be readily lost, shared or duplicated. These regions of hypervariability are known as regions of genomic plasticity, and, due to their flexible nature, are those most likely to undergo alterations, either via deletion, inversion, duplication or insertion (Ogier et al. 2010). Several mechanisms may be responsible for the hypervariability of particular sites within any given genome, including the presence of mobile genetic elements, such as transposons, or the presence of a particular site of preference for phage integration, such as the *attB* locus for λ phage (Shao et al. 2015).

Depending upon the source of the DNA in any one region of genomic plasticity, the GC content may vary due to the differing GC content of the originating organism. Thus, the GC content, paired with bioinformatic information, is a useful method for the determination of regions of genomic plasticity, and may be used to predict the presence of novel pathogenicity islands (PAIs) where clusters of virulence genes have been transferred from another organism and integrated into the host genome (Hacker & Kaper 2000).

A list of LF82 regions of genomic plasticity was obtained from Miquel et al. (2010) and their locations mapped onto the LF82 genome alongside the LF82 unique genes. Mapping was performed using CGView, and can be seen in Figure 3.3 (Grant & Stothard 2008). Where PAIs are located, an alteration in GC can also often be seen, suggesting that the PAI was obtained from an external source (other bacteria, phage etc.). An example of this can be seen with PAI-2 in Figure 3.3, where GC content varies from the genomic average considerably. Many of the LF82 unique genes identified align with either regions of genomic plasticity or PAIs, supporting the authenticity of these genes as 'unique' to LF82 as they are likely not part of the core *E. coli* genome, and originate from other organisms.

It is interesting to note that there are several small clusters of LF82 unique genes which do not align with known regions of genomic plasticity. One such region lies with the 4600-4700 kbp sector (see Figure 3.3, pink arrow). This region also appears to display a lower GC content than surrounding areas, perhaps suggesting that it is a previously unidentified region of genomic plasticity and/or the DNA in this region was obtained from an external source.

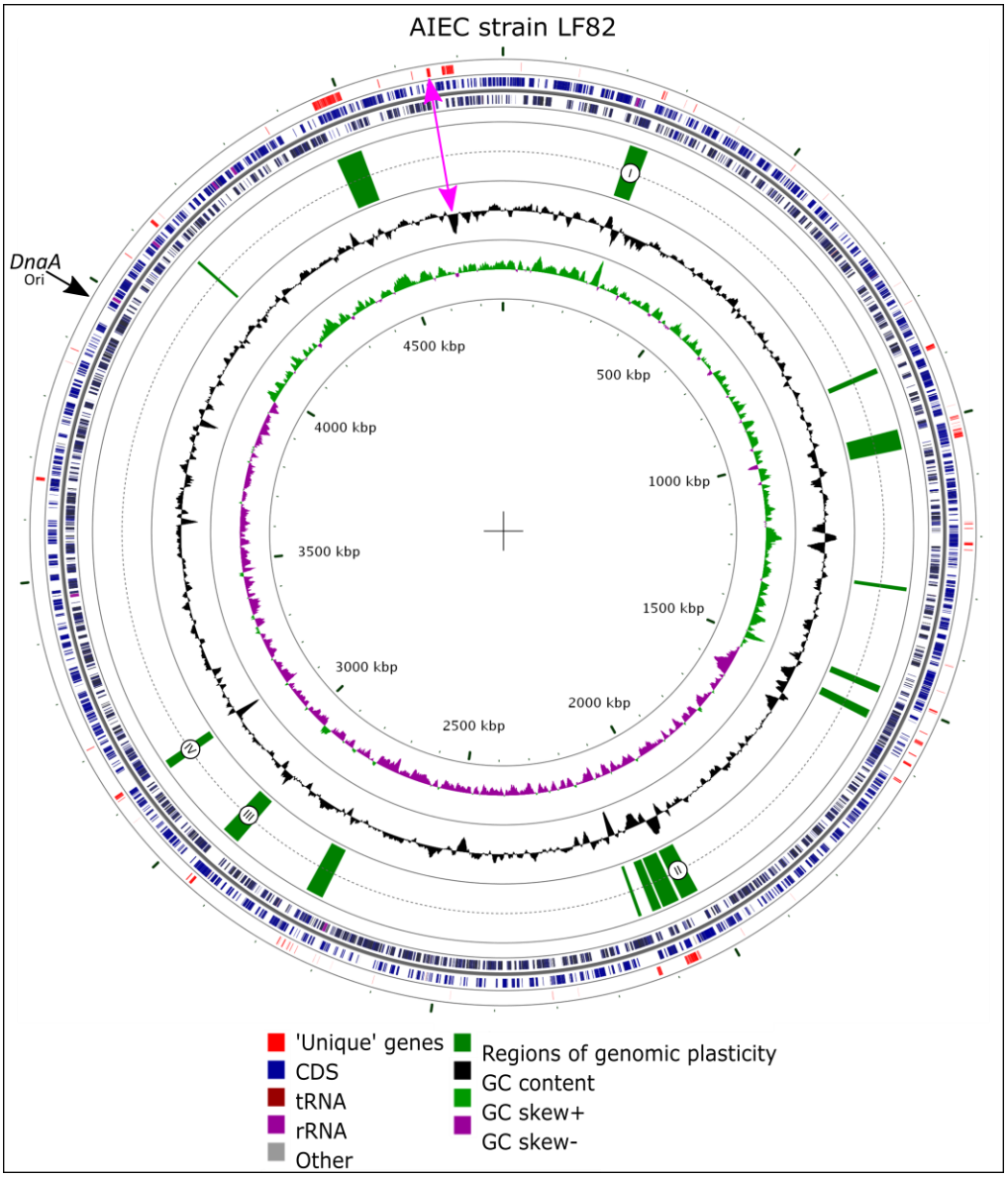


Figure 3.3. **LF82 unique genes mapped onto the LF82 genome**. This map displays, from inside to outside: GC skew, GC content, regions of genomic plasticity and PAIs (identified by Miquel et al. (2010)), LF82 CDSs, and LF82 unique genes identified during this research. Predicted PAIs are labelled with their designated number (PAI I, II, III and IV). Note the change in GC content present for the PAIs, suggesting that they were acquired from another organism. Genes identified as unique to LF82 in the current research frequently cluster in alignment with regions of genomic plasticity. However, small clusters exist, such as that within the 4600-4700 kbp sector (pink arrow), which have no associated region of genomic plasticity. The GC content for this region may, however, suggest that these genes were acquired from another organism. DnaA indicates the origin of replication. Figure created using CGView (Grant & Stothard 2008).

3.5 Preparation of LF82 library for analysis

3.5.1 Preparation of LF82 unique genes for SLiCE

To analyse the LF82 unique gene set, each gene was cloned into an expression vector as an mEmerald-GFP (commonly referred to as mEm, or GFP) fusion. To do this, all genes were amplified via two PCRs. PCR R1 was used to amplify each gene from the LF82 genome. PCR R2 was used to add extensions onto each amplified gene CDS, necessary for the SLiCE reaction (see below). While this cloning is possible with one single PCR for each gene, the oligonucleotide length would be >60 bp, requiring an increased synthesis scale from IDT, resulting in a substantial increase in cost.

Due to the number of genes in the LF82 unique gene set, primers were ordered in three 96-well plates which were referred to frequently as 'Set 1', 'Set 2', and 'Set 3'. A previous version of the iterative InParanoid strategy erroneously resulted in the identification of non-unique genes, and some of these were subsequently taken forward to act as a 'control' set as part of Set 3.

A fourth set, 'Set 4', was also created at a later time. This set consisted of those members of the LF82 unique gene set which failed PCR R1, PCR R2, SLiCE or final confirmation by sequencing. See: Revisiting SLiCE failures for the creation of Set 4, for more details.

3.5.1.1 PCR R1

PCR primers were designed to amplify each individual CDS in the LF82 unique gene set from the LF82 genome. During amplification, primers simultaneously added extensions for utilisation in PCR R2 (below).

PCR products were subjected to gel electrophoresis to help validate amplification of the correct genes. Gel images from Set 2 PCR R1 can be seen

in Figure 3.4 a, and show apparent successful amplification of 66 out of 96 attempted amplifications – a success rate of 69%.

The range of gene sizes in the LF82 unique gene set varied from 105 bp to 3903 bp, which presented several challenges for PCR amplification, gel electrophoresis and DNA purification. To accommodate amplification of the longest ORFs, the PCR R1 extension step was increased over the recommended time. While there was potential for this to introduce nonspecific amplification for smaller products, DNA sequencing was used to ensure correct sequence presence before use in the transfection screen.

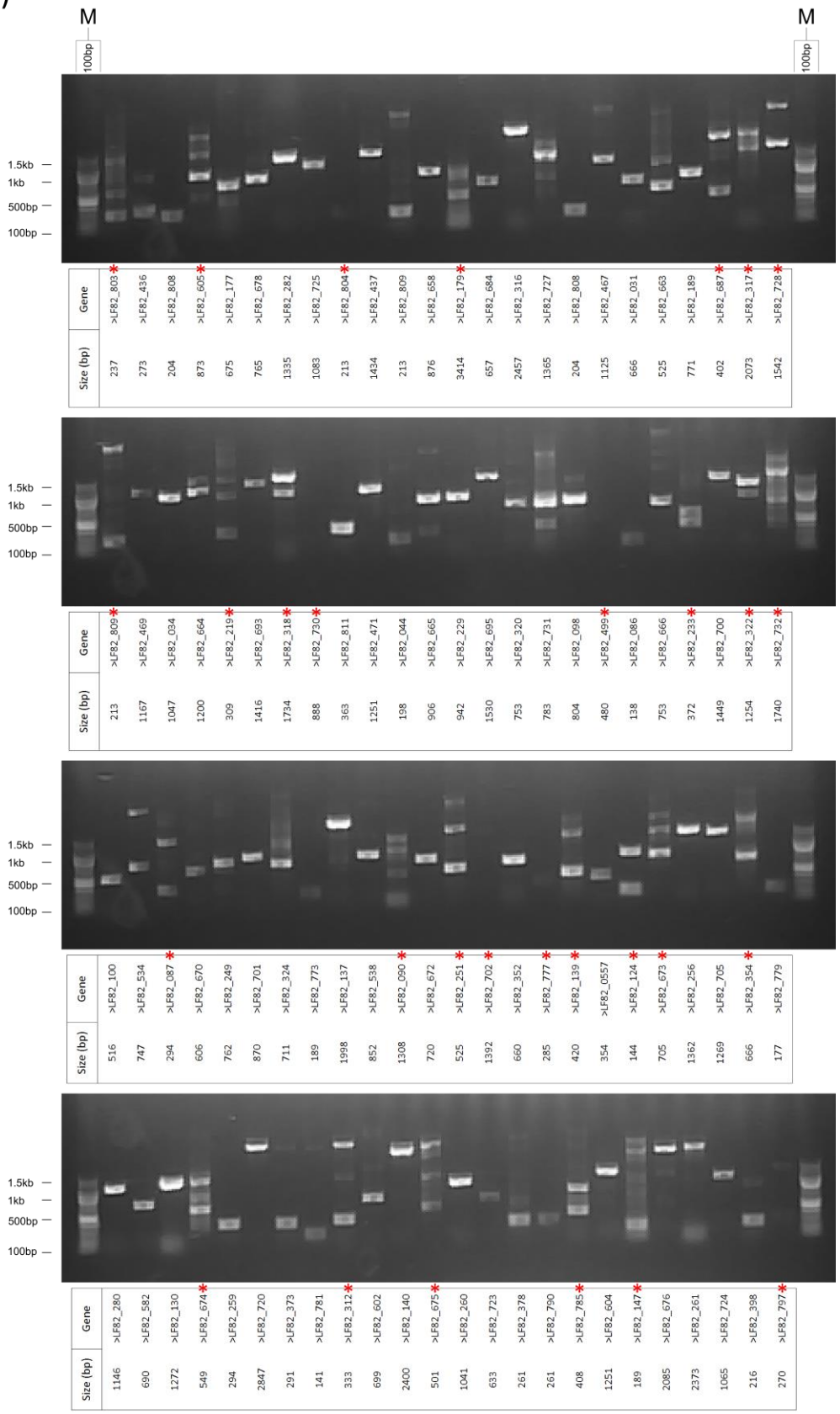
The products resulting from smaller CDSs in the LF82 unique gene set became increasingly difficult to identify after PCR R1 due to their size. However, after the PCR R2 step (below), these products increased in size by 52 bp, allowing a marginal increase in ease of analysis. In addition to this, the colony PCR performed post-SLiCE (see later) and transformation provided further evidence of the amplification success of these CDSs.

3.5.1.2 PCR R2

PCR R2 was used to prepare amplified genes from PCR R1 for the SLiCE reaction. Primers were designed to contain homology to the extensions added to amplified genes in PCR R1, thus only two PCR R2 primers were required (forward and reverse).

Extensions added during PCR R2 provided homology regions for the restriction digested destination vector, pCMV-mEm-4GS. PCR R2 therefore, once complete, resulted in a library of LF82 unique gene CDSs ready for the SLiCE reaction, and insertion into pCMV-mEm-4GS. Gel images from Set 2 PCR R2 can be seen in Figure 3.4 b, and show apparent successful amplification of 78 out of 96 attempted amplifications – a success rate of 81%.

a)



Gene	Size (bp)
>LF82_803*	237
>LF82_436	273
>LF82_808	204
>LF82_605*	873
>LF82_177	675
>LF82_678	765
>LF82_282	1335
>LF82_725	1083
>LF82_804*	213
>LF82_437	1434
>LF82_809	213
>LF82_658	876
>LF82_179*	3414
>LF82_684	657
>LF82_316	2457
>LF82_727	1365
>LF82_808	204
>LF82_467	1125
>LF82_031	666
>LF82_663	525
>LF82_189	771
>LF82_687*	402
>LF82_317*	2073
>LF82_728*	1542

Gene	Size (bp)
>LF82_809*	213
>LF82_469	1167
>LF82_034	1047
>LF82_664	1200
>LF82_219*	309
>LF82_693	1416
>LF82_318*	1734
>LF82_730*	888
>LF82_811	363
>LF82_471	1251
>LF82_044	198
>LF82_665	906
>LF82_229	942
>LF82_695	1530
>LF82_320	753
>LF82_731	783
>LF82_098	804
>LF82_499*	480
>LF82_086	138
>LF82_666	753
>LF82_233*	372
>LF82_700	1449
>LF82_322*	1254
>LF82_732*	1740

Gene	Size (bp)
>LF82_100	516
>LF82_534	747
>LF82_087*	294
>LF82_670	606
>LF82_249	762
>LF82_701	870
>LF82_324	711
>LF82_773	189
>LF82_137	1998
>LF82_538	852
>LF82_090*	1308
>LF82_672	720
>LF82_251*	525
>LF82_702*	1392
>LF82_352	660
>LF82_777*	285
>LF82_139*	420
>LF82_0557	354
>LF82_124*	144
>LF82_673*	705
>LF82_256	1362
>LF82_705	1269
>LF82_354*	666
>LF82_779	177

Gene	Size (bp)
>LF82_280	1146
>LF82_582	690
>LF82_130	1272
>LF82_674*	549
>LF82_259	294
>LF82_720	2847
>LF82_373	291
>LF82_781	141
>LF82_312*	333
>LF82_602	699
>LF82_140	2400
>LF82_675*	501
>LF82_260	1041
>LF82_723	633
>LF82_378	261
>LF82_790	261
>LF82_785*	408
>LF82_604	1251
>LF82_147*	189
>LF82_676	2085
>LF82_261	2373
>LF82_724	1065
>LF82_398	216
>LF82_797*	270

Key:

* Unexpected fragment size or presence of multiple fragments

b)

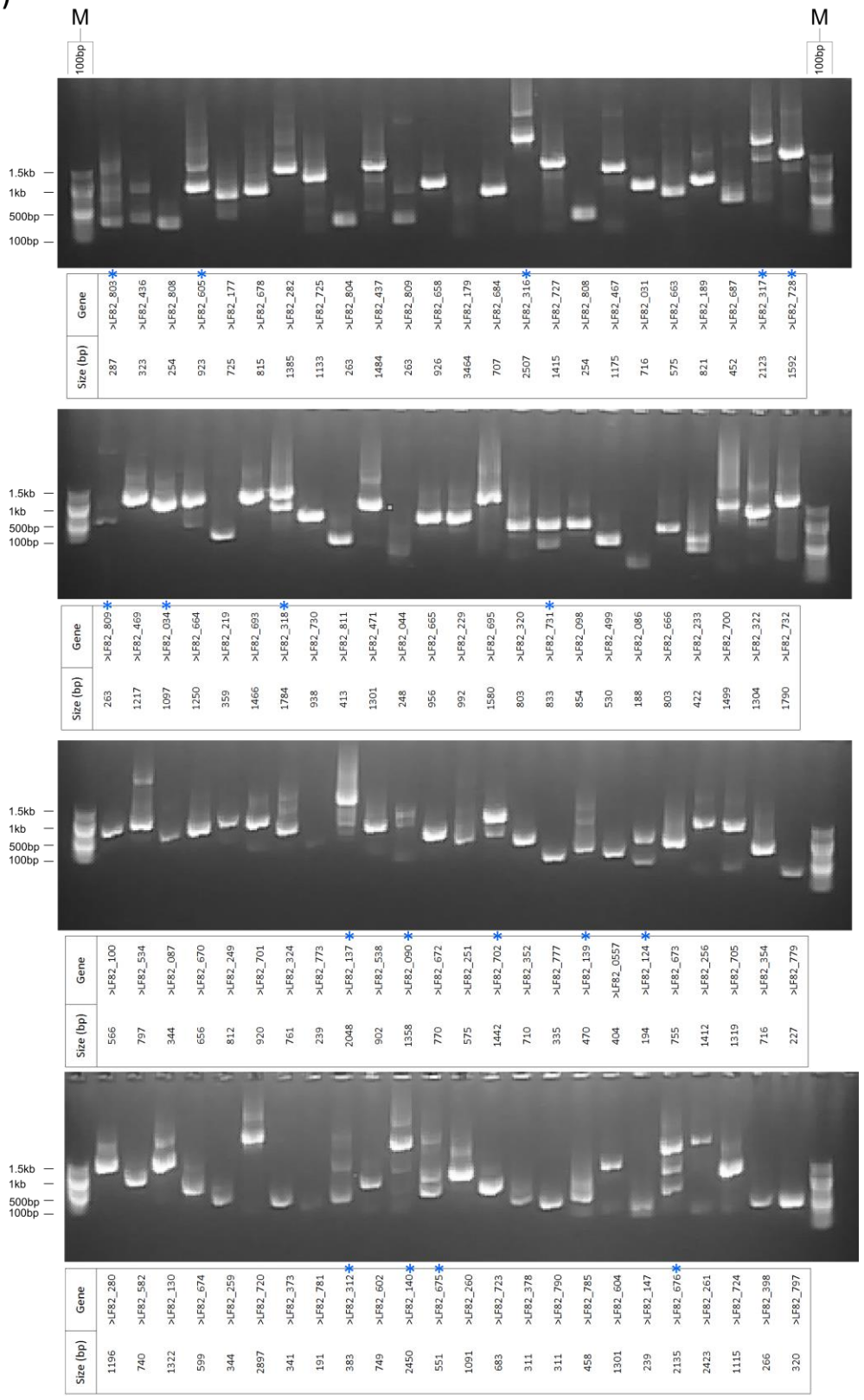

M
100bp
M
100bp
1.5kb
1kb
500bp
100bp
Gene
Size (bp)
>LF82_803 287
>LF82_436 323
>LF82_808 254
>LF82_605 923
>LF82_177 725
>LF82_678 815
>LF82_282 1385
>LF82_725 1133
>LF82_804 263
>LF82_437 1484
>LF82_809 263
>LF82_658 926
>LF82_179 3464
>LF82_684 707
>LF82_316 2507
>LF82_727 1415
>LF82_808 254
>LF82_467 1175
>LF82_031 716
>LF82_663 575
>LF82_189 821
>LF82_687 452
>LF82_317 2123
>LF82_728 1592
>LF82_809 263
>LF82_469 1217
>LF82_034 1097
>LF82_664 1250
>LF82_219 359
>LF82_693 1466
>LF82_318 1784
>LF82_730 938
>LF82_811 413
>LF82_471 1301
>LF82_044 248
>LF82_665 956
>LF82_229 992
>LF82_695 1580
>LF82_320 803
>LF82_731 833
>LF82_098 854
>LF82_499 530
>LF82_086 188
>LF82_666 803
>LF82_233 422
>LF82_700 1499
>LF82_322 1304
>LF82_732 1790
>LF82_100 566
>LF82_534 797
>LF82_087 344
>LF82_670 656
>LF82_249 812
>LF82_701 920
>LF82_324 761
>LF82_773 239
>LF82_137 2048
>LF82_538 902
>LF82_090 1358
>LF82_672 770
>LF82_251 575
>LF82_702 1442
>LF82_352 710
>LF82_777 335
>LF82_139 470
>LF82_0557 404
>LF82_124 194
>LF82_673 755
>LF82_256 1412
>LF82_705 1319
>LF82_354 716
>LF82_779 227
>LF82_280 1196
>LF82_582 740
>LF82_130 1322
>LF82_674 599
>LF82_259 344
>LF82_720 2897
>LF82_373 341
>LF82_781 191
>LF82_312 383
>LF82_602 749
>LF82_140 2450
>LF82_675 551
>LF82_260 1091
>LF82_723 683
>LF82_378 311
>LF82_790 311
>LF82_785 458
>LF82_604 1301
>LF82_147 239
>LF82_676 2135
>LF82_261 2423
>LF82_724 1115
>LF82_398 266
>LF82_797 320
Key:
* Unexpected fragment size or presence of multiple fragments

Figure 3.4. **Gel electrophoresis of Set 2 LF82 unique genes**. Gels show PCR products after a) R1 PCR amplification of the target gene from the LF82 genome, and b) R2 PCR amplification of R1 PCR products to prepare genes for the SLiCE reaction. Asterisks represent PCR amplifications predicted to have failed.

3.5.2 Seamless ligation in cell extract (SLiCE)

SLiCE is a novel cloning method developed by Zhang et al. (2012) that utilises bacterial cell extract to induce homologous recombination *in vitro*. A wide variety of bacterial strains – typically *E. coli* – that express the λ Red recombination system could potentially be used for SLiCE and the recombination of a desired fragment into a desired vector location. Here, however, the PPY strain was utilised (courtesy of Yongwei Zhang, Albert Einstein College of Medicine, New York, USA). This strain contains an optimised λ Red recombination system, with *redβ* and *gam* constitutively expressed, and *redα* under arabinose control.

When the PPY strain is subjected to arabinose induction and cells harvested for their cytoplasmic contents, the resulting extract therefore contains high levels of the λ Red recombinase machinery. When subsequently added to a mixture of DNA, homologous regions undergo recombination. This system is ideal for high-content screens such as the one presented here; SLiCE is incredibly cost effective, has good efficiency, and is easy to utilise for large numbers of cloning reactions.

To create an mEmerald-GFP (mEm-GFP) LF82-gene fusion, the pCMV-mEm-4GS vector was linearised using EcoRV-HF, gel purified, and combined in a SLiCE reaction with the PCR R2 product corresponding to each LF82 unique gene. The extensions added to each unique gene using PCR R2 underwent recombination during the SLiCE reaction. This resulted in the insertion of the respective unique gene into the pCMV-mEm-4GS vector at the Eco-RV-HF cut site in the multiple cloning site (MCS) (see Figure 3.5). A library of vectors was therefore created, containing each LF82 unique gene fused to mEm-GFP via a flexible linker at their C-terminal ends. The aim of the flexible linker was to allow a higher degree of movement for the fused protein, thus reducing interference with potential interactions or function/s (Chen et al. 2013).

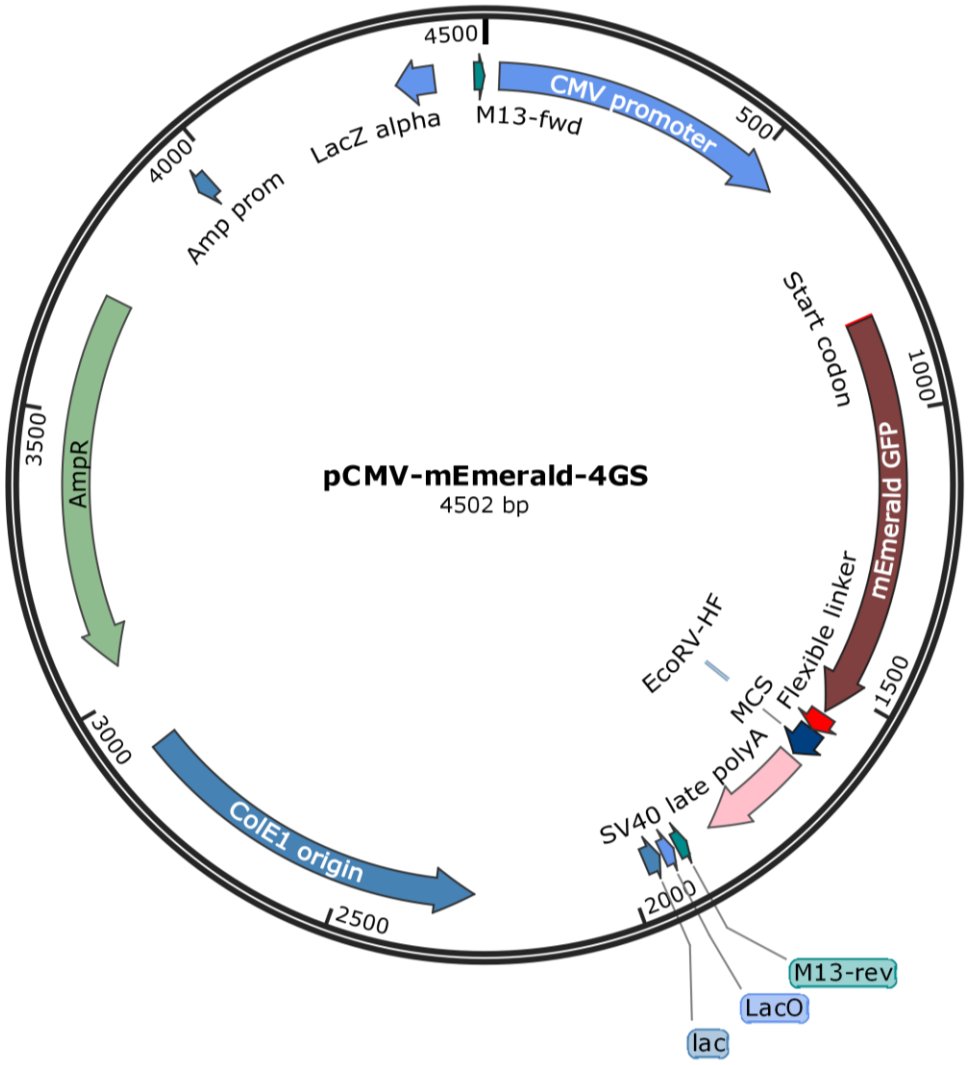


Figure 3.5. **pCMV-mEm-4GS vector map**. This map shows mEmerald GFP on the right with the flexible linker below (brown and red, respectively). LF82 unique genes were inserted – using SLiCE – in the multiple cloning site (MCS) at the Eco-RV-HF cut site. See Figure 2.4 (in Seamless Ligation in Cell Extract (SLiCE)) for more information.

3.5.3 Transformation of library DNA

Transformation of the entire SLiCE library of vectors was performed using chemical transformation of XL1-blue *E. coli*. However, to limit the time that this took – and to limit the time taken in future high-content experiments – a bead-guide was designed for high-throughput spreading.

3.5.3.1 3D bead-guide design

The online 3D design software, TinkerCad (https://www.tinkercad.com/) was used to design a rectangular guide to fit inside a rectangular 'agar plate' (made via the pouring of agar into an OmniTray (Nunc)) (Figure 3.6 a). Prototypes of the bead-guide were produced using 3D printing to ensure the final product would function as desired. While 3D printing is relatively inexpensive, there are restrictions on which plastics can be used, making sterilisation of equipment difficult. Final bead-guides were therefore milled from PTFE – a plastic which can be ethanol treated, and even autoclaved if desired (although the repeated use of the latter is not recommended) – and can be seen in Figure 3.6 b and c.

Bead-guide usage enabled 24 XL1-blue transformations to be spread at once, reducing workload and saving time. Glass beads were dispensed into and rolled along each lane, spreading transformation broth as they travelled. From the 20 µL of transformation broth applied to each bead, an average of 25 transformant colony forming units (CFU) were observed per lane.

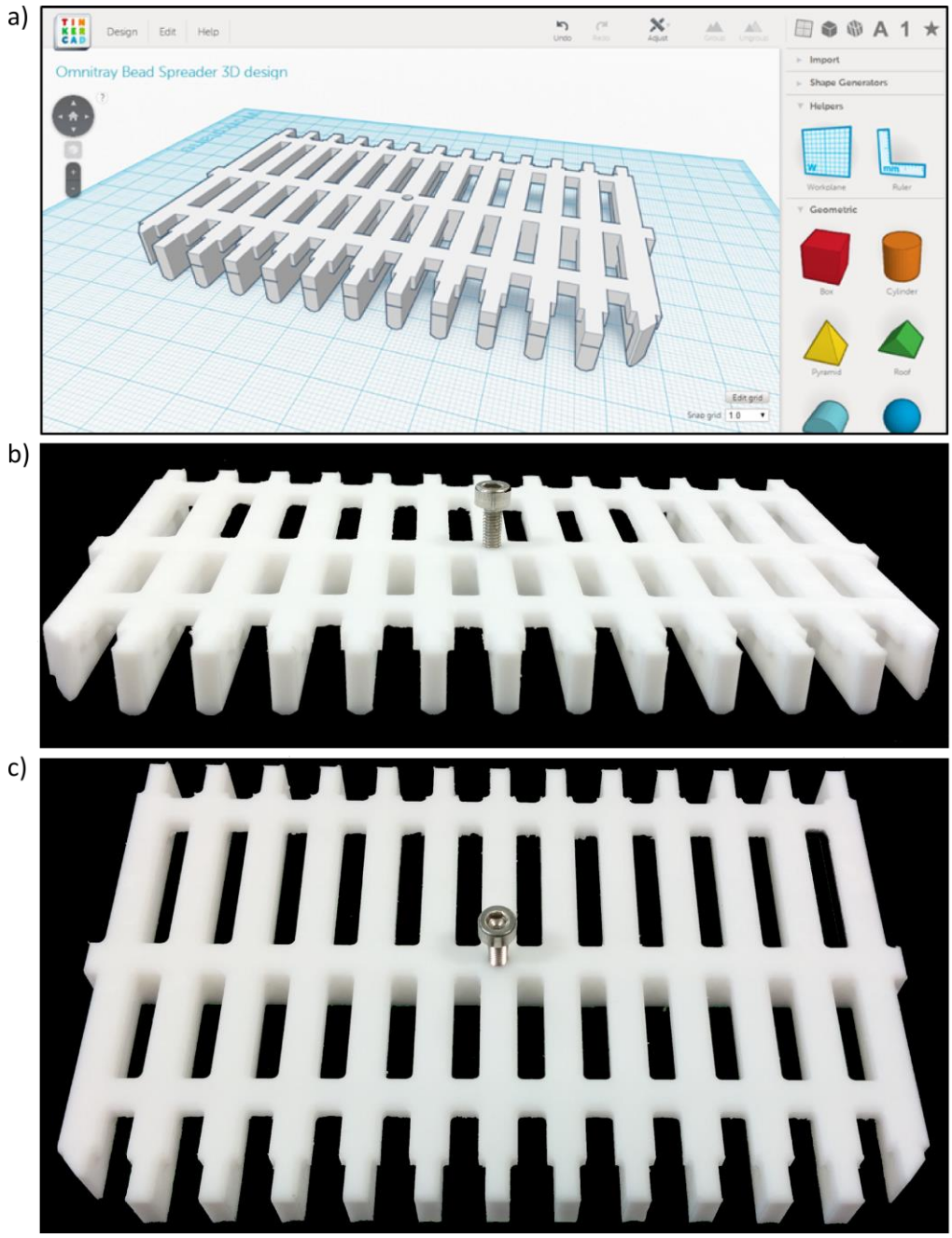


Figure 3.6. **Custom designed bead-guide**. The bead-guide was used for spreading transformation broths in a high-throughput manner via glass beads in each lane. a) 3D design software, TinkerCad (https://www.tinkercad.com/) was used to design the spreader. b) and c) show the final product milled from PTFE, complete with screw to enable direct placement and removal from rectangular agar plates.

3.5.3.2 Cross contamination as a result of bead-guide usage

During the colony PCR step (see below), little-to-no cross-contamination between lanes was observed. Bands present on the electrophoresis gels displayed no obvious signs of cross-contamination, with predicted sizes of amplified fragments showing no resemblance to their neighbouring wells (see Figure 3.7).

3.5.4 Confirmation of successful LF82 unique gene insertion into pCMV-mEm-4GS

To confirm that each LF82 unique gene was successfully inserted into the pCMV-mEm-4GS vector via SLiCE, two methods were utilised: agarose gel electrophoresis, and Sanger sequencing. Agarose gel electrophoresis was used as an inexpensive method to estimate the presence of the correct LF82 insert, while Sanger sequencing was used to ensure insertion of the correct CDS at the correct site in the vector.

3.5.4.1 Colony PCR

Multiple colonies were selected from each lane of the rectangular agar plate to test for the presence of successful SLiCE vectors. Colonies were picked, subcultured onto agar and then utilised for a colony PCR reaction. Primers used were specific to binding sites either side of the MCS gene insert region on the pCMV-mEm-4GS vector, and can be seen in Table 2.4 (Oligonucleotides).

Amplification using this arrangement resulted in PCR products corresponding to their desired insert gene size plus 914 bp. Unsuccessful SLiCE reaction vector presence resulted in PCR products of just 914 bp, allowing quick and simple differentiation between a majority of successful and unsuccessful SLiCE recombinations. An example of the colony PCR electrophoresis gels obtained can be seen in Figure 3.7.

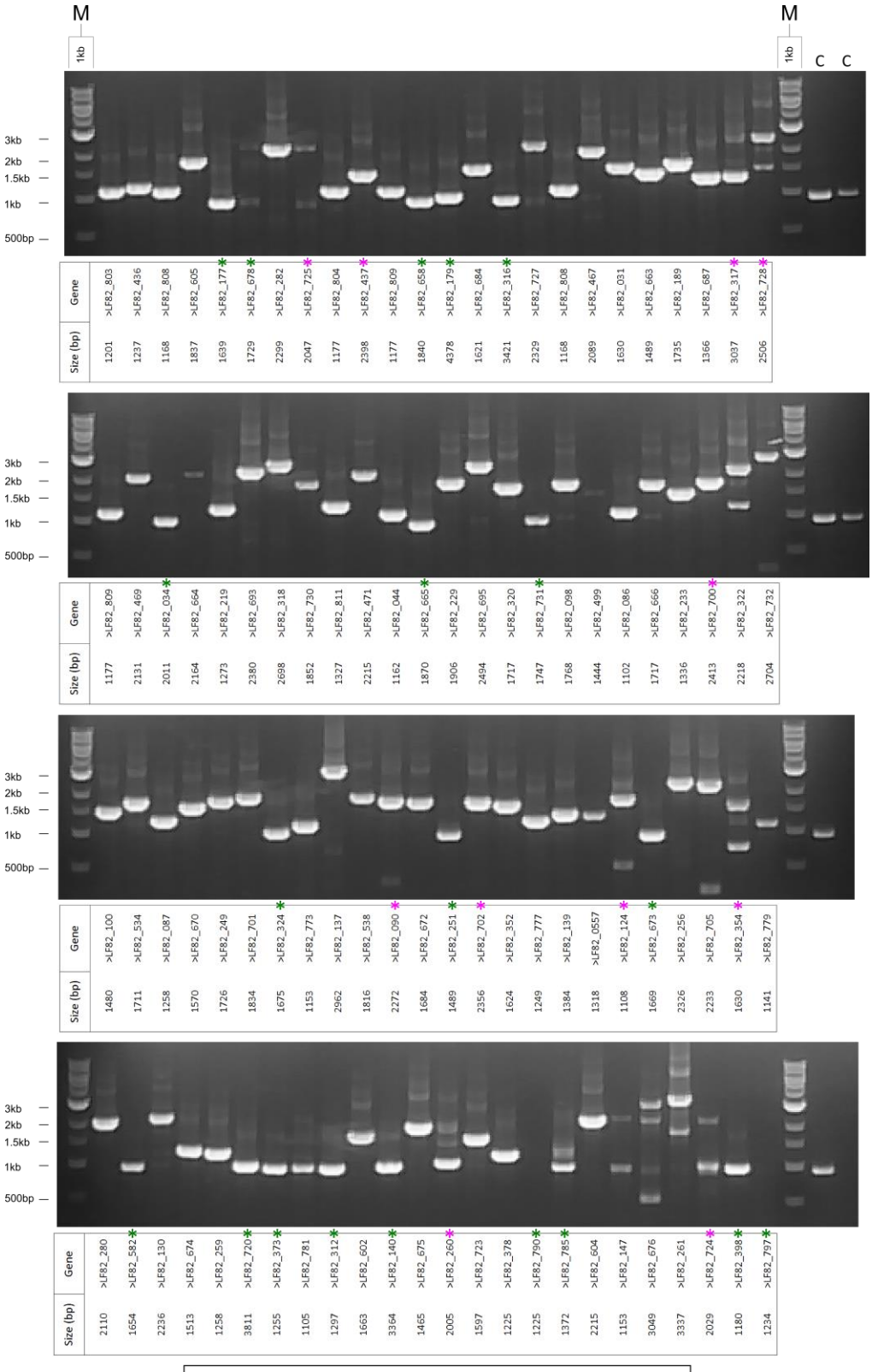

M
1kb
M
1kb
C
C
3kb
2kb
1.5kb
1kb
500bp
Gene
Size (bp)
>LF82_803 1201
>LF82_436 1237
>LF82_808 1168
>LF82_605 1837
>LF82_177 1639
>LF82_678 1729
>LF82_282 2299
>LF82_725 2047
>LF82_804 1177
>LF82_437 2398
>LF82_809 1177
>LF82_658 1840
>LF82_179 4378
>LF82_684 1621
>LF82_316 3421
>LF82_727 2329
>LF82_808 1168
>LF82_467 2089
>LF82_031 1630
>LF82_663 1489
>LF82_189 1735
>LF82_687 1366
>LF82_317 3037
>LF82_728 2506
>LF82_809 1177
>LF82_469 2131
>LF82_034 2011
>LF82_664 2164
>LF82_219 1273
>LF82_693 2380
>LF82_318 2698
>LF82_730 1852
>LF82_811 1327
>LF82_471 2215
>LF82_044 1162
>LF82_665 1870
>LF82_229 1906
>LF82_695 2494
>LF82_320 1717
>LF82_731 1747
>LF82_098 1768
>LF82_499 1444
>LF82_086 1102
>LF82_666 1717
>LF82_233 1336
>LF82_700 2413
>LF82_322 2218
>LF82_732 2704
>LF82_100 1480
>LF82_534 1711
>LF82_087 1258
>LF82_670 1570
>LF82_249 1726
>LF82_701 1834
>LF82_324 1675
>LF82_773 1153
>LF82_137 2962
>LF82_538 1816
>LF82_090 2272
>LF82_672 1684
>LF82_251 1489
>LF82_702 2356
>LF82_352 1624
>LF82_777 1249
>LF82_139 1384
>LF82_0557 1318
>LF82_124 1108
>LF82_673 1669
>LF82_256 2326
>LF82_705 2233
>LF82_354 1630
>LF82_779 1141
>LF82_280 2110
>LF82_582 1654
>LF82_130 2236
>LF82_674 1513
>LF82_259 1258
>LF82_720 3811
>LF82_373 1255
>LF82_781 1105
>LF82_312 1297
>LF82_602 1663
>LF82_140 3364
>LF82_675 1465
>LF82_260 2005
>LF82_723 1597
>LF82_378 1225
>LF82_790 1225
>LF82_785 1372
>LF82_604 2215
>LF82_147 1153
>LF82_676 3049
>LF82_261 3337
>LF82_724 2029
>LF82_398 1180
>LF82_797 1234
Key:
* Band size of ~ 914bp indiciative of a failed SLiCE
* Band size not as expected - wrong product SLiCEd
C Control lane - empty pCMV-mEm-4GS vector of size 914bp

Figure 3.7. **Colony PCR of transformed *E. coli***. Electrophoresis gel showing Set 2 colony PCR products after subjecting Set 2 PCR R2 gene products to the SLiCE reaction, transforming the subsequent DNA into XL1-blue, and selecting one colony per LF82 unique gene for colony PCR testing. Few fragments in neighbouring lanes appear of similar size, suggesting little-to-no cross contamination between transformation lanes from which these colonies were selected. Asterisks represent PCR amplifications predicted to have failed. C = control lane/s using empty pCMV-mEm-4GS for PCR.

The LF82 unique gene set contained CDSs of varying sizes, with some CDSs as small as 105 bp. Smaller members of the gene set were therefore difficult to analyse using this colony PCR, as the addition of 105 bp to the 914 bp seen with empty vector was difficult to detect. For these products, an alternative colony PCR was performed using an alternative primer (pCMV_s_small_F) which resulted in PCR products corresponding to the desired insert size plus 223 bp. Using this and the method above, colony PCRs for the first pass of LF82 unique gene amplification and SLiCE showed a success rate of 62% (60/96), 68% (65/96), and 74% (32/43) for Sets 1, 2 and 3, respectively. A total of 67% of the LF82 unique CDSs were therefore successfully cloned via SLiCE on the first attempt. Cultures corresponding to these were frozen.

As multiple colonies were selected and subcultured from each transformation lane, new clones were selected for each sample that failed the first colony PCR attempt. These colonies were subjected to colony PCR and samples which appeared successful added to the frozen sample library. For those that failed this colony PCR, the initial amplification PCR R1, PCR R2 and the SLiCE reaction was re-attempted once. Any LF82 unique genes that appeared to have failed this process were abandoned until a later date, where they were handled separately (see: Revisiting SLiCE failures for the creation of Set 4). In total, these colony PCRs indicated the successful SLiCE cloning of 92% of the LF82 unique gene library, although these were yet to be confirmed via sequencing.

Interestingly, purification of PCR R2 products pre-SLiCE did not increase SLiCE efficiency – in fact it appeared to reduce it. It is not clear why this was the case, although it may simply be due to a loss of PCR R2 product and a lower concentration of insert entering the SLiCE reaction.

3.5.4.2 Sanger sequencing

Frozen glycerol cultures of XL1-blue containing the library of vectors were purified as described below (see: Purification of library DNA). DNA from each of the library vectors was sent for sequencing by Source Bioscience (Nottingham, UK). The returned sequencing data was analysed using a local BLAST against the LF82 genome. This resulted in the confirmation of 170 LF82 unique genes out of the original 235 genes – a success rate of 72%. This rate was not as high as desired, however, and those LF82 unique genes that previously failed the SLiCE reaction were revisited (see below).

3.6 Revisiting SLiCE failures for the creation of Set 4

LF82 unique genes which failed the three attempts of PCR R1, PCR R2 and SLiCE, were revisited to reattempt PCR amplification and SLiCE. Alternative PCR primers were designed independently for each of these CDSs, and they were all once again subjected to PCR R1, PCR R2 and SLiCE. However, the 3D bead-guide was not utilised for this experiment, with standard 10 cm agar plates used instead. 50 µL of each sample was spread onto each plate, resulting in an increased the likelihood of at least once successful SLiCE clone being present on each plate.

65 LF82 unique genes were re-entered into this final round of library preparation. Of these, 39 were successful and were confirmed by sequencing. This resulted in a final library completion rate of 209 unique genes out of the original 235 – a total library coverage rate of 89%.

3.7 Creation of screen reporter cell line

3.7.1 HeLa mCherry-LC3 creation

3.7.1.1 LC3 as a reporter of autophagosome number

LC3 is commonly used as a reporter to estimate cellular autophagosome number – typically via its fusion to a fluorescent protein such as GFP or mCherry (Pugsley 2016). As mEm-GFP was already in use as part of this screen, mCherry was selected for LC3 fusion.

3.7.1.2 Cell line creation via lentiviral transduction

To create the reporter cell line, the modification of a previously created lab vector containing mCherry-GFP-LC3 (pBABE-mCherry-EGFP-LC3) was attempted. A PCR was designed to amplify around the entirety of the vector, with the exclusion of the EGFP-encoding fragment. The resulting PCR product was ligated to form a vector containing mCherry-LC3, from which the mCherry-LC3 fragment was amplified, gel extracted and ligated into a lentiviral vector (pWPXLd). However, the final step of this process proved difficult, with it failing multiple times for unknown reasons. Alternative destination vectors were also tried, but to no avail.

An alternative strategy using gBlocks (IDT) was used to create a ligateable mCherry-LC3 fragment with restriction sites for EcoRV and NotI at 5' and 3' ends, respectively. This fragment was subjected to oligonucleotide annealing, and subsequently restriction digested using EcoRI-HF and NotI-HF, before being ligating into a linearised pWPXLd vector to form pWPXLd-mCherry-LC3. Lentiviral transduction was then used to integrate the mCherry-LC3 fragment into the HeLa CCL2 genome. Successfully transduced cells were identified via mCherry expression and fluorescence microscopy. Transduced cells were then diluted to allow the isolation of individual cells displaying strong mCherry fluorescence phenotype. From this, a homogenous population of transduced cells was obtained, resulting in a HeLa CCL2 cell line stably expressing the mCherry-LC3 reporter: HeLa-mCherry-LC3.

3.8 Screen of LF82 unique gene library

3.8.1 Transfection of plasmid library for screen

3.8.1.1 Purification of library DNA

XL1-blue containing successfully SLiCE library DNA were used to inoculate 1 mL of broth and grown overnight. These broths were subjected to a 96-well miniprep using the GenElute HP 96-well plasmid miniprep kit (Sigma), which utilises silica-based columns for purification. Library DNA was then used in for transfection of HeLa cells process outlined below.

Initial experiments revealed that transfection efficiency was not as desired for many samples. This was attributed mostly to low purity of the library DNA, but also to low DNA concentration in several cases. To combat this, an alternative miniprep kit was used which assured higher purity final DNA via the use of anion exchange columns and endotoxin removal: the NucleoBond Xtra Midi EF Midiprep Kit (Macherey-Nagel). Furthermore, the overnight growth of the XL1-blue library was repeated three times for each sample, and all cultures pooled for use with the new miniprep kit. This not only ensured a higher purity of final plasmid product, but also resulted in higher concentrations of library DNA.

3.8.1.2 Transfection reagents

Lipofectamine 2000 is routinely used in conjunction with HeLa cells for DNA transfection due to its efficiency across a wide cell range, and its simple protocol. However, a certain level of toxicity is often observed with Lipofectamine 2000 (Thermo Fisher Scientific) treatment. This is particularly undesirable when working with low cell numbers – such as those present in a single well of a 96 well plate – but also when utilising an automated process for image capture, as adjustments to disregard regions of cell death are difficult. Furthermore, it may be difficult to differentiate cell death due to Lipofectamine 2000 treatment from cell death due to the expression of a

toxic protein. For these reasons, optimisation of the Lipofectamine 2000 transfection protocol was performed

Varying volumes of both Lipofectamine 2000 reagent and a range of DNA concentrations (10-500 ng) were used to test reagent toxicity, and to determine the best possible transfection conditions. In addition to this, Lipofectamine 3000 (Thermo Fisher Scientific) and JetPRIME (Polyplus) reagents were trialled for toxicity comparison. While neither of these reagents demonstrate as simple a protocol as Lipofectamine 2000, they do claim to offer a lower level of toxicity. Comparisons were performed using reverse transfections, as this transfection method is less labour intensive in a high-content setting and reduces the total time taken to perform the experiment.

Lipofectamine 2000 was shown to produce good transfection efficiency vs toxicity balance when utilising $1x10^5$ HeLa mCherry-LC3 mL^{-1}, 0.25 µL Lipofectamine 2000 and 50 ng of DNA per well. In comparison to this, however, Lipofectamine 3000 resulted in better transfection efficiency – with the use of $2.5x10^5$ cells mL^{-1}, 120 ng DNA and 0.8 µL of P3000 reagent – and a reduction in the levels of cell death, despite the use of the same total volume of transfection reagent.

In comparison to both Lipofectamine transfection reagents, JetPRIME was inferior in terms of transfection rate and levels of toxicity. Lipofectamine 3000 was therefore used for transfecting the library DNA into the HeLa mCherry-LC3 reporter cell line.

3.8.1.3 Transfection

To perform the screen, 96-well transfection plates required a modified layout to ensure that the ImageXpress Ultra microscope could obtain consistent images during screening. The outer wells of 96-well plates were therefore left

blank, and only the 60 remaining internal wells used. Transfections were performed and resulted in a transfection efficiency of ~40-60% for LF82 genes, and ~60% for empty pCMV-mEm-4GS vector. An average of 183 cells were present per site, with ~90 successfully transfected cells for LF82 genes and ~112 successfully transfected cells for control wells sites. These values were in line with those seen with during Lipofectamine 3000 testing.

Each LF82 unique gene set – Set 1, 2 and 3 – was screened in duplicate, and split into two subsets: Part 1 and Part 2. An example of the layout of the genes from Set 3 Part 1 can be seen in Figure 3.8 a, below, while the layout of the rest of the sets can be found in Appendix III: Layout of screen transfection plates. Figure 3.8 b shows the layout of Set 4 (see: Revisiting SLiCE failures for the creation of Set 4 for more information).

3.8.1.4 Transfection plate controls

The bottom row of each transfection plate (row G) was reserved for controls. Cells in control wells were, from left to right on each plate: transfected with empty pCMV-mEm-4GS vector (mEm), treated with ammonium chloride (NH_4Cl), treated with rapamycin (Rap), treated with ammonium chloride and rapamycin (NH_4Cl + Rap), transfected with a pCMV-GFP-p62 (p62), transfected with pCMV-3xHA-SseD (a vector containing unlabelled *Salmonella typhimurium* secreted effector protein SseD) (SseD), and finally, untransfected and untreated cells (Untr.). The layout of these controls can be seen in Figure 3.8.

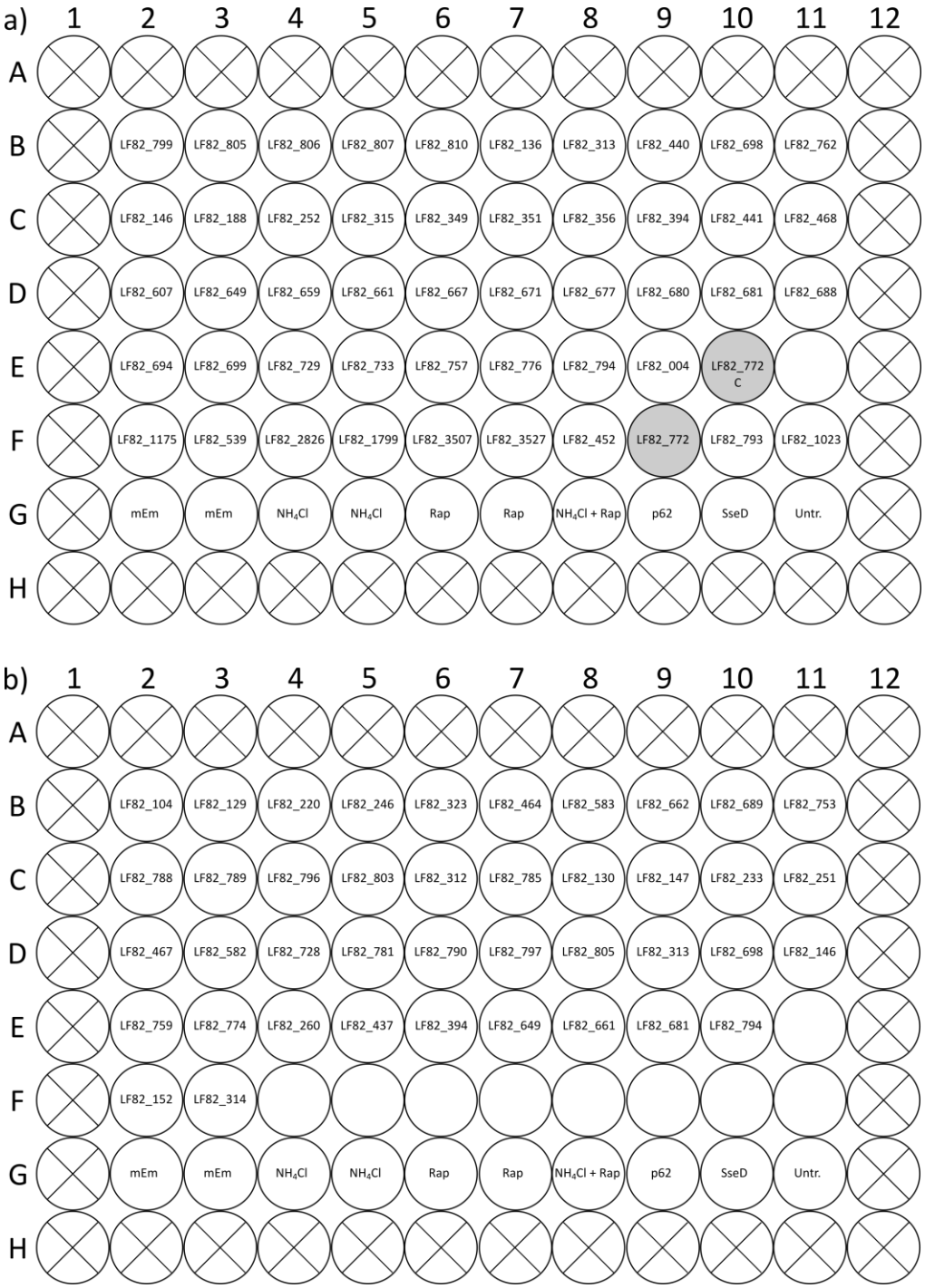


Figure 3.8. **Example layouts of LF82 unique gene transfections for screening**. Layout of a) Set 3 Part 1, and b) Set 4 for screen transfections. Empty wells are blank, and wells not used due to limitations of automated microscopy are marked with an X. Control wells can be seen at the bottom of each respective figure. Grey wells are those thought to be involved in cross contamination (see: Radial zone 4), 'C' designates the contaminated well (Set 3 Part 1, well E10). For the layouts of the entirety of Sets 1, 2, 3 and 4, see Appendix III: Layout of screen transfection plates. For an explanation of control well contents, see: Transfection plate controls.

Empty vector was used to ensure that pCMV-mEm-4GS alone (or transfection reagent) did not induce any strong phenotypes. Ammonium chloride prevents lysosomal enzyme function resulting in an increase in the number of autophagosomes, while rapamycin induces autophagosome production via the stimulation of nutrient and oxidative stress responses. Ammonium chloride and rapamycin were used together in an effort to stimulate autophagosome production, but prevent their breakdown, theoretically resulting in an increase in the number of autophagosomes. p62 was used to induce autophagy via its overexpression (see: Artificial manipulation of autophagy) and is also known to co-localise with LC3, thereby acting as a positive control for the detection of autophagosomes – this can be seen on graphs displaying GFP/LC3 co-localisation (see: GFP/LC3 and GFP/Actin). SseD has been previously identified as an inducer of autophagy when over expressed (A. Huett, Personal communication). Finally, untransfected cells were used to gauge the effects of transfection reagents on HeLa mCherry-LC3 cells, and also to ensure that cells did not undergo any unusual phenotypic changes due to other external stimuli.

3.8.1.5 Preparation of cells for microscopic analysis

In order to analyse the effects of LF82 unique gene expression upon HeLa cells, several different reporting strategies were used in parallel. Fluorescent markers have already been mentioned, with GFP allowing the visualisation of expressed LF82 protein localisation, and LC3 allowing autophagosomes to be enumerated and their location determined. However, also used were the fluorescently labelled bicyclic peptide, Phalloidin 647, which allows visualisation of F-actin, and the fluorescent dye, Hoechst 33342, which allows visualisation of DNA.

Hoechst 33342 was used as a substitute for DAPI (4',6-diamidino-2-phenylindole) – which also stains DNA – and the term ‘Hoechst’ was used throughout this research in reference to Hoechst 33342. This was to simplify

CellProfiler pipeline naming strategies, and to simplify code when analysing screen data (see: Software for image analysis).

Cells were fixed using 4% formalin in PBS, before permeabilising and applying Phalloidin 647 and Hoechst 33342. Once stained, plates were stored at 4°C for < 7 days before imaging cells.

3.8.2 Screen image collection

3.8.2.1 Automated image capture

The IX Ultra captured images at 4 different wavelengths, at 6 different sites for each well. For each 60 transfection wells, this therefore resulted in 1440 images. Sets 1, 2, and 3 were split between two 96-well plates to enable the use of 60 wells per plate plus controls, and each set was performed in duplicate. Set 4 was not split across two 96-well plates, but was performed in duplicate. This therefore resulted in a final image dataset of ~20,000 images. While a screen of this magnitude is possible via manual microscopy, it would take an excessive amount of time to perform in comparison to the ~18 h that was necessary using the IX Ultra.

3.9 Future perspectives

This study was primarily concerned with the AIEC type strain LF82, however, it would be of interest to investigate the presence of unique genes in other AIEC strains, such as NRG857c (O83:H1) through a similar approach as used here. Comparisons between the LF82 and NRG857c gene sets, via genetic analysis and through a phenotypic screen, may aid identification of any CDSs likely to be functional homologues.

InParanoid was used in this research to compare relatively few strains of *E. coli*. Despite the small size of the comparison set, this process was time consuming and some steps led to confusion due to the way in which InParanoid presents comparison data. It would therefore be advantageous to

develop a more automated or streamlined process via which InParanoid is used to compare strains and remove suspected orthologs. This process was not optimised in the current research due to the relatively limited use of InParanoid, but may be of benefit for any future use involving multiple comparisons or larger comparison strain lists.

Tracking the number of autophagosomes using mCherry-LC3 is just one of a range of phenotypes that may help elucidate the methods through which AIEC manipulates host cell biology. The use of alternative reporters may therefore be used to understand how AIEC is able to invade and survive within mammalian cells. For example, the association of a particular LF82 protein with microtubules or actin may suggest that it is involved in LF82 invasion, as first suggested by Boudeau et al. (1999). However, the use of any other localisation markers than those used in this research would have been difficult, due to the number of fluorescent markers already in use. Despite this, mCherry could be reassigned to other reporters in the future.

Bray et al. (2016) have recently described the use of cell image analysis software, CellProfiler, for the measurement of cells (size, shape, texture etc.) in a setting similar to that described here. An example of this can be seen in Figure 3.9, in which a morphological profiling strategy is outlined. While this example does not consider the use of CellProfiler for the identification of bacterial effectors, it does display the usefulness of CellProfiler in a high-content screen.

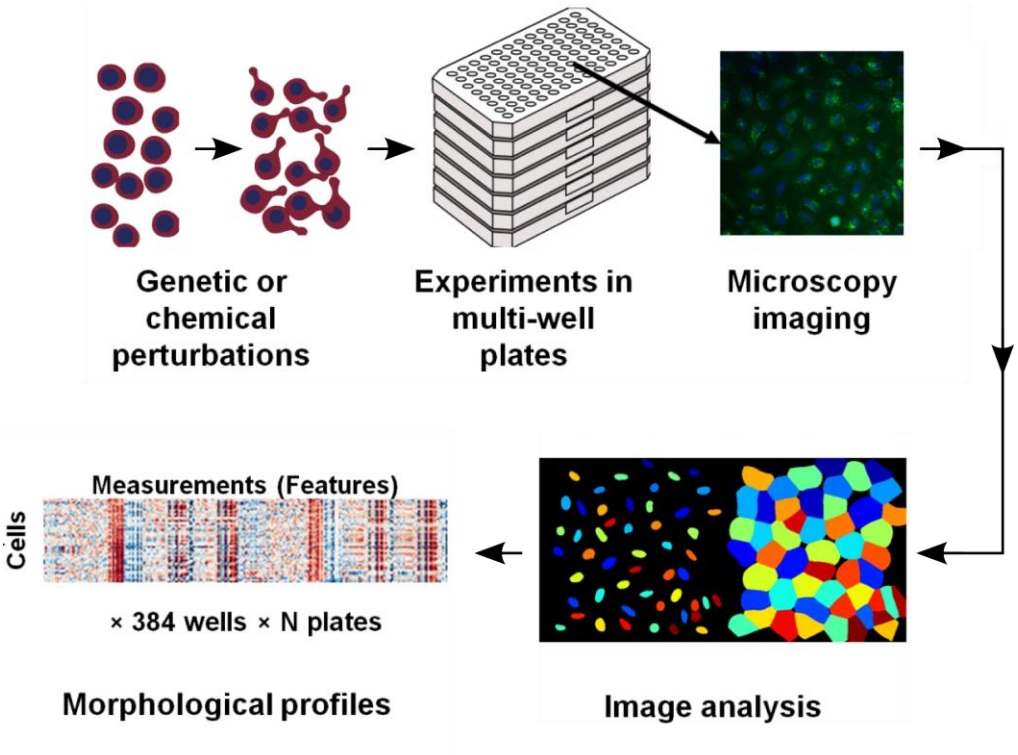


Figure 3.9. **Overview of the morphological screen developed by Bray et al. (2016)**. This strategy shows the use of a high-content image-based screen for the discovery of new drugs, as well as the analysis of drugs currently in use. Figure adapted from Bray et al. (2016).

Since the LF82 analysis presented here, Zhang et al. (2015) have also predicted LF82 'unique' genes, via comparison of AIEC vs non-invasive *E. coli* (NIEC). Comparative RNAseq was used to identify genes expressed during exponential and stationary phases of bacterial growth. From this, LF82 was found to express 166 genes which were not present in NIEC. When compared to the InParanoid/Miquel et al. (2010) gene set used in the present research, 73% of the Zhang et al. (2015) gene set were present (see: Appendix II: LF82 unique genes for comparisons of these sets). Despite the elegant work of Zhang et al. (2015) and Miquel et al. (2010) in identifying CDSs unique to LF82, the expression library created during the present research is the only one currently known to exist.

3.10 Summary

InParanoid can be used to compare the proteome of AIEC LF82 with those of multiple bacterial strains to obtain a list of non-orthologous CDSs. These identified CDSs can be considered unique to LF82 within this comparison set. The use of SLiCE, along with an in-house designed bead-spreader proved successful in allowing the cloning of 89% of LF82 unique genes for a fluorescence microscopy based screen for LF82 bacterial effectors.

Chapter 4: Development of an automated pipeline for the analysis of GFP-tagged proteins in a HeLa cell population

4.1 High-content techniques for cell imaging

4.1.1 Background

The use of high-content screening has proven itself invaluable for the discovery of novel drugs, and for the functional classification of genes and proteins over the past 10-15 years (Abraham et al. 2004). As software has been optimised and techniques refined, the potential scale and use of high-content screens has increased.

High-content, multi-phenotypic screens have shown the ability to identify factors affecting bacterial virulence as well as invasion. An example of this is the identification of the invasin, OmpA, from *Coxiella burnetii* which was observed through the screening of a library of *C. burnetii* transposon mutants (Martinez et al. 2014). Invading/non-invading bacteria were identified using their EGFP expression, and this information used to determine transposons containing CDSs likely important for bacterial invasion.

High-content screens have been used in the past to identify actin regulators, host proteins necessary for the invasion of HIV, and tyrosine kinase signalling regulators, amongst others (Rohn et al. 2011; Liu et al. 2009; Brass et al. 2008; Friedman & Perrimon 2006). Many of these screening approaches utilise an RNA interference (RNAi) based approach, however, as these rely on the inhibition of gene expression, much information is lost (such as localisation or protein-protein interaction data).

The screen presented in the following chapter can be considered an extension of that described by Slagowski et al. (2008). In said screen, Slagowski et al. (2008) analyse the effects of the expression of a multitude of bacterial virulence factors in *Saccharomyces cerevisiae* (yeast). In this example, the effects of expression upon growth of the yeast was examined. However, while this method involves expression of potential virulence genes

within a eukaryotic cell – similar to the screen described here – it only considers one primary phenotype: growth.

Here, a screen was developed to analyse the expression of >200 potential effectors from LF82. Host cell phenotypes displayed as a result of this were then analysed using a method able to measure a multitude of phenotypes.

4.2 Aims

The development of an automated screen for the measurement of a range of cell phenotypes will allow a quick, and unbiased evaluation, of phenotypes observed during the eukaryotic expression the LF82 unique gene set. This will allow the identification of putative LF82 effectors, and help to provide an insight into the role of LF82 in CD. The aims of this chapter were to:

- Create a CellProfiler pipeline which can be used to process all images from the previous chapter in an automated fashion.
- Analyse CellProfiler data using CellProfiler Analyst and R, making a semi-automated analysis process for the identification and validation of potential screen hits.
- Verify the use of graphs produced using R for their ability to present useful and meaningful data that can then be interpreted by an end user for manual phenotype validation.
- Select one or several CDSs of interest from LF82 whose phenotype/s provoke further investigation.

4.3 Screen pipeline development

4.3.1 Software for image analysis

4.3.1.1 CellProfiler

CellProfiler is a free open-source image analysis software package (available at: http://cellprofiler.org/) designed for cell microscopy image analysis (Carpenter et al. 2006). It can be used to analyse batches of hundreds of thousands of images and can output data in a number of program specific or open formats (e.g. CellProfiler Analyst files, Microsoft Excel files, SQL databases, .txt files etc.)

CellProfiler functions using a custom pipeline to perform measurements or functions on an image before passing the resulting measurement/image onto another subsequent module in the pipeline. Modules are pre-packaged with CellProfiler and are geared towards biological analysis of microscopy images, however they can often be tailored to an individual user's specifications. The resulting quantitative analysis data can be exported for further analysis in several different external applications.

CellProfiler has been cited in over 3500 publications to date, and has been utilised in high-impact research into Ebola, *Mycobacterium tuberculosis* and leukaemia amongst other research areas (Sakurai et al. 2015; Stanley et al. 2014; Hartwell et al. 2013; Wen et al. 2012).

4.3.1.2 CellProfiler alternatives

Other software packages exist which can be used to achieve similar results to CellProfiler, however they are often not free, or are not of an open-source nature. Open-source software allows any user to download and reproduce an experiment as long as they have compatible hardware, and also allows for greater customisability and flexibility in comparison to closed-sourced alternatives.

4.3.1.3 CellProfiler pipeline testing

The designed pipeline accepted four image channels: Hoechst (for Hoechst 33342), GFP (for mEmerald GFP), mCherry (for mCherry-LC3), and far-red (for actin detection via phalloidin 647). Images from each corresponding wavelength were split into channels based upon their file metadata, resulting in four 'common names' for the channels: Hoechst, GFP, LC3 and Actin. The pipeline was designed and tested using several randomly selected image sets, control image sets, and manually identified 'extremes' (e.g. wells whose images presented just noise, or wells in which no cells were successfully transfected).

4.3.1.4 Alteration of untransfected controls

Untreated control wells G10 and G11 (containing SseD transfected cells, and untransfected cells, respectively) contained no GFP-encoding vector for GFP expression, and therefore did not provide any GFP signal. The CellProfiler pipeline relies upon this signal to identify cells from which to take measurements (see below). In order to take any measurements from these control wells, it was therefore necessary to duplicate Hoechst channel images, and rename them as GFP channel images.

Images from wells G10 and G11 were therefore considered by CellProfiler as wells where every cell was transfected, due to the Hoechst channel being considered as the GFP channel. These wells can be seen on the graphs displaying GFP/Hoechst correlation and concerning radial zones 1 and 4, as examples of 'perfect' GFP localisation to the nucleus (Figures 4.10, 4.11, and 4.27, respectively).

4.4 CellProfiler pipeline

A visual representation of the CellProfiler pipeline described below can be seen in Figure 4.1.

Nuclei were identified as 'objects' from the Hoechst channel based upon their intensity, size and typical shape. Each identified nuclei 'object' was then used as a 'seed' within the Actin channel; each seed was expanded until edges of cells were detected. This method showed superior results over direct attempts to identify cells without nuclei seeds, as nuclei are far easier for CellProfiler to identify. This was due to Hoechst 33342's specificity to such a small, localised regions (such as nuclei) and said region's isolated nature from other nuclei, as few nuclei overlapped one another. Furthermore, as each seed required one object (a cell in this case) to be associated with it, it was less likely that two cells were clumped together unless multiple nuclei were erroneously identified as one nucleus object.

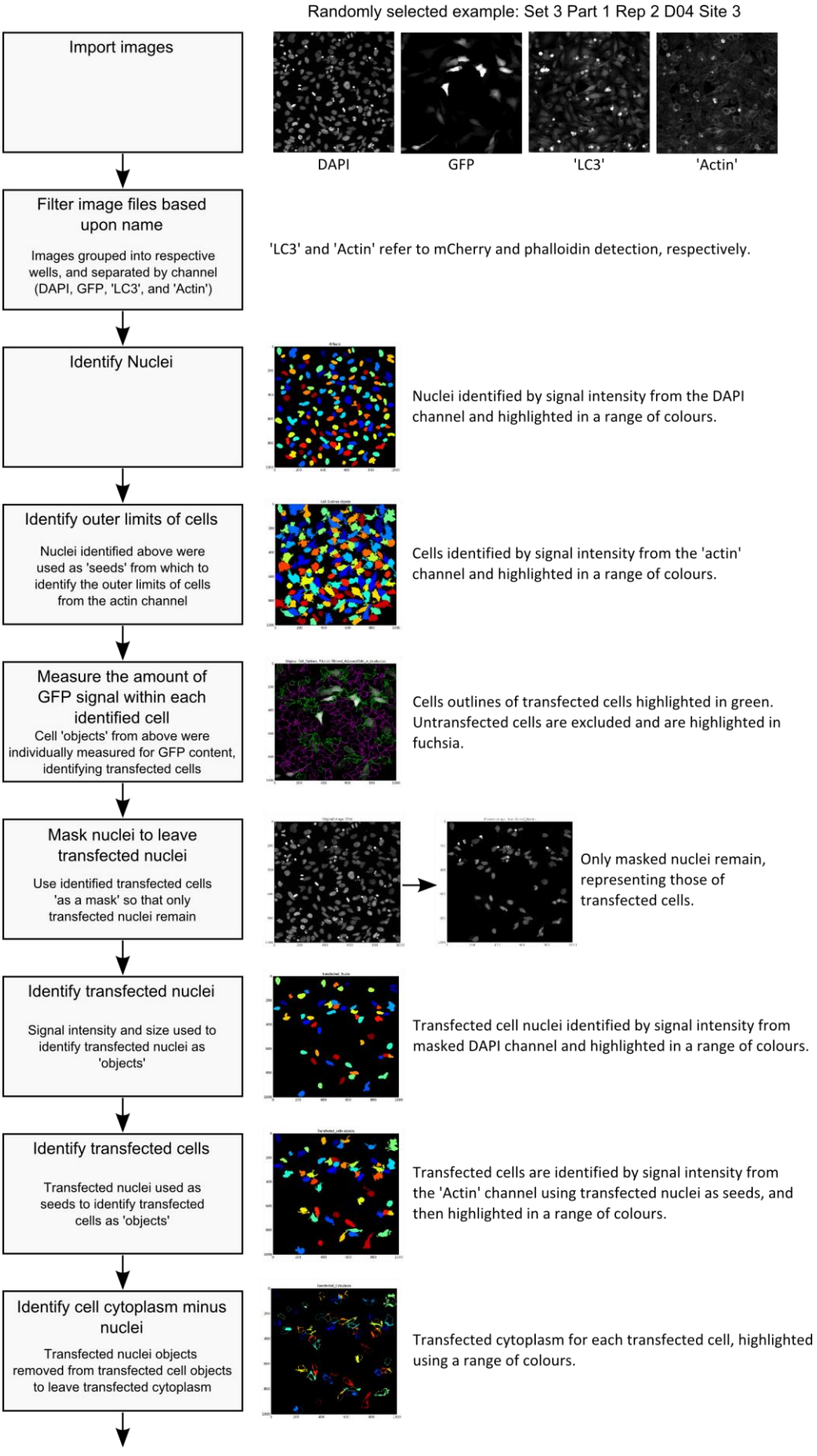
Once cells were identified, their internal intensity was measured through the GFP channel. Successfully transfected cells contained GFP and would therefore obtain a higher intensity value due to their brightness. This allowed identification and differentiation of transfected cells from non-transfected cells. Despite this, misidentification of transfected/non-transfected cells was sometimes evident, however, where this did occur, it typically appeared to be due to abnormally low GFP expression, resulting in some transfected cells not being detected.

Original images were 'masked' with the newly identified transfected cell objects, in effect removing non-transfected cells from each image. From these 'transfected only' cells, nuclei objects were subtracted to leave only the cytoplasm of transfected cells. Measurements were then taken of each individual cell and nucleus, with many different parameters measured, including diameter, perimeter and circumference. The correlation between

GFP, Actin and Hoechst were also measured to evaluate localisation of the expressed gene to those areas.

A measurement of the radial distribution of GFP was then performed (see Figure 4.2). This worked by separating each cell into 4 concentric regions (starting from the centre of each cell) after which the mean fractional intensity was calculated for each resulting zone. This was calculated by determining the fraction of total GFP intensity in each zone, which was normalised by the number of pixels within each zone. In essence this module provided a good representation of where the GFP (and therefore the protein of interest) was localised within an individual cell – be it nuclei, cytoplasm or membrane. While this module was useful here, it could prove to be even more useful at higher magnifications and resolutions by, for example, allowing localisation to the endoplasmic reticulum or the nuclear envelope to be more readily identified.

The mCherry channel was subjected to a tophat transformation, which resulted in the enhancement of small details of the mCherry channel images. The tophat transformation works by reducing and normalising background noise via the enhancement of regions of high intensity (high contrast), from regions of low intensity (low contrast). In this case, the basal expression of mCherry-LC3 was considered a low intensity region, and mCherry-LC3 foci (indicative of autophagosomes) were considered high intensity regions. Therefore the tophat enhancement resulted in an increase in brightness of each mCherry-LC3 foci compared to its background. Autophagosomes were then identified based upon pixel intensity and size before being enumerated and measured. Correlation between GFP and mCherry was determined to identify any localisation to autophagosomes, before exporting the final data as an Analyst database and as a SQLite database for analysis in CellProfiler Analyst and R, respectively.


Randomly selected example: Set 3 Part 1 Rep 2 D04 Site 3
Import images
DAPI
GFP
'LC3'
'Actin'
Filter image files based upon name
Images grouped into respective wells, and separated by channel (DAPI, GFP, 'LC3', and 'Actin')
'LC3' and 'Actin' refer to mCherry and phalloidin detection, respectively.
Identify Nuclei
Nuclei identified by signal intensity from the DAPI channel and highlighted in a range of colours.
Identify outer limits of cells
Nuclei identified above were used as 'seeds' from which to identify the outer limits of cells from the actin channel
Cells identified by signal intensity from the 'actin' channel and highlighted in a range of colours.
Measure the amount of GFP signal within each identified cell
Cell 'objects' from above were individually measured for GFP content, identifying transfected cells
Cells outlines of transfected cells highlighted in green. Untransfected cells are excluded and are highlighted in fuchsia.
Mask nuclei to leave transfected nuclei
Use identified transfected cells 'as a mask' so that only transfected nuclei remain
Only masked nuclei remain, representing those of transfected cells.
Identify transfected nuclei
Signal intensity and size used to identify transfected nuclei as 'objects'
Transfected cell nuclei identified by signal intensity from masked DAPI channel and highlighted in a range of colours.
Identify transfected cells
Transfected nuclei used as seeds to identify transfected cells as 'objects'
Transfected cells are identified by signal intensity from the 'Actin' channel using transfected nuclei as seeds, and then highlighted in a range of colours.
Identify cell cytoplasm minus nuclei
Transfected nuclei objects removed from transfected cell objects to leave transfected cytoplasm
Transfected cytoplasm for each transfected cell, highlighted using a range of colours.

Measure objects

Transfected nuclei and cells measured

Object	Feature	Mean	Median	STD
Transfected_Nuclei	Eccentricity	0.76	0.80	0.13
Transfected_Nuclei	MajorAxisLength	59.11	57.88	12.12
Transfected_Nuclei	MinorAxisLength	35.89	33.51	7.59
Transfected_Nuclei	Orientation	4.48	6.05	44.46
Transfected_Nuclei	Compactness	1.22	1.19	0.16
Transfected_Nuclei	Area	1595.36	1480.00	505.14
Transfected_Nuclei	Center_X	536.91	539.00	261.59

Example of some measurements obtained from the object measurement module.

Measure correlations

Correlation between transfected cell GFP-actin, and GFP-DAPI measured

Measure radial distribution

Measurement of the mean fractional intensity of GFP in each of 4 bins[1]

[1] See figure legend

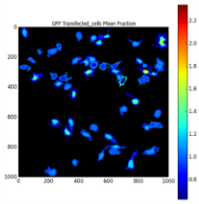
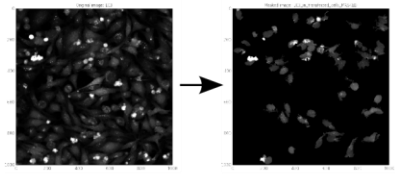


Image	Objects	Bin	Fraction	Intensity	COV
GFP	Transfected_cells	1	0.0331	1.1558	0.0406
GFP	Transfected_cells	2	0.1318	1.1332	0.0722
GFP	Transfected_cells	3	0.2828	1.0867	0.115
GFP	Transfected_cells	4	0.5523	0.9414	0.233

Intensity of GFP in each given fraction is shown in a heat map: a colour closer to red is displayed when GFP is localised towards the nuclear region. The table shows average fractional intensity within each bin for all transfected cells.

Mask LC3 in transfected cells

Transfected cell objects used to mask LC3 channel

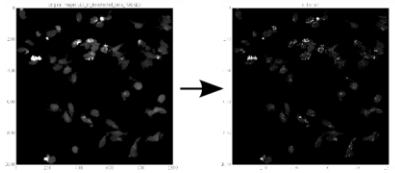


LC3 channel masked so that only transfected cell LC3 remains.

Morph transfected LC3 with Tophat function

Transfected LC3 masked image treated with Tophat filter[2]

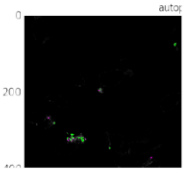
[2] See figure legend



Transfected cell LC3 subjected to a Tophat transformation, highlighting puncta in cells.

Identify autophagosomes from LC3 channel

Transfected cell LC3 image used to identify suspected autophagosomes using intensity and size



Autophagosomes identified by signal intensity from LC3 channel and highlighted in a range of colours. Due to their size, a zoom of the CellProfiler output is shown.

Measure autophagosomes

Based on the morphed LC3 image, the area and perimeter of detectable autophagosomes was calculated

Maths module calculation

Number of autophagosomes are divided by the number of transfected cells identified

Measure correlation

Correlation between transfected cell GFP-LC3 measured

Export to database

All data exported to a SQLite database for analysis in CellProfiler Analyst

Figure 4.1. **Overview of the CellProfiler pipeline used to process microscopy images**. This example uses a randomly selected well from Set 3. The flow diagram describes the function of each module in the pipeline, and the images show the resulting image data which the pipeline passes between modules. On the right are descriptions of what is shown in each resulting image. [1] The radial distribution module divides each cell object into 4 radial zones, increasing in size from the centre of each cell to the outer membrane. The levels of GFP signal in each zone is determined and expressed as a fraction of the total GFP within the entire cell. [2] A tophat transformation reduces and normalises background noise, allowing the enhancement of mCherry-LC3 foci. It therefore aids the detection of autophagosomes against basal mCherry-LC3 expression.

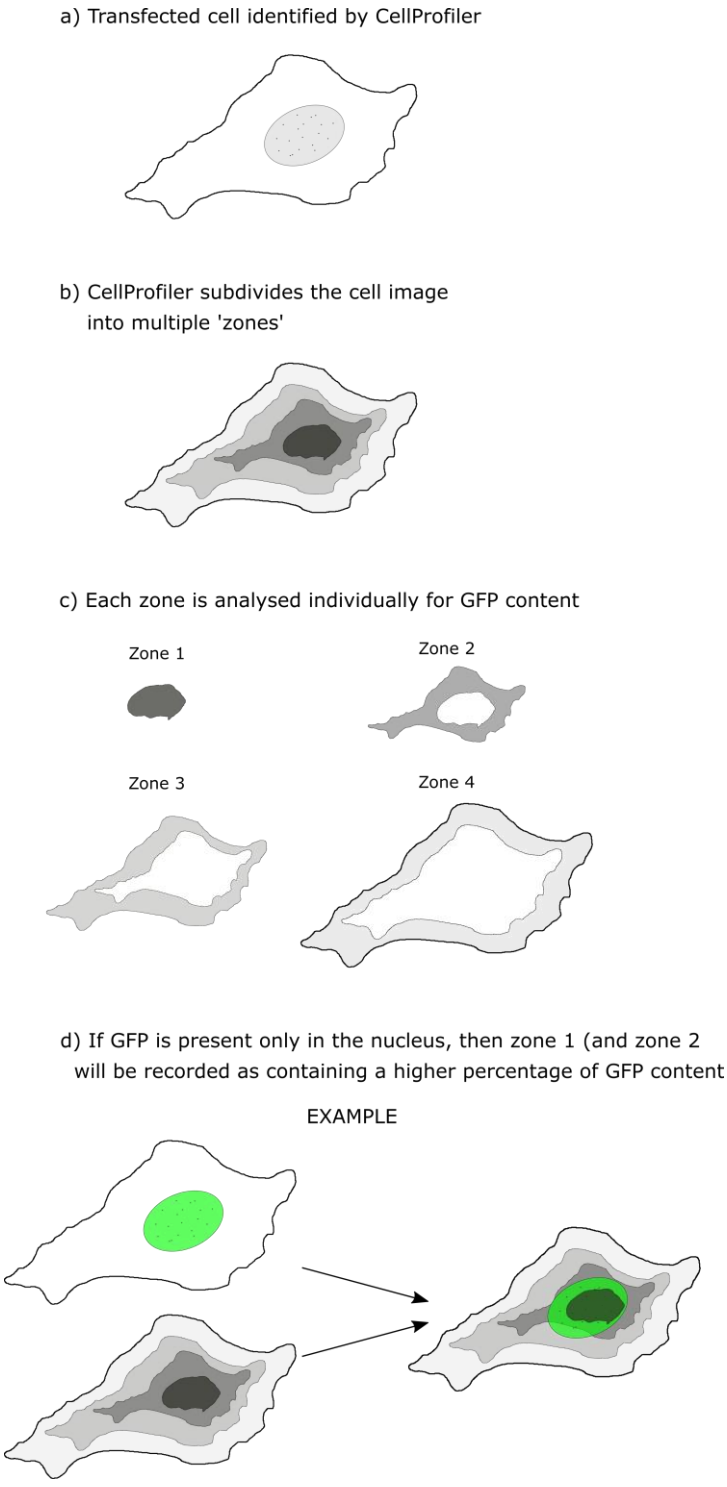


Figure 4.2. **Sketch of radial distribution measurement performed by CellProfiler.** a) CellProfiler evaluates each individual transfected cell using a radial distribution measurement. b) and c) During this measurement, cells are subdivided into 4 'zones', within which the amount of GFP is calculated as a percentage of total cellular GFP. d) Example where the nucleus of a transfected cell contains a majority of the cellular GFP. In this case, radial zones 1 and 2 would be reported as containing a higher proportion of the cellular GFP.

4.5 Analysis of screen data

A vast amount of data was obtained from CellProfiler analysis of the ~20,000 images in this screen, with each image potentially containing hundreds of objects. For this reason, manual analysis of the resulting data would have been extremely difficult. Two software packages were therefore employed to aid data analysis: CellProfiler Analyst – designed for analysing large, image-derived data; and R – designed for statistical and graphical analysis of small and large data. Despite the use of these software packages, subsamples of data were also selected and manually analysed for quality control.

4.5.1 CellProfiler Analyst (Analyst) data analysis

As the software name suggests, CellProfiler Analyst is a software package designed to work in harmony with CellProfiler itself. Like CellProfiler, Analyst (available from http://cellprofiler.org/) is a free, open-source software package, and is designed for the exploration and analysis of large datasets (Jones et al. 2008).

4.5.1.1 CellProfiler Analyst use with CellProfiler

Data from a CellProfiler pipeline can be exported as an SQLite database, which is not inherently compatible with Analyst. However, the creation of a CellProfiler Analyst Properties file during the execution of the final pipeline export module, allowed a relationship between Analyst and the data within the SQLite database to be maintained.

Through the Analyst-SQLite database relationship, Analyst was able to link database data with its original source images and screening plates. Each 60-wells of every 96-well plate used in this screen was therefore represented as a 96-well plate layout on-screen, allowing the selection of any desired well and the display of its associated screen images (see Figure 4.3, later).

4.5.1.2 Data exploration

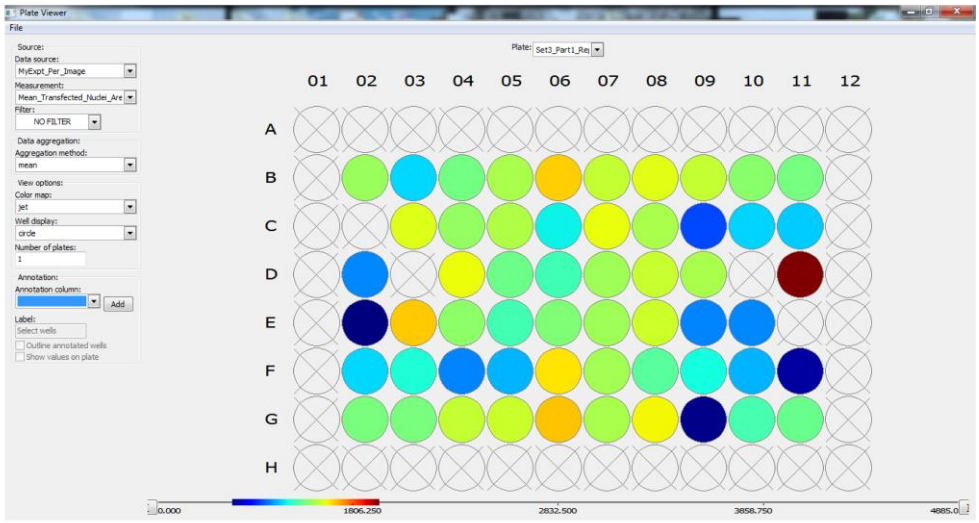
Every measurement taken during image analysis by CellProfiler (> 300 measurements) could be represented as a heat map of the wells of each respective plate. For example, the average value for each measurement of nuclei area could be displayed for each well, allowing any significant effects upon this measurement to be detected via a change in colour. An example of a heat map for average nuclei area can be seen in Figure 4.3 a), along with an example of average radial zone 4 measurements in Figure 4.3 b.

The use of Analyst in this way therefore allowed basic data analysis and casual data browsing, allowing the user to view images associated with a particular well at will. Furthermore, the relationship maintained between screen images and well data also allowed the manual confirmation of phenotypes detected using R (see: R data analysis) via analysis of well images by eye. Simply right clicking on a well allowed the selection of any desired images from any site for viewing. Furthermore, the included image browser automatically displayed the captured images as merges, with simple controls over which channel was displayed and basic brightness/contrast controls.

4.5.1.3 Data analysis

The plate view provided with Analyst allowed the analysis and detection of any plate-wide characteristics that may not have otherwise been detected. For example, the observation of any edge effects (lower transformation efficiency towards the edges of the plates, for example) would likely have been identified using this view. Furthermore, cross-contamination between wells would also have likely been detected, with similar phenotypes being observed in localised regions. No obvious edge effects were observed in this screen, likely due to the exclusion of the outer wells of each 96-well plate, and no evidence of cross-contamination was seen via this view.

a)



b)

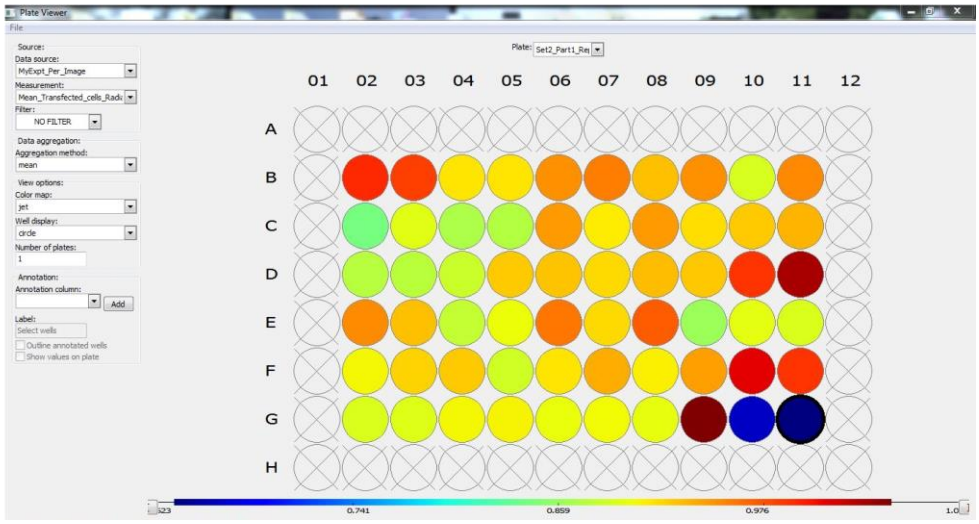


Figure 4.3. **CellProfiler Analyst heat map examples**. a) Set 3 Part 2 Replicate 2 average transfected cell nuclei area, and b) average GFP fraction present in Set 2 Part 1 Replicate 2 transfected cell radial zone 4. Colour scales can be seen at the bottom of each window. Measurements were selected via the dropdown boxes on the left, and plates selected using the dropdown box at the top of the window. Wells G09, G10 and G11 contain p62 transfected, SseD transfected and untransfected cells, respectively. p62 is known to shuttle cargo between the nucleus and cytoplasm, which could lead to the change in nuclear area observed in a). In b) controls wells G10 and G11 report low GFP intensity in zone 4 due to a deliberate manipulation of their GFP/Hoechst channels discussed in section 4.1.3.4: Alteration of untransfected controls.

As well as plate-views, Analyst can also produce some basic data graphs, such as scatterplots and boxplots. However, manipulation of these graphs was difficult and often not possible, and the statistical tests available for application to the data was inadequate. The discrimination between screen hits and normal well-to-well variation was therefore difficult, and alternative methods were sought.

Analyst contains powerful machine learning software called Classifier, with which a user can manually begin to classify phenotypes into two or more bins. From this, Classifier attempts to identify patterns between the images in each bin and expand this to the rest of the image set. This, in theory, provided an extremely powerful method through which phenotypes could be identified. However, when tested for use with this screen, results were inaccurate, likely due to the incorporation of untransfected cells into Classifier's considerations. An alternative solution for data analysis was therefore found: R.

4.5.1.4 CellProfiler Analyst use in this research

Overall Analyst was an indispensable tool for data analysis and phenotype confirmation, particularly due to its ability to relate data with images. However, as a standalone package for hit detection, its ability was lacking for this screen.

4.5.2 R data analysis

R is a statistics and graphics package routinely used for the analysis of large data. It was utilised here to handle and display CellProfiler data, as well as to perform statistical analysis upon said data.

4.5.2.1 R use with CellProfiler

R is not specifically designed to work with CellProfiler, however, it can handle most data or databases via the use of user-installed, community-created packages. One such package, RSQLite, was utilised here to import the SQLite database created by the CellProfiler pipeline into R (Wickham et al. 2015).

Importing an SQLite database into R is not trivial, but is necessary to increase the ease of use of the data downstream. Once imported, SQLite databases were exported as text files, which R can utilise without additional software packages. This conversion process therefore made both the requirement of the RSQLite package, as well as the use of complex coding, unnecessary for any other downstream users.

4.5.2.2 Data processing using R

Text files were processed with the aid of two R packages – ggplot2, and RcmdR – as well as built in R functions (Wickham 2009; Fox 2005). The ggplot2 and Rcmdr packages were utilised to draw graphs, and to aid analysis, respectively. Ggplot2 is a well-known package for the production of complex and multi-layered publication quality graphics, and has been used for publications in multiple fields (Ito & Murphy 2013). Rcmdr, on the other hand, provides a more user-friendly graphical user interface than that provided with R.

Each table within the database created by CellProfiler was assigned its own text file. Three of these tables contained data relating to transfected cells, and were combined to create one large table once imported. Each table contained measurements of individual transfected cells, nuclei and cytoplasm objects. Key measurements of interest were then selected for each object (see Table 4.1.). These consisted of: cell perimeter; nucleus area and perimeter; correlations between GFP and Hoechst, GFP and phalloidin, and GFP and mCherry; radial distributions of GFP in cell fractions 1, 2, 3 and 4.

Autophagosome measurements were also considered and are discussed later (see: Limitations of autophagosome enumeration in the CellProfiler pipeline).

Table 4.1. **List of screen parameters of interest**. A list and short description of the main parameters of interest obtained from the CellProfiler screen database.

Parameter	Desired use
Total autophagosomal area	Detection of increased autophagy
Correlation between GFP and DNA	Suggestive of GFP/expressed CDS localisation to the nucleus
Radial zone 1	Detection of GFP localisation to the nuclear/sub-nuclear region
Perimeter of transfected cells	Detection of an increase or decrease in the area of transfected cells
Perimeter of transfected cell nuclei	Detection of an increase or decrease in the perimeter of transfected cell nuclei
Area of transfected cell nuclei	Detection of an increase or decrease in the area of transfected cell nuclei
Correlation between GFP and LC3	Suggestive of GFP/expressed CDS localisation to autophagosomes
Correlation between GFP and actin	Suggestive of GFP/expressed CDS localisation to cell cytoskeleton
Radial zone 4	Detection of GFP localisation to the outer cell cytoplasm region

4.5.2.3 Data presentation using R

The ggplot2 package was used to present data as boxplot graphs. Boxplots were chosen as they allowed an 'at a glance' overview of the data, and, when combined with statistical data (see below), the subsequent identification of potential screen hits. Each measurement per object (transfected cell object, or nucleus object) was plotted against its respective well. Both sets of well data from each replicate were then pooled into one graph. Each well that had been transfected with pCMV-mEm-4GS (wells G02 and G03) within each plate were pooled together. These were plotted on graphs labelled as "S#P#_Rep1" and "S#P#_Rep2", where 'S#' refers to the current Set number,

and ‘P#’ refers to the current Part number. Examples of these graphs can be seen throughout the chapter, starting with Figure 4.10.

From each set of graphs, R was used to calculate the medians and standard deviations of data from each well. Sample values from each well were first divided by the mean of the values of ‘untreated cell’ control wells, in line with the procedures recommended by Bray & Carpenter (2012) for inter-plate normalisation. From this data, medians and standard deviations were calculated and sorted, resulting in the top 20 hits from each graph set. This provided an accessible list of top hits for each measurement of interest. Examples of these tables can be seen throughout the chapter, starting with Table 4.2.

4.5.2.4 Statistical analysis

Data from each well within each replicate was subjected to a Wilcoxon-Mann-Whitney test, comparing each test well against the pooled mEm-only control wells from the same source plate. For example, each set of object measurements from well B04 of Set 1 Part 2 Replicate 1 was compared to the pooled object measurements from wells G02 and G03 of Set 1 Part 2 Replicate 1. Comparisons were limited to control wells within the same source plate in an attempt to prevent the identification of false positives due to inter-plate variance. For example, comparing a well from Set 1 Part 2 Replicate 1 with mEm-only controls from Set 1 Part 2 Replicate 2 may have resulted in the mis-reporting of a significant difference purely due to a slight difference in seeding densities between replicates.

The above method was used to determine unadjusted p-values, which were then subjected to the Benjamini & Hochberg p-value adjustment, thereby controlling for false detection rate. Only if *both* replicates were determined to be significantly different from their respective mEm-only control wells ($p < 0.05$) were they annotated as such on graphs. Two confidence levels were

used for this: $p < 0.05$, and $p < 0.01$. These are represented on graphs in as either red or blue asterisks, respectively.

For each set of duplicate plates, the 'sample-wide' (both replicates) median was determined for each measurement, along with the upper and lower quartiles. These were displayed on each graph as black, blue and red dashed lines, respectively. Together with the Wilcox-Mann-Whitney test results, these statistics allowed reliable interpretation of the data presented on most graphs. Where boxes on each boxplot graph lay outside the plate-wide upper and lower quartiles, and were shown to be significantly different from controls, manual examination of the corresponding well images was performed with the aid of CellProfiler Analyst, allowing confirmation or rejection of the reported phenotype.

4.5.2.5 Graph inconsistencies

The identification of strong phenotypes was not expected from every LF82 CDS expressed, with a majority of CDSs expected to show no clear phenotype at all. From any given graph, a majority of boxplots were therefore expected to lie between the upper and lower quartile lines, and display no significant difference compared to control wells. While this was the case for most measurement graphs, it was not the case for others.

One such measurement graph was that of GFP/Actin correlation. In this example, seen in Figure 4.25 (see: GFP/LC3 and GFP/Actin), the majority of boxplots can be seen to lie within the upper and lower quartile lines, however, for some sets, a majority of wells appear to have resulted in being reported as significantly different from control wells. While this kind of discrepancy is also seen in other graphs, it is largely possible to identify true hits by their reported significance and their appearance above or below their respective upper or lower quartile lines. With GFP/Actin, however, this was not the case; even where significance and presence outside relevant quartiles

were in agreement, no observable correlation between GFP and actin was seen.

An example of this can be observed with well C05 in Set 3 Part 2. This well displays one of the highest correlations for GFP/Actin in the entire screen. However, as seen in Figure 4.4, no obvious correlation can be seen with GFP and actin.

It is thought that the resolution and magnification utilised for the screen was simply not great enough to observe the actin channel in enough detail for comparisons to be made. While future screens of this nature should look to rectify this, they may also consider the use of positive controls for actin localisation, perhaps using mEm fused to actin-related protein-3 (ARP3) which is known to localise to actin.

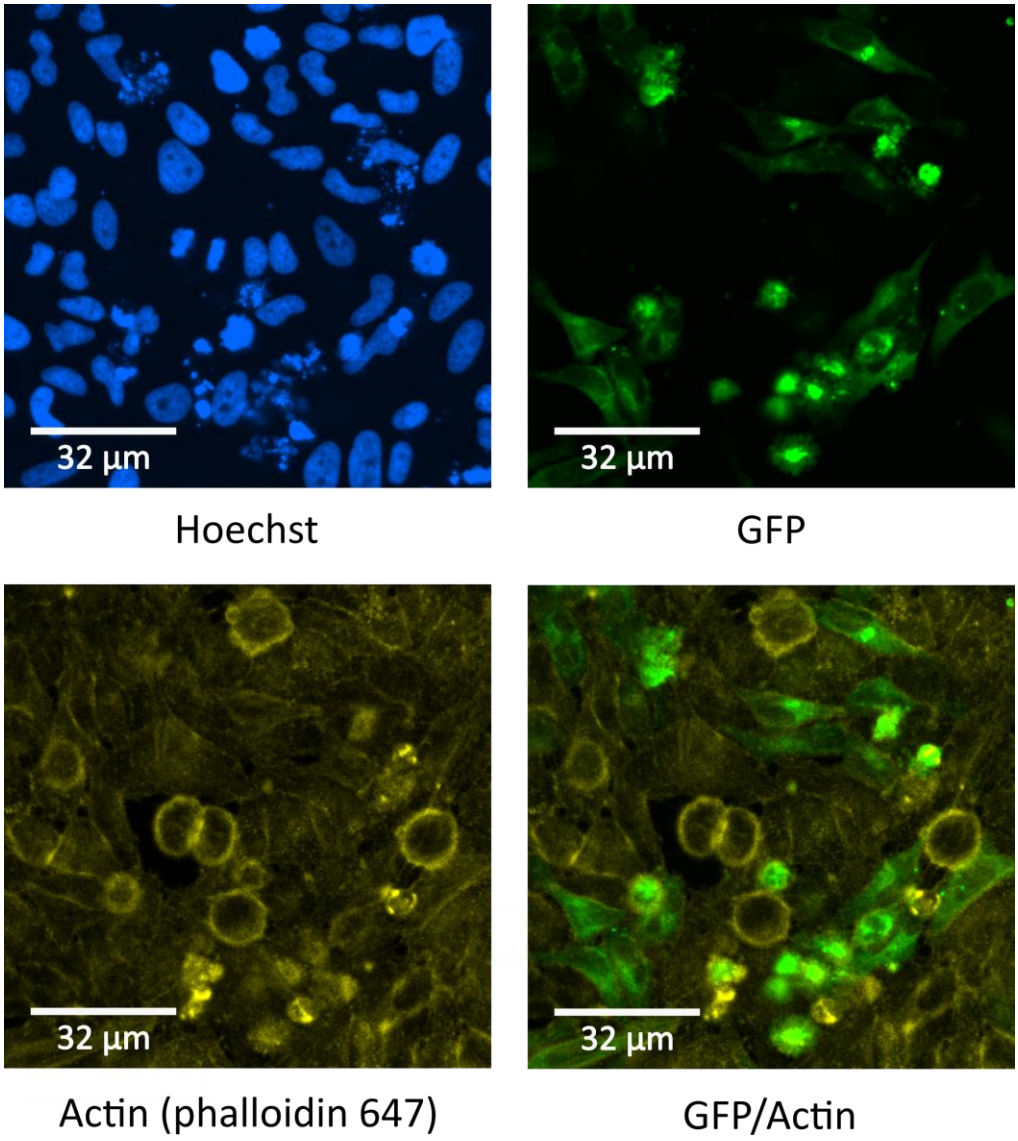


Figure 4.4. **Example of well images displaying a high level of GFP/Actin correlation.** Example images from well C05 of Set 3 Part 2 Replicate 1 (site 4). GFP/Actin measurements identified this well as one in which GFP displayed a high level of co-localisation with actin, however this is difficult to confirm using the images obtained during the screen.

4.5.3 Detection of autophagosomes

The use of LC3 for the detection of autophagosomes is well established (Pugsley 2016; Kabeya et al. 2004). However, while detection of autophagosomes by eye is/was of relative ease, their identification in a high-content setting was more challenging.

4.5.3.1 Successful detection of autophagosomes using an automated pipeline

Autophagosomes are subcellular inclusions that can vary in size from 300 nm to several μm (Jin & Klionsky 2014). This variability in autophagosome size, as well as number, can be used to help determine autophagy status. Labelling autophagosomes with mCherry via fusion to LC3 should therefore allow autophagosomes to be identified, and subsequently enumerated and measured, thus allowing an estimation of the autophagy status of an individual cell.

Prior to performing analysis of the screen data using CellProfiler, the ability to detect and enumerate autophagosomes during an automated pipeline was evaluated. This was performed using two methods: firstly via transfection of p152 (see below), and secondly using selected control wells from the screen.

LF82_152 (p152) is an LF82 CDS studied previously in the lab. It was initially identified due to its apparent effect upon autophagy, causing an increase in the number and size of autophagosomes when overexpressed in comparison to negative controls.

HeLa mCherry-LC3 cells were transfected with pCMV-3xFLAG-4GS-p152 and empty pCMV-3xFLAG-4GS vector. FLAG tags were detected via immunofluorescence, and images were acquired from each set of transfections, as seen in Figure 4.5. Images from these transfections were then subjected to a similar CellProfiler pipeline as used for the main screen,

and, as seen in Figure 4.6, this resulted in the successful identification of autophagosomes. An increase in number and size of autophagosomes was detected in the cells expressing p152.

While p152 transfection led to an apparent increase in the number of autophagosomes and/or an increase in autophagosome size, the images used for this experiment were not representative of those obtained from the screen as they were taken at a higher magnification, and transfections were performed using midiprepped DNA. Therefore the ability of the pipeline to detect autophagosomes was again tested using control wells containing p62 (see below).

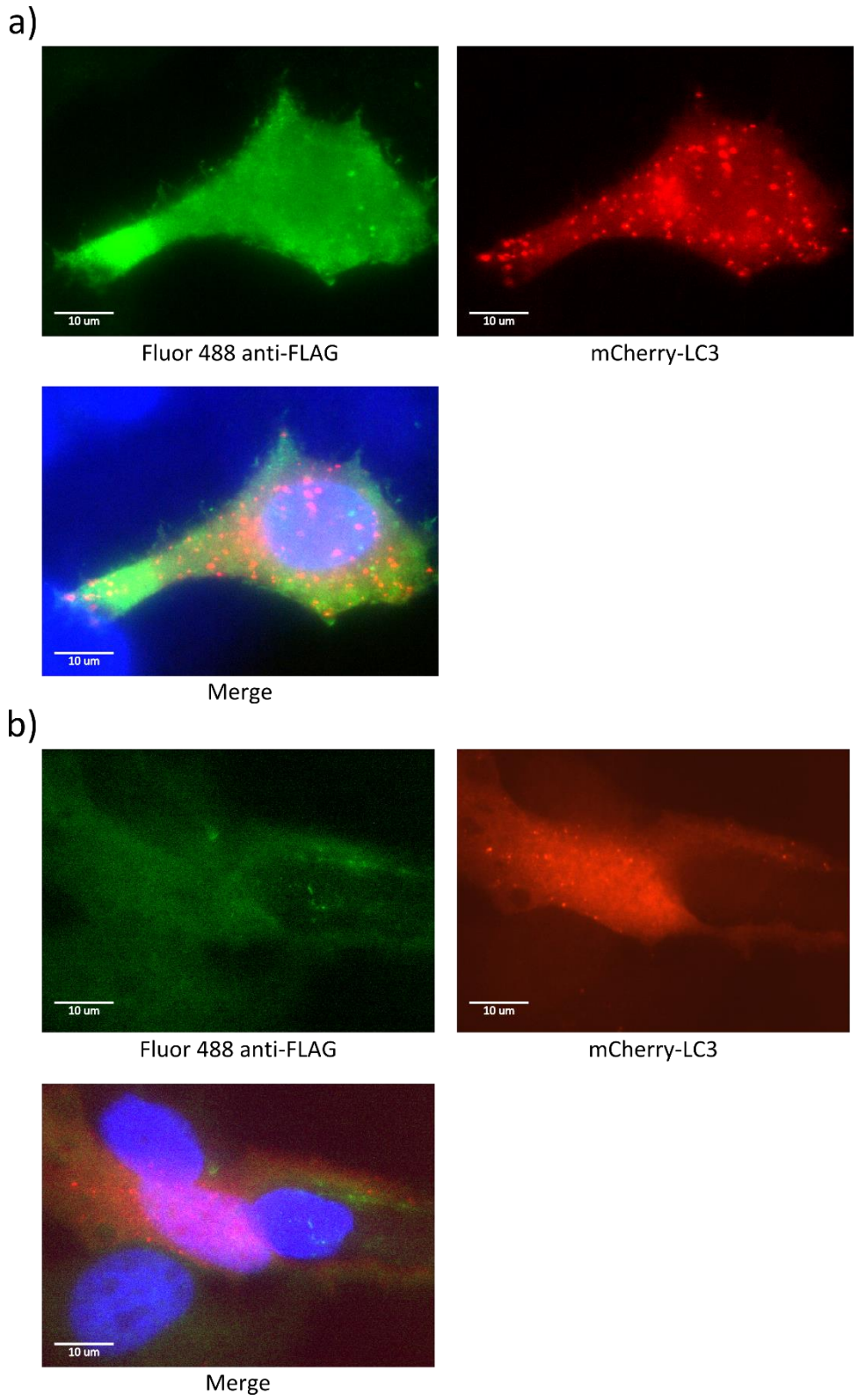
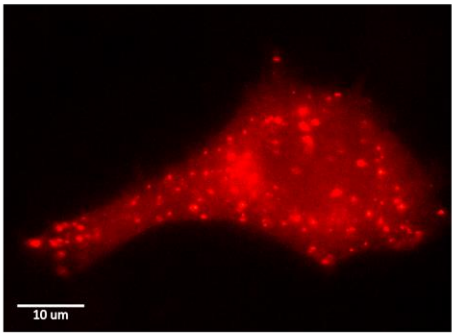
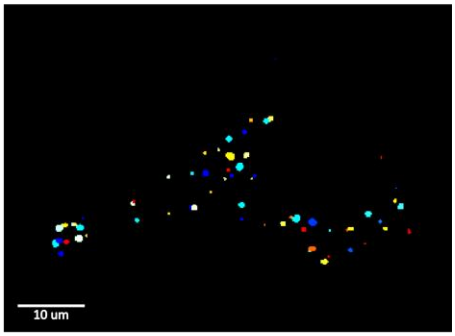


Figure 4.5. **Images of HeLa cells used to test CellProfiler for the detection of autophagosomes**. HeLa mCherry-LC3 cells were transfected with a) pCMV-3xFLAG-4GS-p152 and b) empty pCMV-3xFLAG-4GS vector.

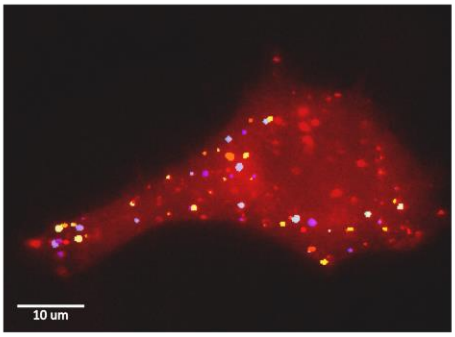
a)



mCherry-LC3



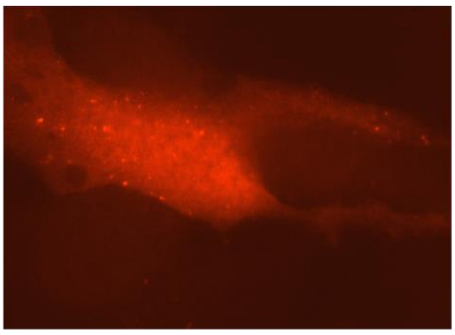
CellProfiler identified autophagosomes



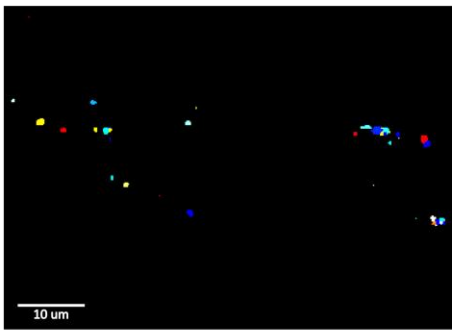
Merge

$p152^{+}$ autophagosome object measurements	
#objects identified	61
Total object area (pixels)	10032
Total cell area (pixels)	660085
Percentage covered by objects	1.5%

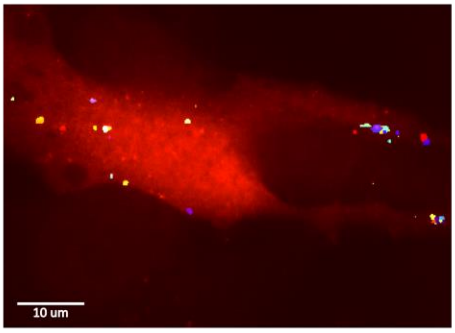
b)



mCherry-LC3



CellProfiler identified autophagosomes



Merge

$p152^{-}$ autophagosome object measurements	
#objects identified	23
Total object area (pixels)	5661
Total cell area (pixels)	558068
Percentage covered by objects	1%

Figure 4.6. **Images from a CellProfiler pipeline showing its ability to detect autophagosomes.** Autophagosomes were detected in both p152-expressing cells (a), and mEm-only expressing cells (b). Tables show respective number of autophagosomes identified, along with area of respective cells covered by objects.

4.5.3.2 Autophagosome identification during pipeline development

While it had previously been shown that CellProfiler was capable of identifying autophagosomes at higher magnifications (and via manual microscopy), this capability was evaluated for use during the development of the screen pipeline, due to the use of different microscopic techniques. To do this, control well images were tested for ease of autophagosome identification.

Well G09 can be seen to have succeeded in this testing via the significant increase in autophagosome area seen compared to other wells on graphs (Figure 4.9). This shows that a strong phenotype regarding autophagosomes, such as that exhibited by p62 controls in well G09, can be detected with the pipeline used for screen analysis. An example of this phenotype can be seen in Figure 4.7.

4.5.3.3 Limitations in the identification of autophagosomes

Initial attempts to identify autophagosomes using CellProfiler were therefore successful, with p62 controls resulting in larger and more numerous autophagosomes identified compared to untransfected controls. However, when this autophagosome identification module was applied to the pipeline for screen image analysis, it appeared less successful than expected.

The settings tested using the autophagosome identification module worked well for most test images, but it was difficult to find settings which provided accurate autophagosome detection when applied to all screen images. While it was unlikely that one setting could be used to achieve perfect autophagosome detection for all images, the success rate achieved was below expectations, with the module struggling to differentiate between autophagosomes and basal mCherry-LC3 expression. However, it was later realised that it may have not been possible to detect some of the smaller autophagosomes due to restrictions in the resolution of the microscope used.

For example, autophagosomes smaller than 500 nm would be impossible to visualise given the 561 nm wavelength utilised for mCherry excitation.

Despite these potential limitations, however, it does appear that p62 controls (well G09 on graphs (see Figure 4.9)) consistently resulted in an increased number and/or size of autophagosomes detected. While this data suggests that p62 expression was usable as a successful control, other control wells did not display this success. Wells treated with NH_4Cl (wells G04 and G05) – a chemical known to result in larger autophagosomes due to interrupted downstream processing – did not result in any data to suggest an increase in autophagosomal area. Similarly, wells G06 and G07 inconsistently resulted in an increase in normalised autophagosome area, despite treatment with rapamycin, a known inducer of autophagy. Data for NH_4Cl and rapamycin treated cells was therefore not usable as control data in the comparison of normalised autophagosome areas between test wells.

From the autophagosome data, it appeared that very few LF82 unique CDSs affected autophagosome numbers or sizes, as seen in graphs in Figure 4.9 (see: Graphs displaying autophagosome data, later). While this was not unexpected, the microscopy data associated with some boxplots (which displayed an average increase in overall autophagosome area) did not support the proposed phenotype when checked manually. In fact, it appeared that some of the images associated with wells for which an increased autophagosome area was reported, were improperly imaged. The GFP channels for each of these wells were noisy, probably due to low GFP signal, and their apparent reported phenotypes were false.

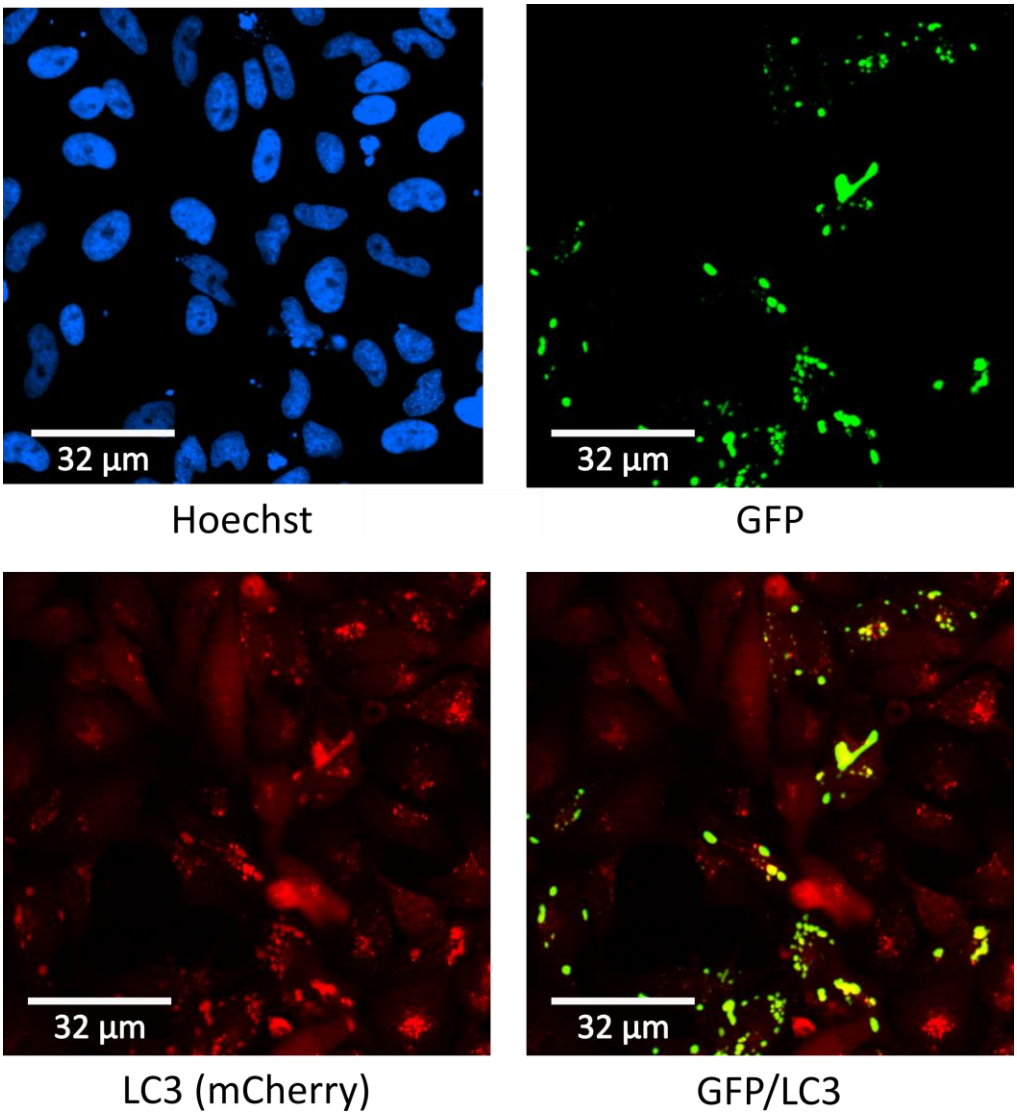


Figure 4.7. **Example of well images displaying p62 expression**. Representative images from well G09 of Set 1 Part 2 Replicate 2 (site 1). The phenotype displayed here was detectable via the Normalised autophagosome area graphs, and can be seen to result in an increased number and size of autophagosomes in comparison to other samples. p62 was expressed in this well, which is known to induce autophagy when overexpressed.

The presence of false positives due to noise in the GFP channel of associated images was true for wells B02 and D11 from Set 1 Part 2, which displayed some of the highest signals for a phenotype affecting autophagy across the entire screen (in terms of boxplot variance from the plate average upper quartile).

Well F02 in Set 3 Part 1 (LF82_1175), however, did not display noise in its GFP channel images and may represent a true phenotype with an effect upon autophagy. Images from this well can be seen in Figure 4.8. However, due to the limitations of CellProfiler's object classification (see below), this phenotype remains difficult to confirm.

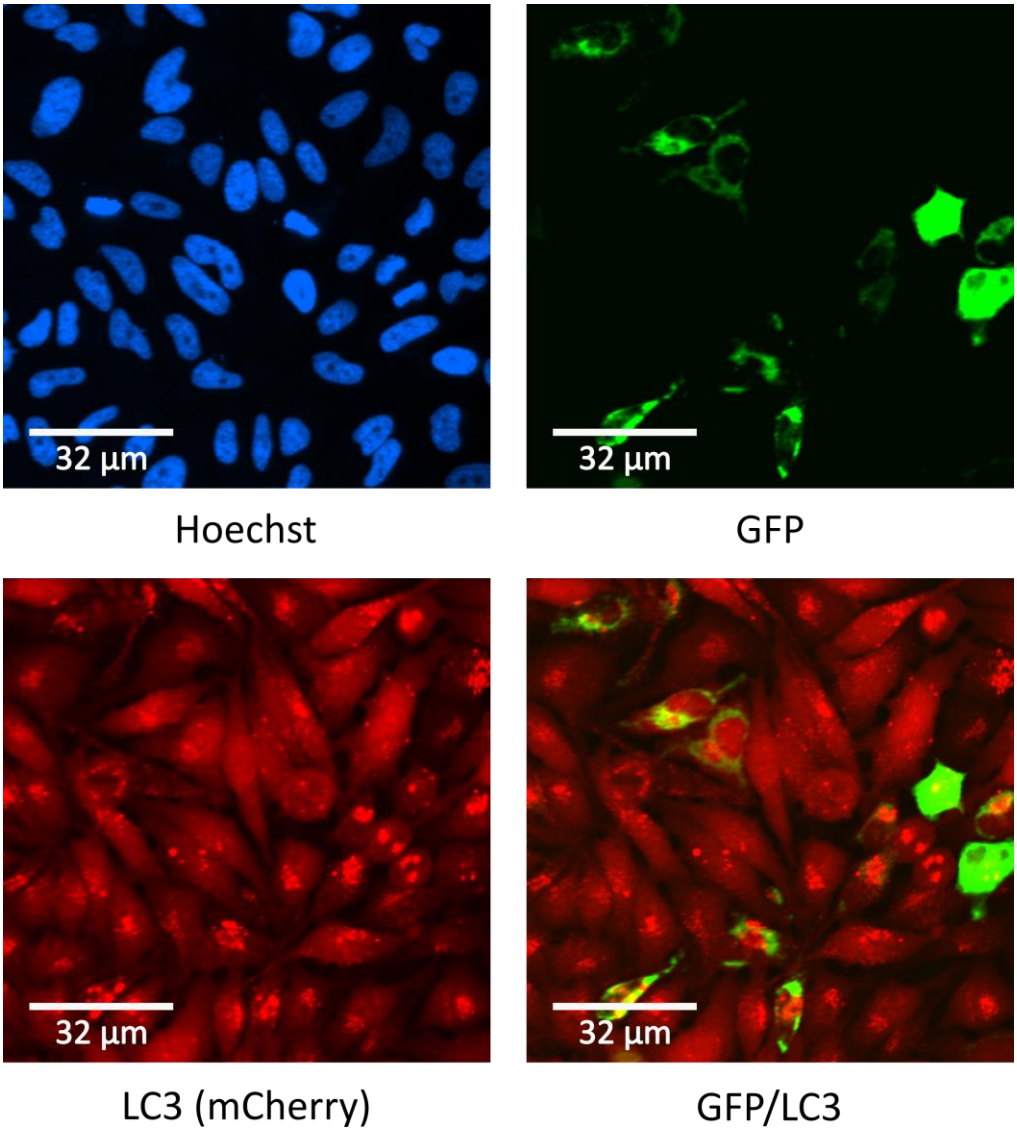


Figure 4.8. **Example of well images displaying a potential effect upon autophagy**. Well F02 from Set 3 Part 1 Replicate 1 (Site 3). This sample was found to show an increased area covered by autophagosomes, when observing the normalised autophagosome area graphs (see Figure 4.9). In comparison to false positives, GFP channels from this well showed no noise in the GFP channel.

4.5.3.4 Limitations of autophagosome enumeration in the CellProfiler pipeline

Problems analysing autophagosomes extended past their simple identification from screen images. As described below, this was due to CellProfiler's 'object' based classification system.

The only module capable of identifying autophagosomes was the 'primary object' identification module, which is usually used to identify objects of interest that may contain other – secondary or tertiary – objects. Primary objects are typically cells in a screen of this nature. Secondary objects are any object associated with a primary object, e.g. nuclei, mitochondria, or endosomes. Multiple secondary objects can belong to a primary object, but multiple primary objects cannot belong to one secondary object.

Autophagosomes fall into the category of secondary objects, but could not be identified as such using the CellProfiler pipeline used here due to limitations in the CellProfiler 'secondary object' module. The best method found for the identification of autophagosomes was via the 'primary object' module. However, while this allowed identification of autophagosomes – albeit inconsistently – they could not be associated with their parent primary object. Thus no numerical data was available for autophagosome counts per transfected cell, only measurement data per total transfected cells within each site.

In an effort to obtain a normalised autophagosome value per site, autophagosome area was considered, not number. This was to account for both small and large autophagosomes, as a low number of large autophagosomes (or vice versa) may skew autophagosome counts. Furthermore, the issues in identifying autophagosomes for enumeration may mean that multiple separate objects were inadvertently grouped together

during identification. Total autophagosome area was therefore divided by the number of transfected cells identified for each site. This resulted in what was termed 'normalised autophagosome area' and is what can be seen plotted in Figure 4.9.

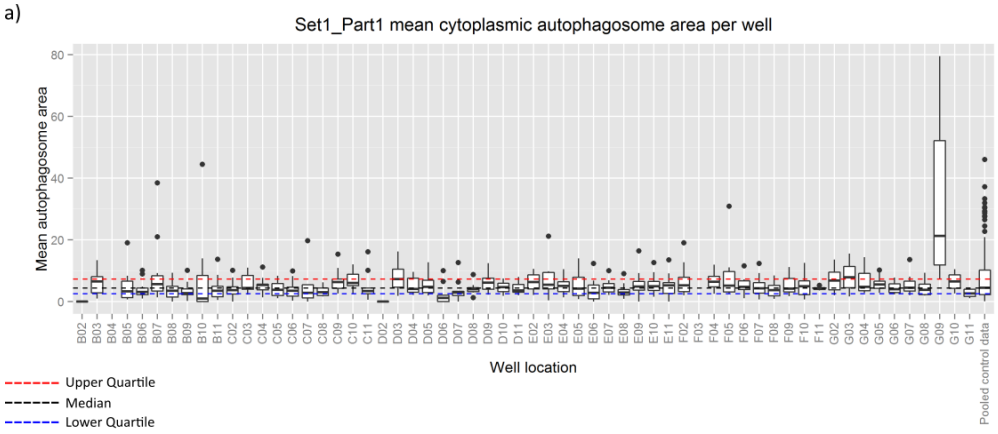
4.5.3.5 Graphs displaying autophagosome data

Despite the efforts put into designing a pipeline for autophagosome identification, enumeration, and measurement, the resulting data was not as useful as desired. The data regarding the ability of this screen to classify factors which may induce or repress autophagy was inconclusive. For these reasons, boxplot graphs representing normalised autophagosome area data were not completed, and their associated data was not subjected to statistical analyses.

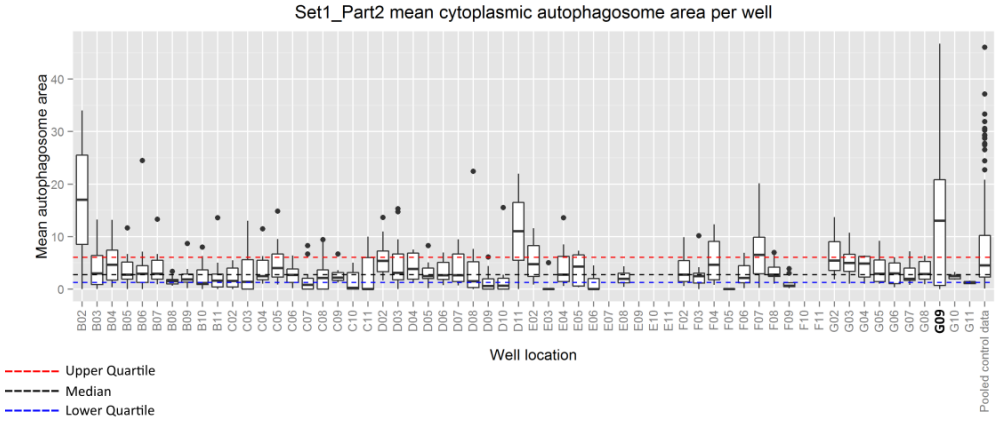
To improve this screen, an increase in magnification and resolution may be of use. However, increasing magnification would require more sites to be imaged per well to acquire enough data to support the observed phenotypes. This would lead to an increase in the amount of time taken to perform the image acquisition process, which may or may not be significant depending on the screen size. An increase in the resolution of microscopy may also increase the time taken to acquire images for any respective screen, and would almost certainly result in a significant increase in cost.

Finally, alternative methods for autophagosome identification and quantification may be used, perhaps via the use of alternative image processing software (such as ImageJ) mid-way through each pipeline cycle – a feature already present in CellProfiler. Nevertheless, this would also likely increase the time taken for screen completion, in-part due to the computational workload increase of analysing each image set in two pieces of software.

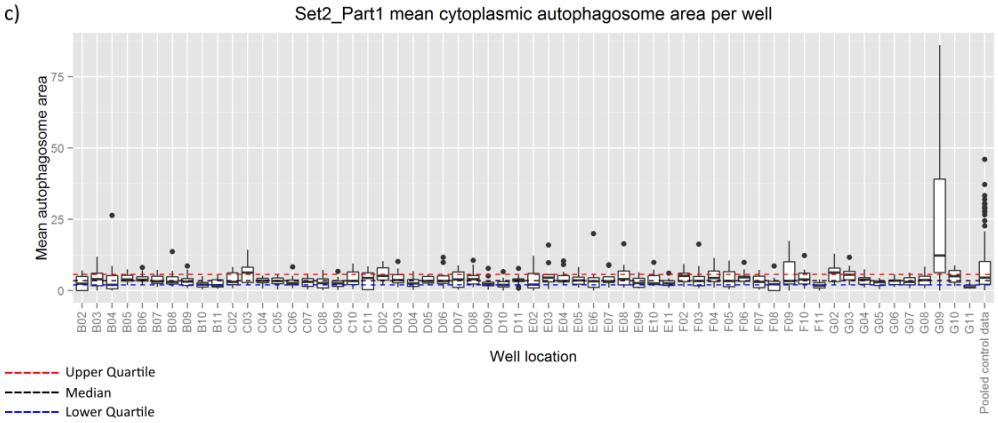
a)


Set1_Part1 mean cytoplasmic autophagosome area per well
Mean autophagosome area
Well location
Pooled control data
Upper Quartile
Median
Lower Quartile

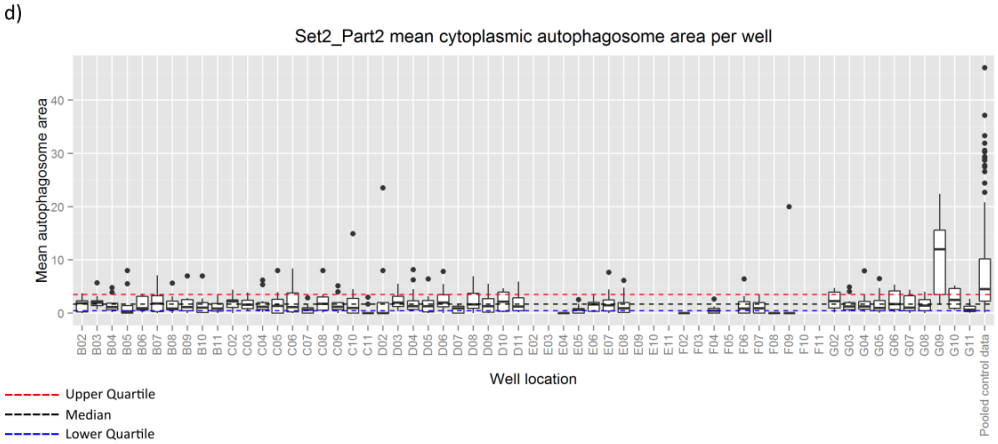
b)


Set1_Part2 mean cytoplasmic autophagosome area per well
Mean autophagosome area
Well location
Pooled control data
Upper Quartile
Median
Lower Quartile

c)


Set2_Part1 mean cytoplasmic autophagosome area per well
Mean autophagosome area
Well location
Pooled control data
Upper Quartile
Median
Lower Quartile

d)


Set2_Part2 mean cytoplasmic autophagosome area per well
Mean autophagosome area
Well location
Pooled control data
Upper Quartile
Median
Lower Quartile

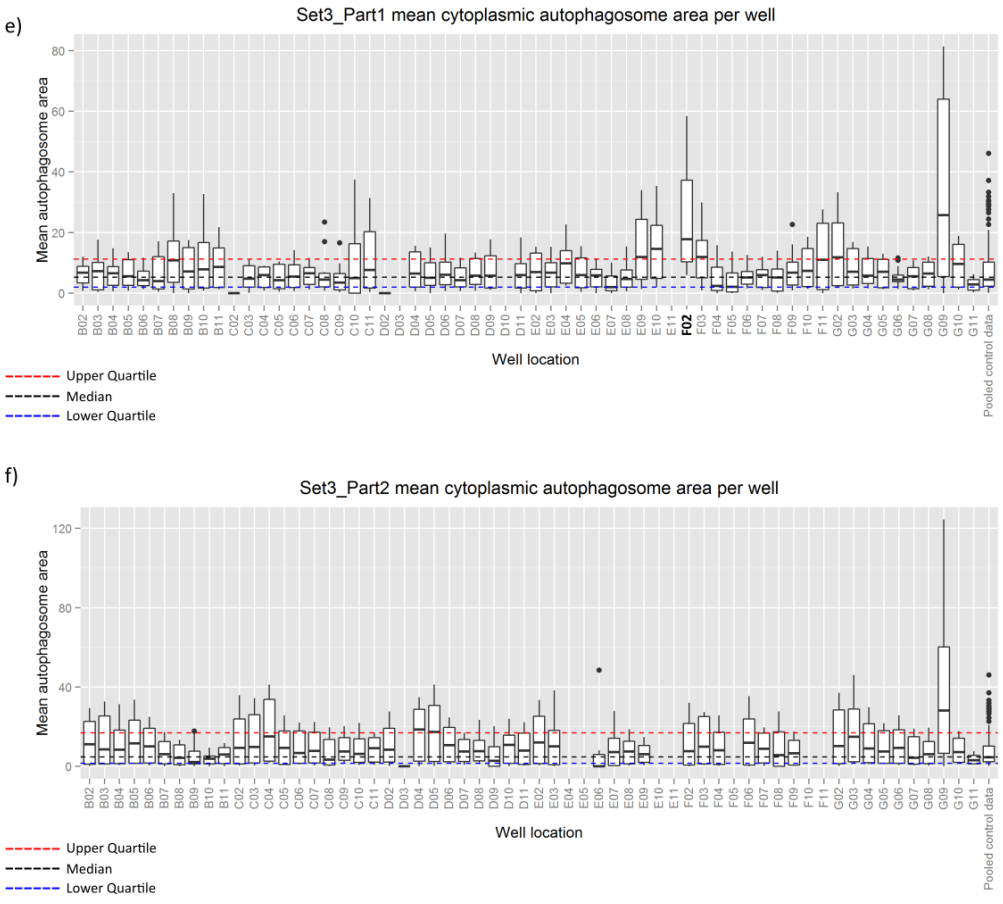


Figure 4.9. **Graphs showing data for normalised autophagosome area.** Normalised autophagosome area was calculated by taking the area of autophagosomes identified in all transfected cells in each site, and divided it by the number of transfected cells. As no individual cell data was available, values are plotted on an averaged individual site bases; there are therefore only 6 data points per well. No markers for statistical significance are present.

Table 4.2. **Top 20 hits from normalised autophagosome data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest. C*: data from contaminated well, discussed in Table 4.11.

Top 20 hits for normalised autophagosome area				
Plate	Well name	Gene	Median	Standard deviation
Set1_Part2	*G09*	*p62*	*18.88*	*18.28*
Set3_Part1	F02	LF82_1175	13.14	3.22
Set1_Part2	F07	LF82_696	10.36	4.63
Set1_Part2	F04	LF82_319	9.30	2.57
Set1_Part2	E02	LF82_102	8.63	3.47
Set1_Part2	D03	LF82_774	7.59	5.08
Set1_Part2	B04	LF82_668	7.48	2.61
Set1_Part2	D02	LF82_770	7.38	2.87
Set1_Part2	B03	LF82_662	6.96	4.69
Set1_Part2	D04	LF82_780	6.90	1.16
Set1_Part2	C05	LF82_719	6.80	3.85
Set1_Part2	D07	LF82_787	6.78	2.16
Set1_Part2	E05	LF82_364	6.71	0.63
Set1_Part2	E04	LF82_314	6.48	3.73
Set2_Part1	F04	LF82_647	6.07	3.38
Set3_Part1	E09	LF82_004	5.86	3.11
Set3_Part1	E10	LF82_772 C*	5.83	2.45
Set3_Part2	C04	LF82_279	5.74	0.84
Set3_Part1	F03	LF82_539	5.71	6.94
Set1_Part2	F02	LF82_104	5.66	2.78
Set1_Part2	B07	LF82_682	5.65	3.61

4.5.4 Identification of genes of interest from screen

The boxplot graphs produced with R were used to identify candidate phenotypes of interest. Plate-wide upper and lower quartiles were considered when determining a potential screen 'hit' from a non-hit, with hits predominantly identified by the shift of their boxplot toward the extremities (or outside) of the plate's respective upper and lower quartiles. Potential hits were further scrutinised by the presence of a red or blue asterisk towards the bottom of the graph, indicating the associated p-value ($p < 0.05$, and $p < 0.01$, respectively) for each result (see: Statistical analysis).

Once individual wells associated with each hit were identified, the corresponding LF82 CDS was looked up in the screen layout diagrams (see: Appendix III: Layout of screen transfection plates). CellProfiler Analyst was then utilised to manually confirm the phenotype observed by displaying the plate from which the hit was found in Analyst, along with the respective measurement of interest (see Figure 4.3). The plate was then scrutinised for any abnormal or localised variance to ensure that the reported phenotype was not due to cross-contamination from other wells, edge effects from excess drying etc. Finally, the microscopy images relating to each well and measurement were viewed individually to determine the credibility of each suspected phenotype.

Key graphs and phenotypes will be considered individually below, while graphs displaying more ambiguous results will be grouped together.

4.5.5 Detection of nuclear localisation

Two key measurements were utilised to detect nuclear localisation of GFP. These measurements were: the correlation between GFP and Hoechst, and the fraction of total GFP within radial zone 1 (radial distribution measurement). Graphs relating to these measurements can be seen in Figures 4.10 and 4.11, respectively.

Control wells G09, G10 and G11 exemplify strong results, with G09 data suggesting exclusion from the nuclear region, and G10 and G11 data suggesting strong nuclear localisation. Well G09 was transfected with a vector containing the p62 construct, resulting in GFP localisation to autophagosomes. This meant that little p62 localised to the nucleus, resulting in a lower nuclear fraction of GFP detected by CellProfiler. Wells G10 and G11 were wells in which Hoechst channels were used to replace GFP channels – as discussed previously (see: Alteration of untransfected controls) – resulting in complete GFP/Hoechst association and a high level of detection within radial zone 1.

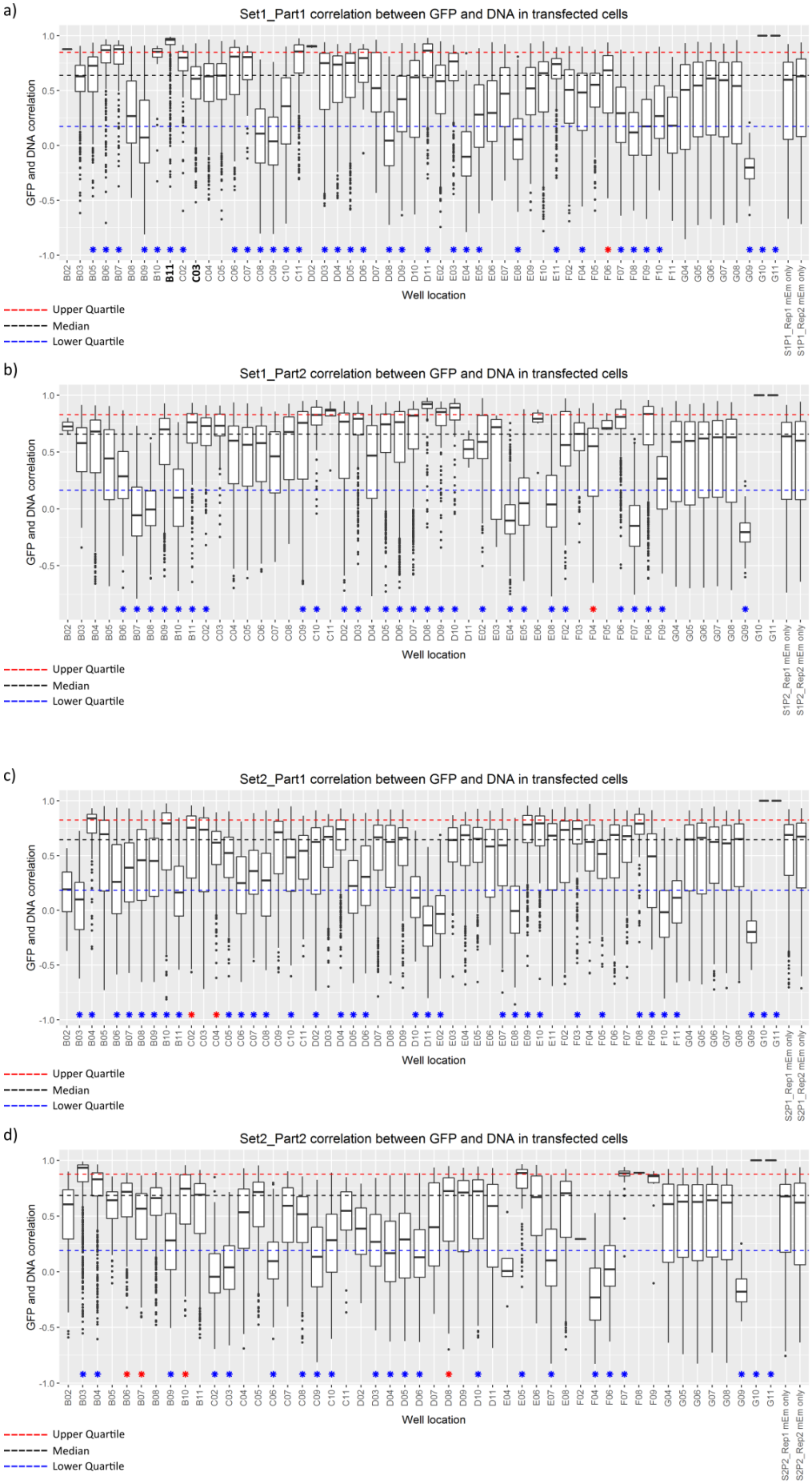
4.5.5.1 GFP/Hoechst and radial zone 1 graph comparison

Both sets of graphs display GFP localised to the region internal of the nucleus. While the GFP/Hoechst correlation graph appears to be far more variable than the radial zone 1 graph, this is likely due to the sub-localisation of GFP within the nucleus. An example of this is well C03 in Set 1 Part 1 which initially appears to be a significant hit from radial zone 1 ($p < 0.01$), but does not display any significant variance from the control wells for GFP/Hoechst correlation. In this case, as seen in Figure 4.12, the GFP localises to sub-nuclear regions, possibly nucleoli. HeLa cells can have multiple nucleoli, as exemplified by Figure 4.13 a. which displays HeLa cells nuclei stained with anti-nucleolin. These images appear similar to those seen in Figure 4.12, with ~6 foci of varying size visible. In contrast, nuclear dots (also known as PML bodies) appear with multiple smaller foci, as seen in Figure 4.13 b.

This example highlights the power of the radial zone measurement (see Figure 4.2), displaying its ability to allow identification of sub-nuclear localisation that would potentially be missed if analysed using the GFP/Hoechst measurement alone. However, in some cases the GFP/Hoechst measurement appears to give a better indication of localisation. For example, well B11 in Set 1 Part 1 (Figure 4.14) appears to be one of the strongest hits observed in the GFP/Hoechst graphs, but does not display as strong a result in the radial zone 1 graph. This suggests that both GFP/Hoechst and radial zone 1 measurements have their place in the detection of GFP localised to nuclei.

In cases where both measurement graphs agree upon localisation to the nuclear/sub-nuclear region, it is likely that this is due to localisation to the nuclear membrane or DNA itself, as opposed to nucleoli or other small sub-nuclear foci. As an example, well B04 in Set 3 Part 2 (Figure 4.10, and Figure 4.11)) appears as a significant hit on both measurement graphs and, as seen in Figure 4.15, does indeed appear to localise to the expected nuclear region.

The inherent method via which GFP/Hoechst correlation is calculated means features such as nucleoli or nuclear pores may not be detected. This is due to Hoechst 33342 only staining DNA, therefore leaving regions which do not contain DNA unstained, showing no correlation between nucleoli-localised GFP and Hoechst signal. The utilisation of the radial zone 1 measurement therefore often appeared useful for the detection of some sub-nuclear localisation phenotypes.


a)
Set1_Part1 correlation between GFP and DNA in transfected cells
GFP and DNA correlation
Well location
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
b)
Set1_Part2 correlation between GFP and DNA in transfected cells
GFP and DNA correlation
Well location
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
c)
Set2_Part1 correlation between GFP and DNA in transfected cells
GFP and DNA correlation
Well location
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
d)
Set2_Part2 correlation between GFP and DNA in transfected cells
GFP and DNA correlation
Well location
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile

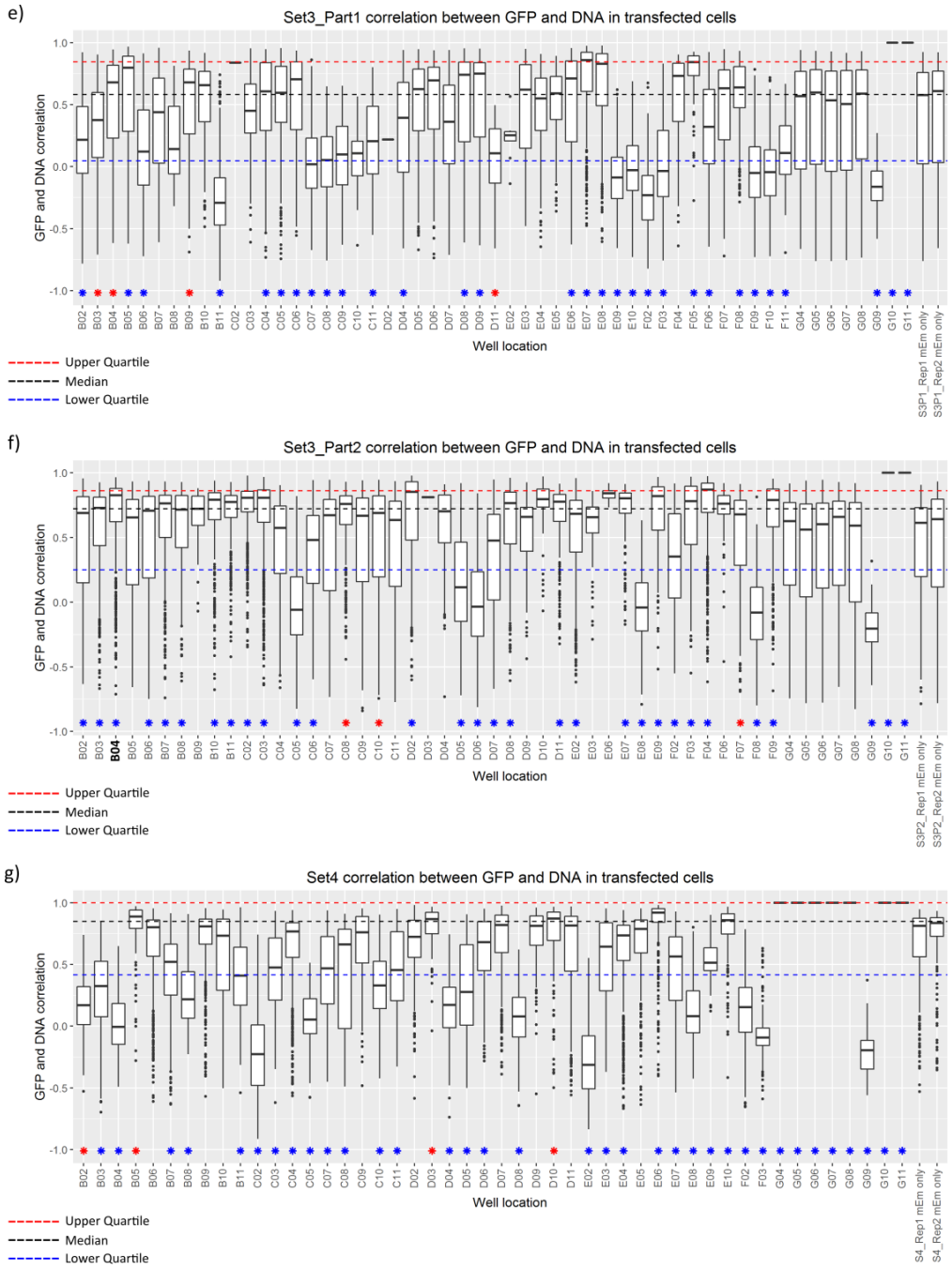
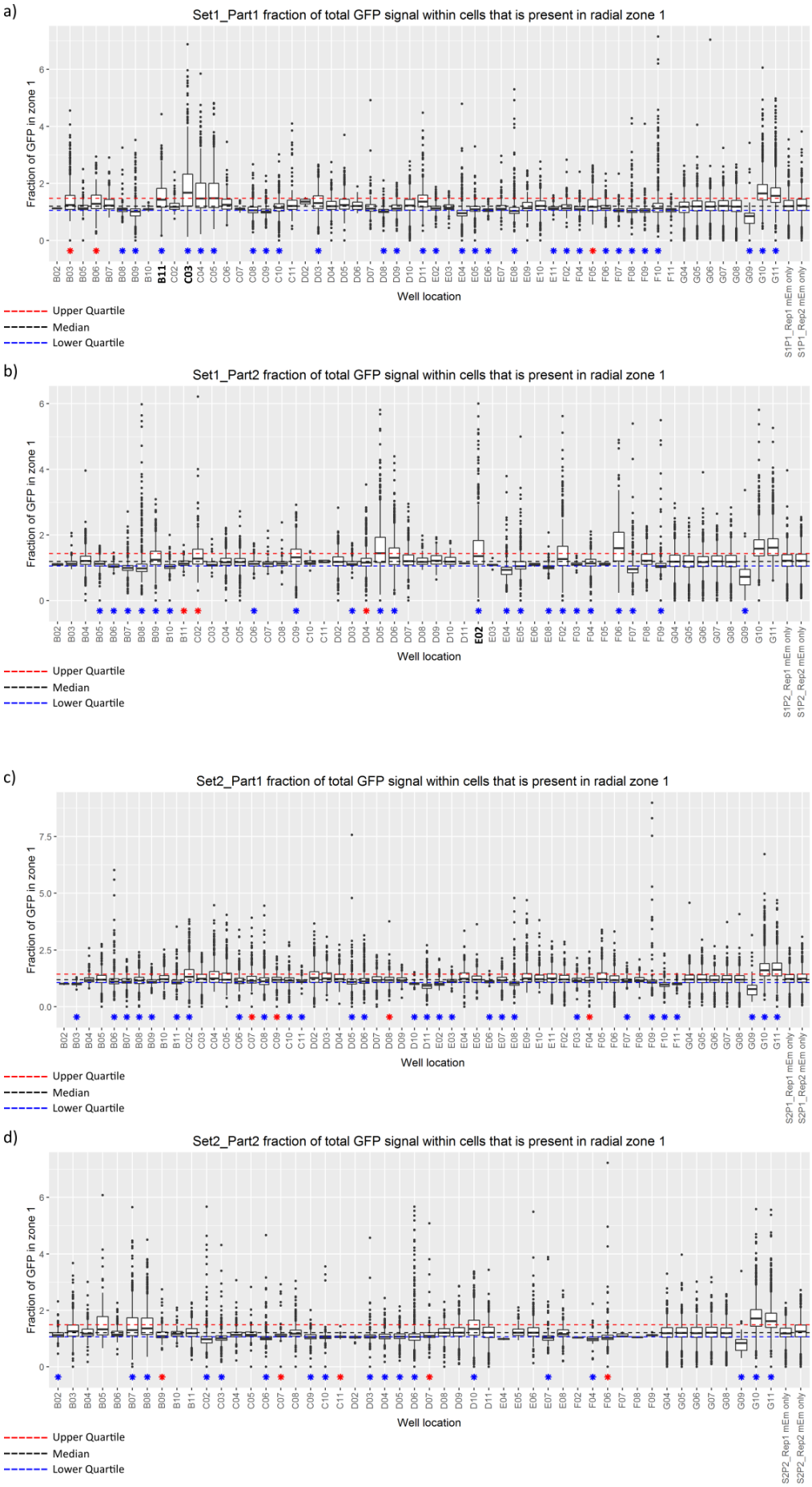


Figure 4.10. **Graphs showing correlation between GFP and Hoechst in transfected cells.** Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.3. **Top 20 hits from GFP and DNA correlation data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest. Note: A median of 1 and standard deviation of 0 was expected due to the manipulations described in section 4.1.3.4: Alteration of untransfected controls.

Top 20 hits for correlation between GFP and DNA in transfected cells				
Plate	Well name	Gene	Median	Standard deviation
Set1_Part1_Rep1	*G10*	*SseD*	*1*	*0*
Set1_Part1	B11	LF82_101	0.96	0.24
Set2_Part2	B03	LF82_658	0.94	0.30
Set4	E06	LF82_649	0.92	0.27
Set1_Part2	D08	LF82_788	0.92	0.29
Set1_Part1	D02	LF82_230	0.90	0.02
Set2_Part2	F07	LF82_676	0.90	0.03
Set1_Part1	D11	LF82_311	0.89	0.26
Set2_Part2	E05	LF82_378	0.89	0.23
Set1_Part1	B06	LF82_071	0.89	0.26
Set4	B05	LF82_246	0.89	0.24
Set2_Part2	F08	LF82_724	0.89	NA
Set3_Part2	D02	LF82_268	0.88	0.34
Set1_Part1	B07	LF82_089	0.88	0.26
Set1_Part2	D10	LF82_792	0.88	0.29
Set1_Part1	B02	LF82_004	0.88	NA
Set3_Part1	F05	LF82_1799	0.87	0.19
Set3_Part2	F04	LF82_691	0.87	0.30
Set2_Part1	F08	LF82_534	0.87	0.23
Set4	D10	LF82_698	0.87	0.33
Set2_Part2	F09	LF82_797	0.87	0.12


a)
Set1_Part1 fraction of total GFP signal within cells that is present in radial zone 1
Fraction of GFP in zone 1
Well location
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
b)
Set1_Part2 fraction of total GFP signal within cells that is present in radial zone 1
Fraction of GFP in zone 1
Well location
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
c)
Set2_Part1 fraction of total GFP signal within cells that is present in radial zone 1
Fraction of GFP in zone 1
Well location
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile
d)
Set2_Part2 fraction of total GFP signal within cells that is present in radial zone 1
Fraction of GFP in zone 1
Well location
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile

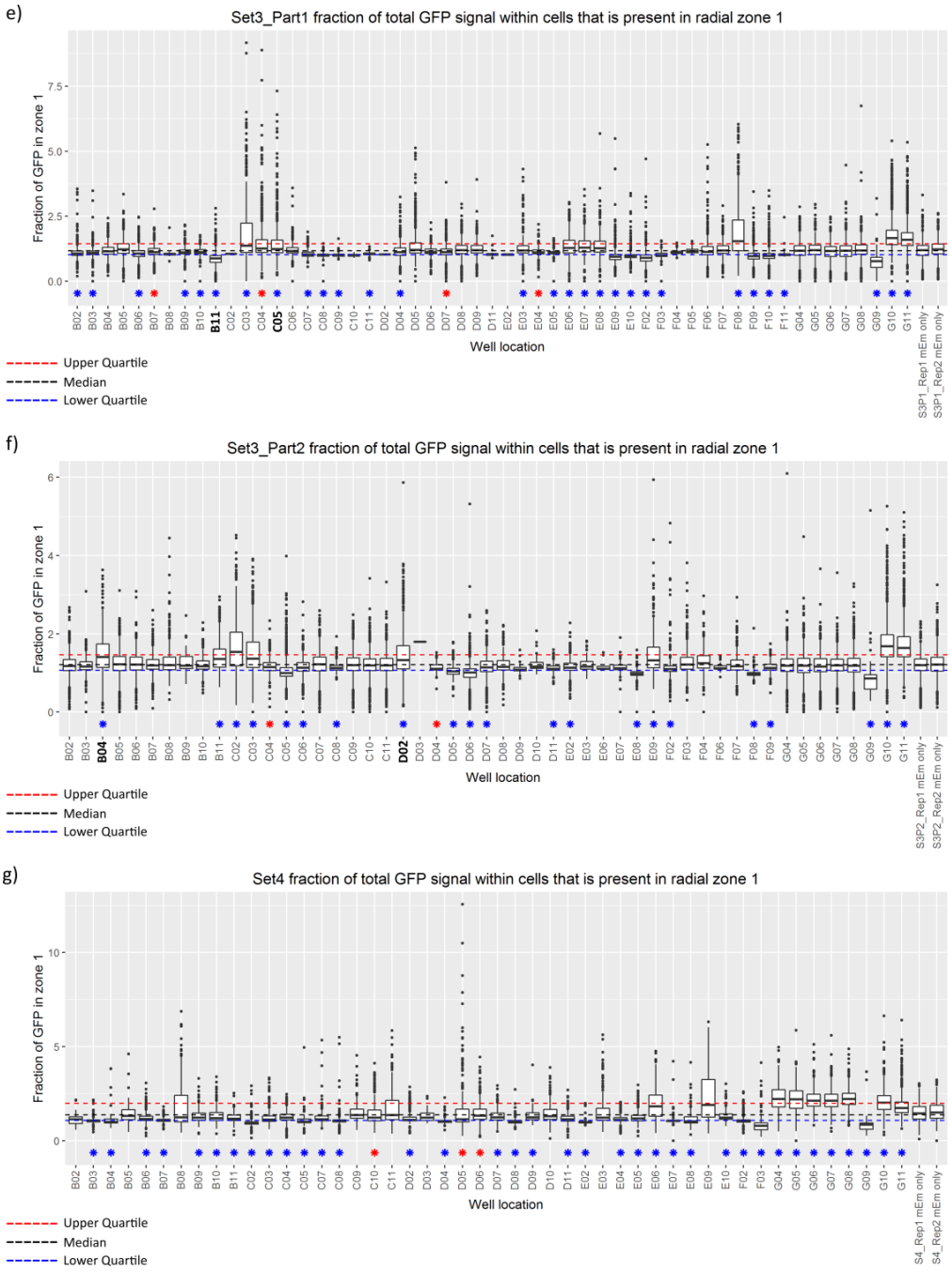


Figure 4.11. **Graphs showing the fraction of GFP signal detected within radial zone 1 of transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.4. **Top 20 hits from radial zone 1 data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest.

Top 20 hits for the fraction of total GFP signal within cells that is present in radial zone 1				
Plate	Well name	Gene	Median	Standard deviation
Set4_Rep1	*G10*	*SseD*	*1.11*	*0.35*
Set1_Part1	C03	LF82_106	1.08	0.59
Set4	E09	LF82_681	1.05	0.82
Set4	E06	LF82_394	1.00	0.46
Set3_Part2	C02	LF82_110	1.00	0.43
Set3_Part2	D03	LF82_312	0.99	NA
Set1_Part1	C04	LF82_114	0.94	0.49
Set1_Part2	F06	LF82_619	0.94	0.38
Set3_Part1	F08	LF82_452	0.93	0.61
Set1_Part1	C05	LF82_129	0.91	0.47
Set3_Part2	C03	LF82_273	0.89	0.38
Set1_Part1	B11	LF82_101	0.87	0.34
Set3_Part2	B04	LF82_0377	0.87	0.34
Set1_Part2	D05	LF82_782	0.85	0.48
Set1_Part1	D11	LF82_311	0.85	0.33
Set3_Part2	D02	LF82_268	0.84	0.38
Set3_Part2	B11	LF82_2825	0.83	0.28
Set1_Part1	D03	LF82_234	0.82	0.24
Set3_Part1	C03	LF82_188	0.82	0.75
Set2_Part1	C02	LF82_808	0.82	0.28
Set1_Part1	D02	LF82_230	0.81	0.13

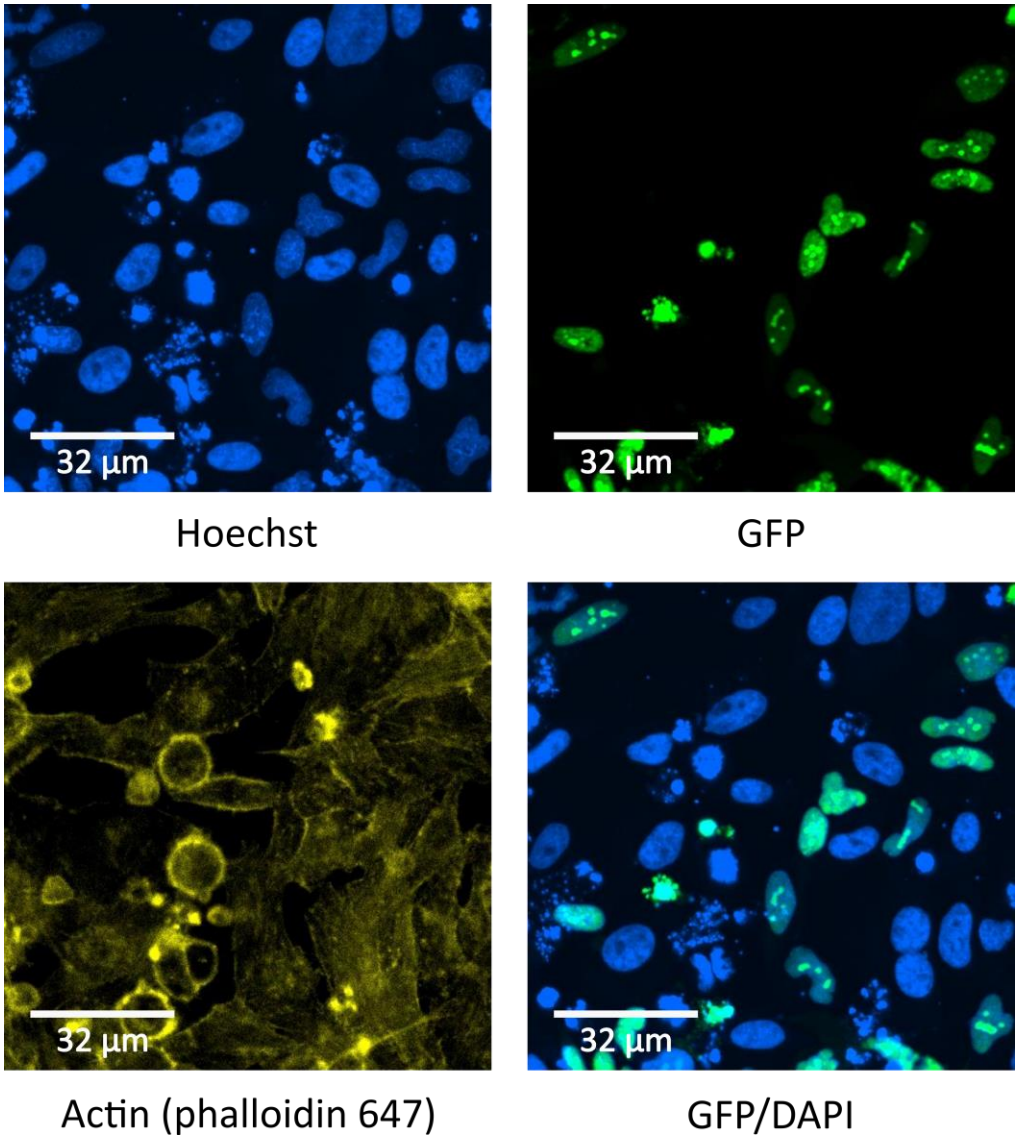


Figure 4.12. **Example of well images displaying sub-nuclear localisation**. Well C03 from Set 1 Part 1 Replicate 2 (site 2). The radial zone 1 measurement was used to identify this well, while the GFP/Hoechst correlation measurement did not report any significant phenotype. The radial zone 1 measurement can be seen to be accurate, however, as GFP can be seen localising to sub-nuclear foci – possibly nucleoli.

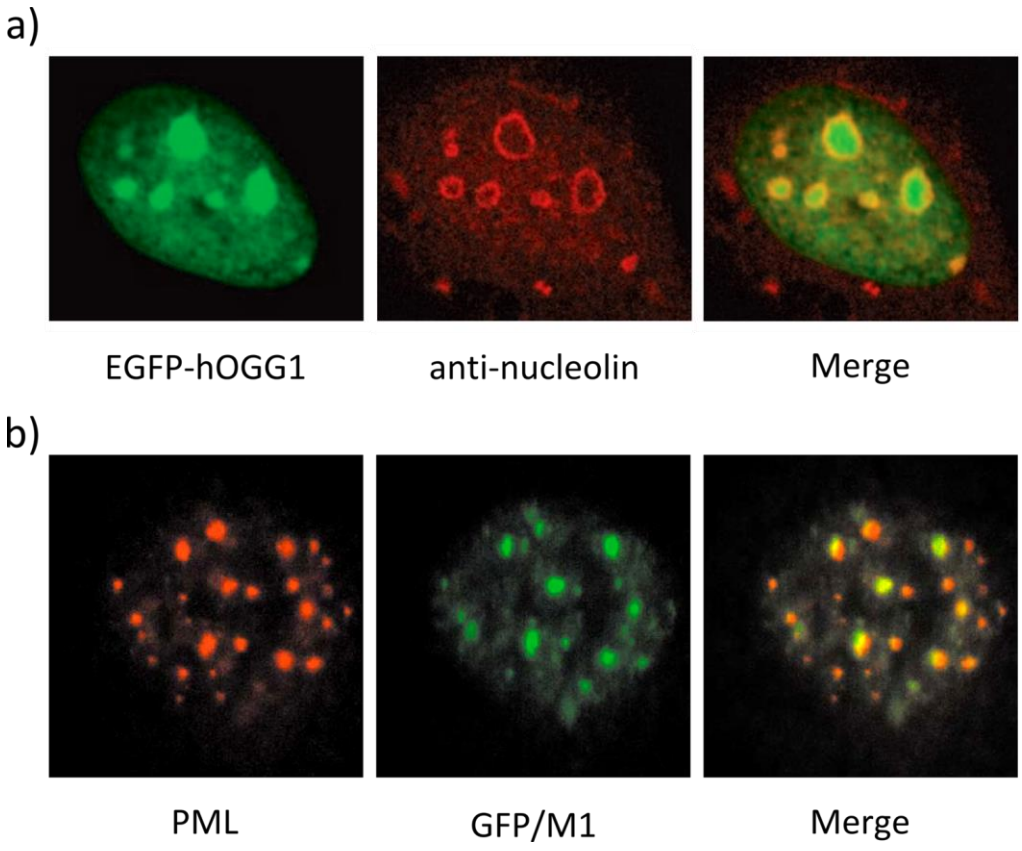


Figure 4.13. **Examples of nucleoli and nuclear dots of HeLa cells**. a) HeLa cells transiently transfected with an EGFP-hOGG1 vector show nuclear localisation with foci at nucleoli, confirmed via detection of nucleoli using anti-nucleolin. Adapted from Luna et al. (2005). b) HeLa cells expressing GFP/M1 – an influenza virus protein – which has been shown to localise to nuclear dots (also known as PML bodies). Adapted from Sato et al. (2003)

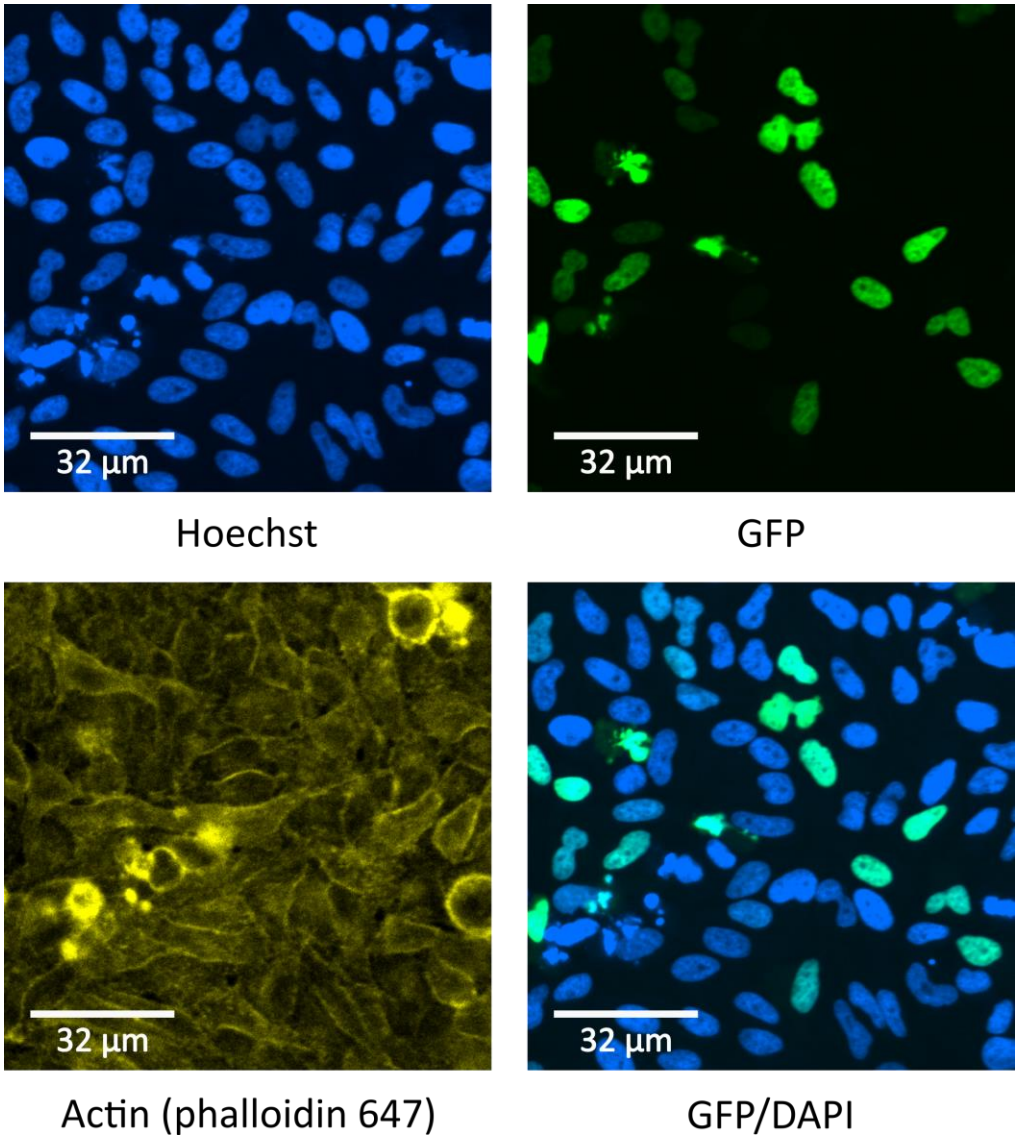


Figure 4.14. **Example well images displaying a nuclear localisation phenotype**. Well B11 from Set 1 Part 1 Replicate 1 (site 5). This phenotype stands out in the GFP/Hoechst graphs as a strong hit, while it is not observed as a strong hit in radial zone 1 graphs.

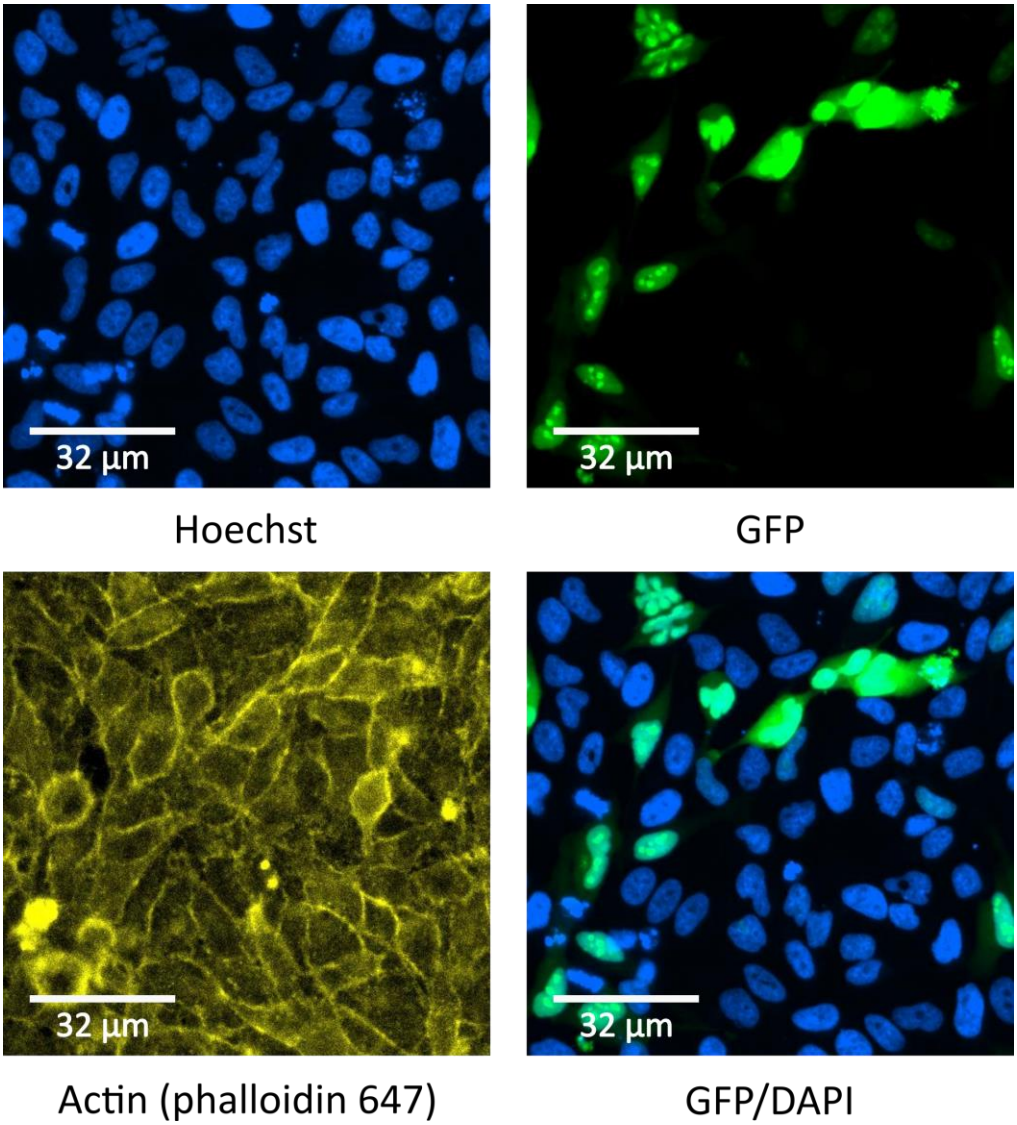


Figure 4.15. **Example well images displaying a nuclear localisation phenotype**. Well B04 from Set 3 Part 2 Replicate 2 (site 4). Both radial zone 1 and GFP/Hoechst graphs were in agreement for the identification of this phenotype. GFP can be seen in this case localising to a sub-nuclear location, possibly to the nuclear envelope and nucleoli.

4.5.5.2 Identification of CDSs localising to nuclei

Due to the limitations in the ability of GFP/Hoechst correlation for identifying CDSs which localised to foci in the sub-nuclear regions, the radial zone 1 measurement was primarily utilised for identifying these phenotypes. A list of CDSs manually confirmed as showing their suspected phenotypes can be seen in Table 4.5. From this table, it can be seen that there were 3 main phenotypes identified: general nuclear localisation, suspected localisation to nucleoli, and suspected localisation to nuclear dots.

Examples of the strongest phenotypes identified are shown in Figures 4.16, 4.17, and 4.18. Figure 4.16 shows well D02 from Set 3 Part 2, displaying a nuclear localisation phenotype, which suggests that the expressed CDS localises to the nuclear membrane. Figure 4.17 shows well C05 from Set 3 Part 1, displaying localisation to several locations in/on the nucleus, resulting in several (typically ~6) foci. This suggests that the expressed CDS localises to sub-nuclear regions such as nucleoli, although this remains to be confirmed. Figure 4.18 shows well E02 from Set 1 Part 2, which also shows localisation to multiple locations in/on the nucleus. However, the number of foci visible here were typically more numerous, with > 10 and up to 50 observed per nucleus. It is unconfirmed where these CDSs are localising, but they possibly resemble nuclear dots (also known as PML bodies), as seen in Figure 4.15 b. (above).

Table 4.5. **Nuclear localisation phenotypes**. Phenotypes observed when manually analysing images from wells whose boxplots were marked as statistically significant and whose boxplot upper quartile exceeded the average upper quartile line for the plate. Localisation to structures resembling endoplasmic reticulum is indicated by "ER", but remains unconfirmed.

Plate	Well	GFP localisation	Respective LF82 CDS
Set 1 Part 1	B03	Punctate	LF82_030
Set 1 Part 1	B06	General nuclear localisation & punctate	LF82_071
Set 1 Part 1	B11	General nuclear localisation	LF82_101
Set 1 Part 1	C03	General nuclear localisation & punctate	LF82_106
Set 1 Part 1	C04	Punctate & "ER"	LF82_114
Set 1 Part 1	C05	Punctate	LF82_129
Set 1 Part 1	D03	Overall cell localisation & punctate	LF82_234
Set 1 Part 1	D11	Grainy punctate localisation	LF82_311
Set 1 Part 2	B09	Overall cell localisation & punctate	LF82_685
Set 1 Part 2	C02	Punctate	LF82_697
Set 1 Part 2	C09	Overall cell localisation & punctate	LF82_750
Set 1 Part 2	D05	Suspected nucleoli	LF82_782
Set 1 Part 2	D06	Punctate (with nuclear blebbing)	LF82_783
Set 1 Part 2	E02	Nuclear dots	LF82_102
Set 1 Part 2	F02	Nuclear dots	LF82_104
Set 1 Part 2	F06	Punctate	LF82_619
Set 2 Part 1	C02	Punctate & "ER"	LF82_808
Set 2 Part 2	B07	Punctate & unidentified cytoplasmic and sub nuclear structure	LF82_666
Set 2 Part 2	B08	Punctate	LF82_670
Set 2 Part 2	D10	Overall cell localisation & punctate	LF82_779
Set 3 Part 1	C03	Nuclear dots	LF82_805
Set 3 Part 1	C04	Punctate	LF82_806
Set 3 Part 1	C05	Punctate	LF82_315
Set 3 Part 1	E06	Overall cell localisation & punctate	LF82_757
Set 3 Part 1	E07	"ER" localisation, (with nuclear blebbing)	LF82_776
Set 3 Part 1	E08	"ER" localisation, (with nuclear blebbing)	LF82_794

Plate	Well	GFP localisation	Respective LF82 CDS
Set 3 Part 1	F08	Punctate	LF82_452
Set 3 Part 2	B04	Punctate	LF82_752
Set 3 Part 2	B11	Punctate	LF82_2825
Set 3 Part 2	C02	General nuclear localisation & punctate	LF82_110
Set 3 Part 2	C03	General nuclear localisation & punctate	LF82_273
Set 3 Part 2	D02	General nuclear localisation	LF82_268
Set 3 Part 2	E09	Punctate	LF82_283
Set 4	E06	Punctate	LF82_394

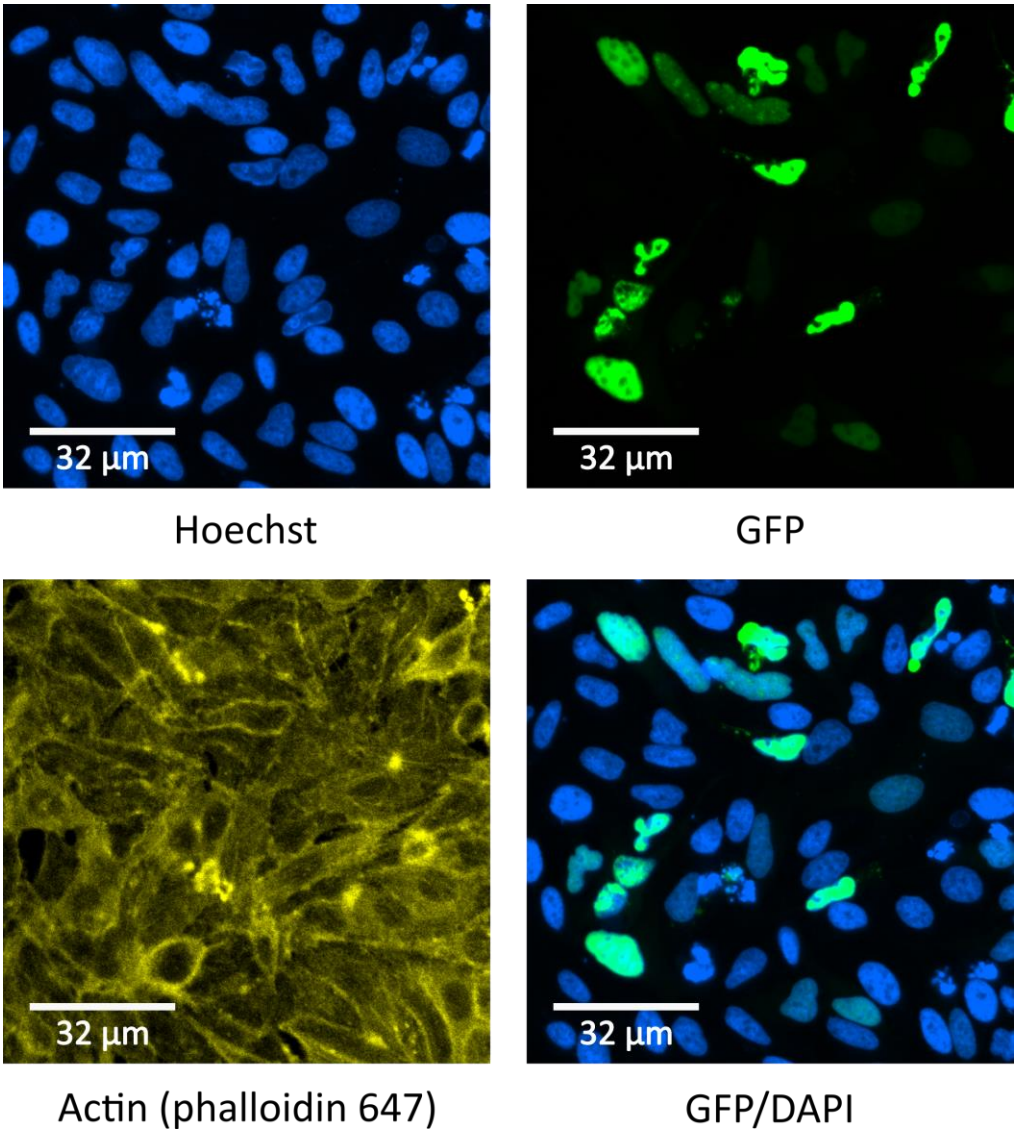


Figure 4.16. **Example well images displaying nuclear localisation**. Well D02 from Set 3 Part 2 Replicate 1 (site 4). GFP can be seen to localise to the nuclei in general.

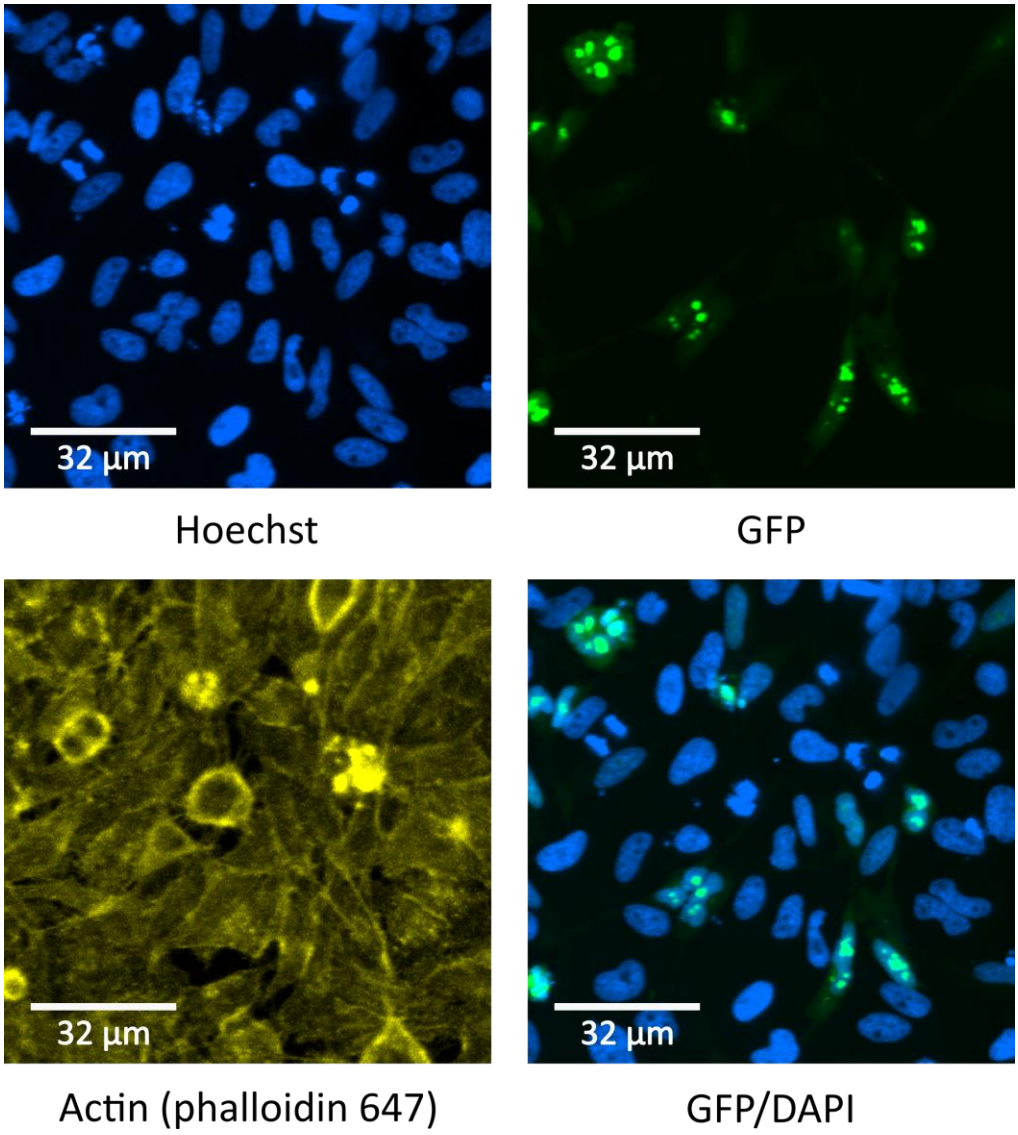


Figure 4.17. **Example well images displaying potential localisation to nucleoli**. Well C05 from Set 3 Part 1 Replicate 1 (site 2). GFP is apparent in foci in the sub-nuclear region, perhaps indicative of localisation to nucleoli.

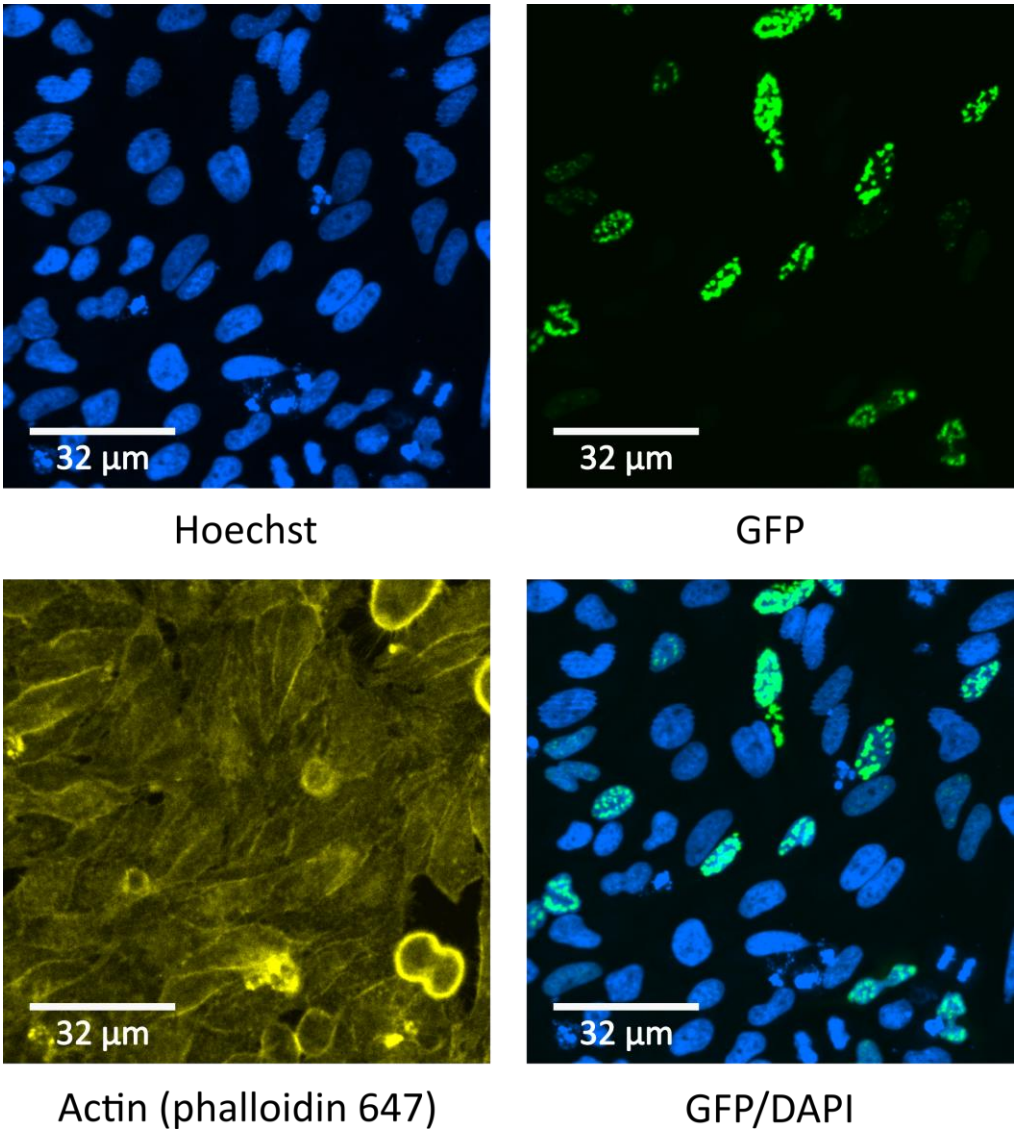
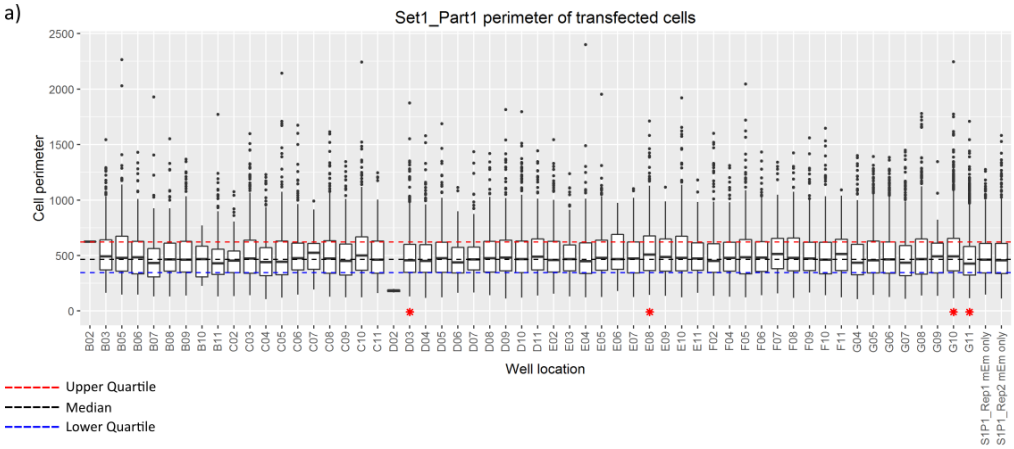


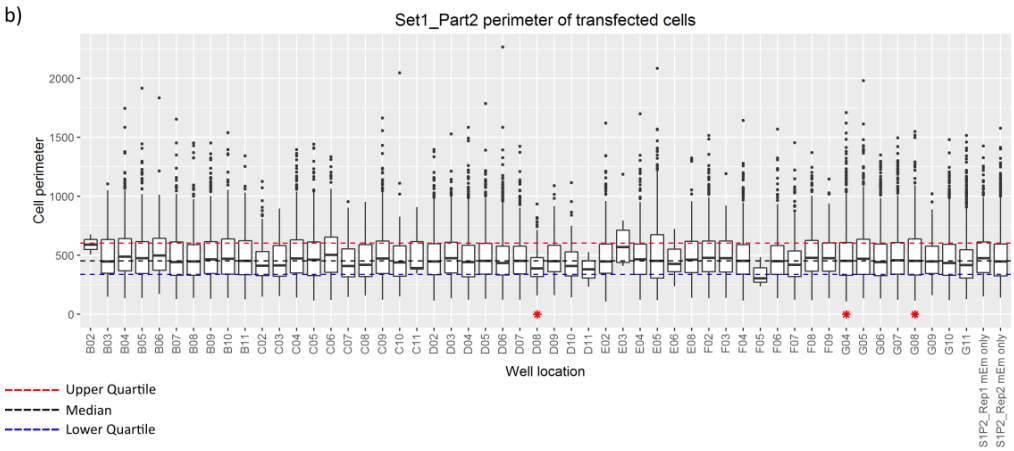
Figure 4.18. **Example well images displaying potential localisation to nuclear dots**. Well E02 from Set 1 Part 2 Replicate 1 (site 3). GFP is apparent in multiple foci in the sub-nuclear region, perhaps indicative of localisation to nuclear dots.

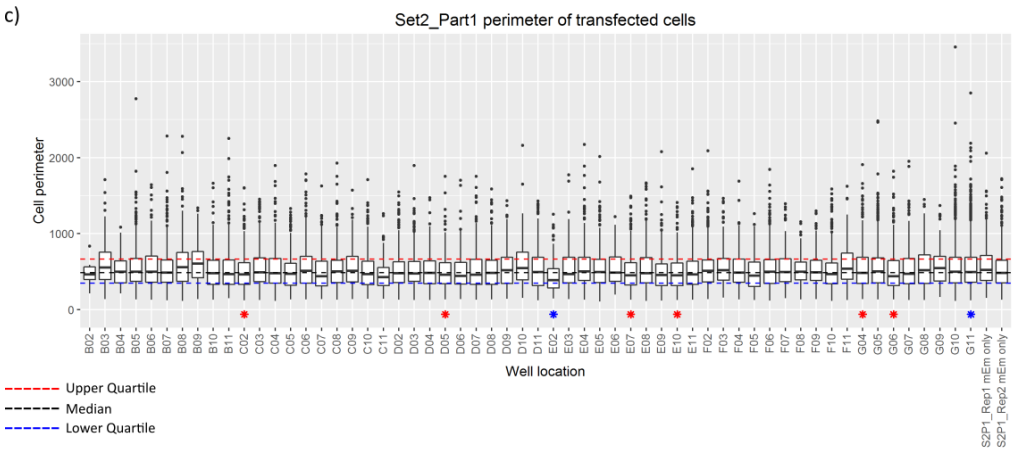
4.5.6 Cell perimeter

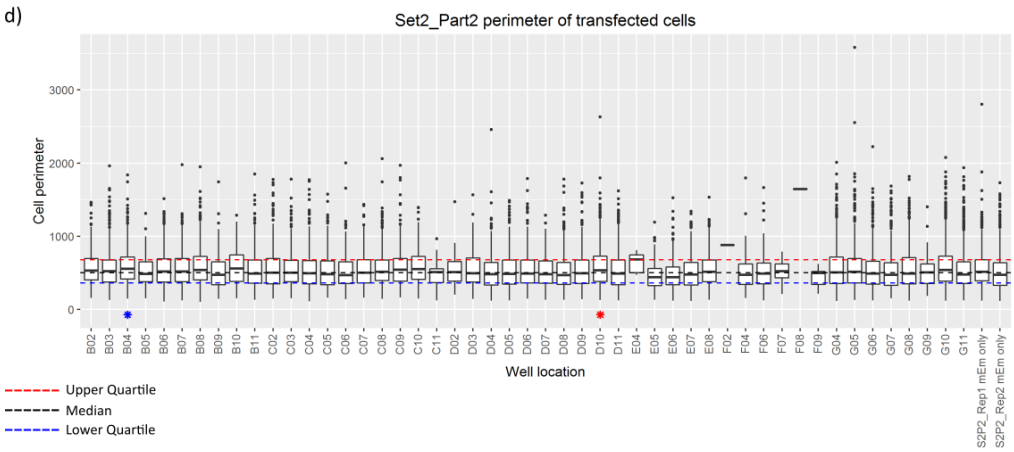
From the cell perimeter graphs shown in Figure 4.19, it was apparent that few wells displayed a statistically significant perimeter change of $P < 0.01$. Those that did were manually checked for a change in cell size/shape. Detection of cell perimeter change was difficult to determine by eye, and no obvious changes in perimeter were seen.

Furthermore, considering graphs for all Sets, wells G10 and G11 – which contained SseD transfected cells, and untreated cells, respectively – show a significant difference in cell perimeter on 3 out of 12 occasions. This finding makes it difficult to trust potential hits identified as significantly different as a quarter of these could be false positives, especially when the phenotype being considered is difficult to identify by eye. This measurement may be more useful in screen where more drastic changes in cell perimeter are expected, such as those looking to detect membrane ruffles, or cell death.


a)
Set1_Part1 perimeter of transfected cells
Cell perimeter
Well location
Upper Quartile
Median
Lower Quartile
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only


b)
Set1_Part2 perimeter of transfected cells
Cell perimeter
Well location
Upper Quartile
Median
Lower Quartile
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only


c)
Set2_Part1 perimeter of transfected cells
Cell perimeter
Well location
Upper Quartile
Median
Lower Quartile
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only


d)
Set2_Part2 perimeter of transfected cells
Cell perimeter
Well location
Upper Quartile
Median
Lower Quartile
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only

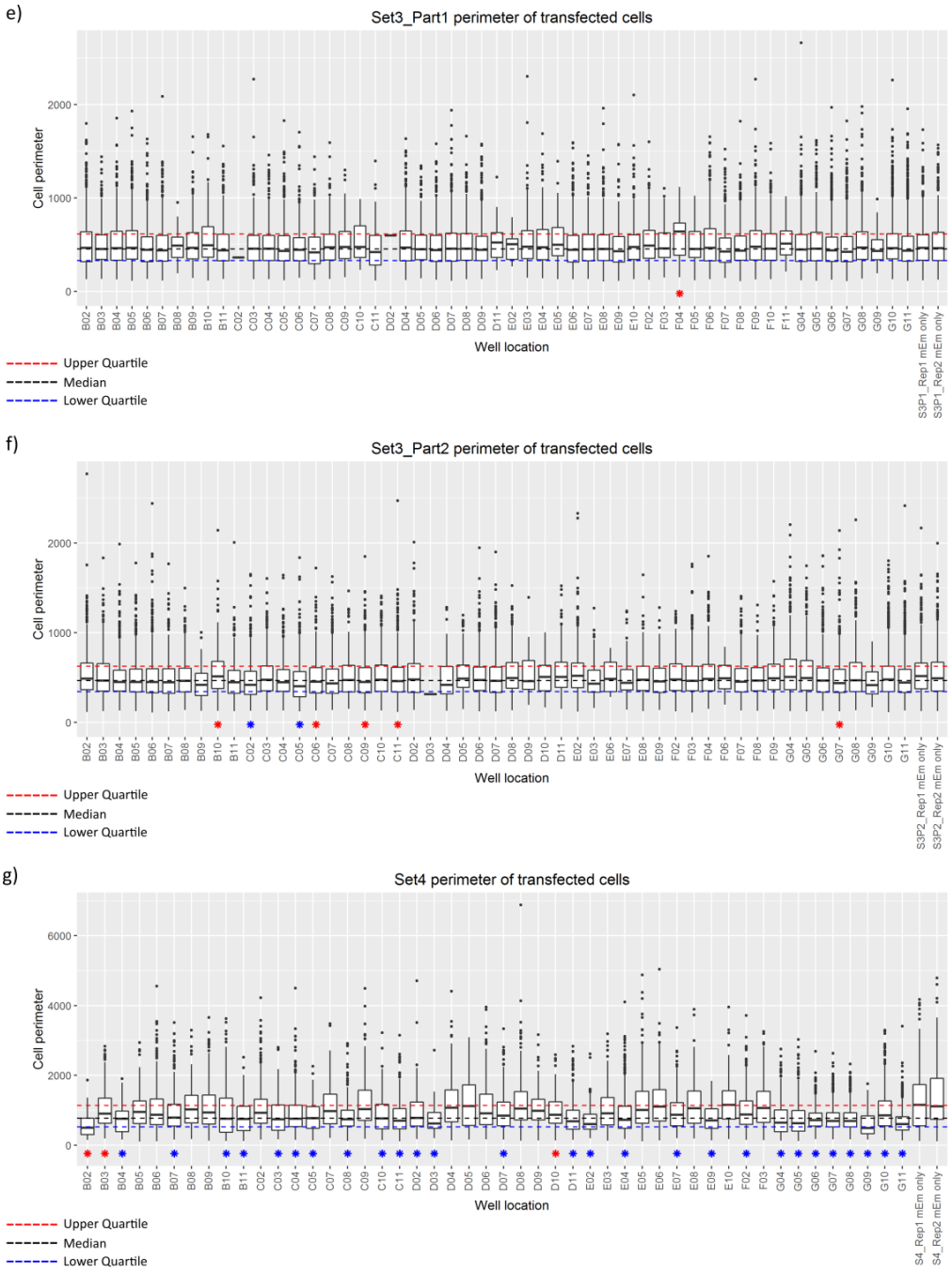


Figure 4.19. **Graphs showing the perimeter of transfected cells**. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.6. **Top 20 hits from transfected cell perimeter data**. Data was normalised between plates using untreated cell data.

Top 20 hits for transfected cell perimeter				
Plate	Well name	Gene	Median	Standard deviation
Set2_Part2	F08	LF82_724	3.16	NA
Set4	E10	LF82_794	1.76	1.00
Set4	D05	LF82_781	1.71	1.10
Set4	E06	LF82_394	1.68	1.03
Set2_Part2	F02	LF82_785	1.65	NA
Set4	D04	LF82_728	1.63	1.12
Set4	F03	LF82_314	1.62	0.97
Set4	E08	LF82_661	1.61	1.02
Set4	D08	LF82_805	1.60	1.20
Set4	C09	LF82_147	1.57	1.05
Set4	B08	LF82_583	1.56	0.87
Set4	E05	LF82_437	1.53	1.14
Set3_Part1	E02	LF82_694	1.53	NA
Set4	D09	LF82_313	1.50	0.86
Set1_Part2	B02	LF82_660	1.49	NA
Set4	C07	LF82_785	1.48	1.00
Set3_Part1	F04	LF82_2826	1.46	0.57
Set4	B05	LF82_246	1.45	0.85
Set4	B09	LF82_662	1.42	0.98
Set4	C02	LF82_788	1.42	0.96

4.5.7 Nuclei perimeter and area

The affinity of the Hoechst 33342 fluorescent stain for DNA resulted in highly specific staining of cell nuclei. This allowed CellProfiler to identify nuclei with high precision. The perimeter and area of nuclei in each well were therefore measured; data for each can be seen in Figures 4.20 and 4.21 (below this section), respectively.

While both perimeter and area are physical measurements of the nucleus, they are both independent measurements and are difficult to compare. For example, the nucleus of a dying cell may experience a high level of blebbing (the occurrence of protrusions that eventually break off), and may also therefore be an odd shape. In this example, the area of the nucleus may be relatively small, whereas the nuclear perimeter may be excessively large due to the many 'folds' present due to blebs in the nuclear envelope. Conversely, there may be occasions where the nucleus is truly more compact, resulting in a smaller perimeter and smaller area.

There are several wells in which nuclei appear to take up less area, as well as having a smaller perimeter. An example of this is well E07 in Set 2 Part 2 (Figure 4.20 d. and 4.21 d.) Images from this well can be seen in Figure 4.22, where transfected cell nuclei can be identified by their associated GFP content. While it is difficult to conclusively state that an effect upon nuclear perimeter is present, nuclei do appear to be physically smaller than their surrounding nuclei and/or an odd shape in some cases. Interestingly, the CDS expressed in this well (LF82_675) also appears to be excluded from the nuclear region of transfected cells, although it is unclear how this would affect nuclear size or shape. To confirm a phenotype of this nature, more repeats would be necessary, and LF82_675 may require further investigation to determine its true associated phenotype.

As shown here, this kind of phenotype is difficult to observe, even with a quantitative screen to aid analysis. It is therefore reasonable to assume that the phenotype seen in this case, if real, would likely be overlooked if screen images were processed without the aid of CellProfiler (or similar software).

Although difficult to observe via the graphs in Figures 4.20 and 4.21, the control well in which p62 was expressed often appears to display a phenotype regarding the perimeter and/or area of transfected cell nuclei (see Figure 4.4. As p62 is known to be associated with nuclei, and has a role in shuttling cargo between the nucleus and cytoplasm, this area/perimeter effect may not be surprising (Lippai & Lőw 2014). Excess p62 may, for example, result in an increased level of nuclear-cytoplasmic shuttling, reducing or increasing the size of the nucleus. In comparison, SseD and untransfected cell controls do not appear to affect nuclei in either Figures 4.3, 4.20, or 4.21.

4.5.7.1 Nuclei perimeter

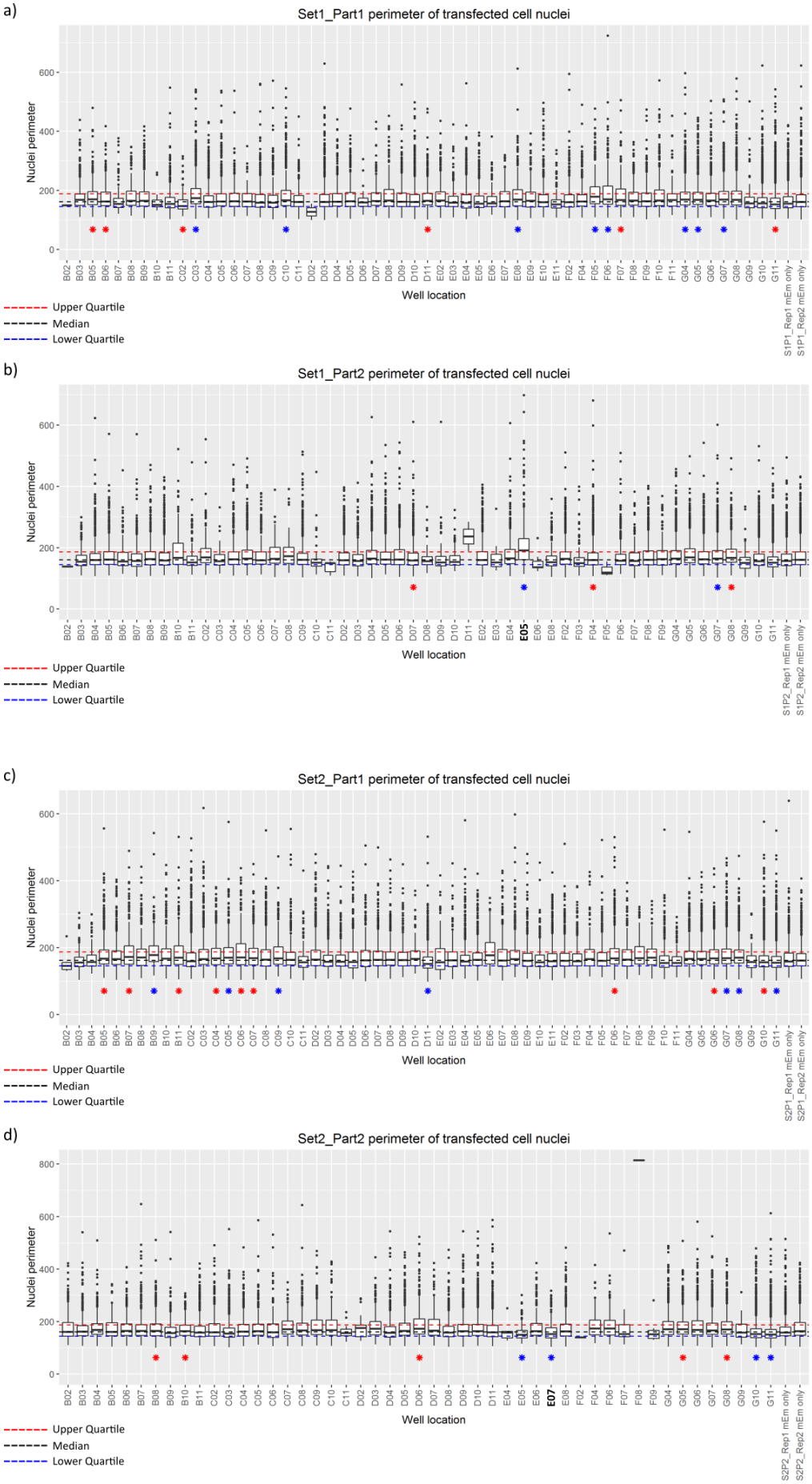
While it is clear that the parameters of one nuclear measurement may change while the other remains relatively consistent, this observation offers little evidence for functional characterisation of a particular CDS. Nonetheless, the detection of these parameters appeared to be successful for the identification of hits from this screen.

Well E05 from Set 1 Part 2 (Figure 4.20 b.) was identified from the nuclear perimeter measurement graphs, and did not appear to present a significant result on the nuclei area graph (Figures 4.20 and 4.21, respectively). As can be seen in Figure 4.23, transfected cells from this well frequently appear to have oddly shaped nuclei in comparison to the more typical oval shaped nuclei seen in non-transfected cells. It remains to be determined what effect the expression of the respective CDS, LF82_364, has upon these cells, though this phenotype may suggest that this CDS has cytotoxic effects.

4.5.7.2 Nuclei area

Many hits were identified using the nuclei area measurement, however they were difficult to confirm via manual analysis of well images.

It should be noted that some wells, for example well F05 from Set 1 Part 2 in Figure 4.21 b., appear to show a large shift in nuclei area, but lack significance. After checking the data to support this result, it was determined that the average number of transfected cells per site identified by CellProfiler in these cases were very low (3 cells in this case), hence the lack of significance.


a)
Set1_Part1 perimeter of transfected cell nuclei
b)
Set1_Part2 perimeter of transfected cell nuclei
c)
Set2_Part1 perimeter of transfected cell nuclei
d)
Set2_Part2 perimeter of transfected cell nuclei
Nuclei perimeter
Well location
Upper Quartile
Median
Lower Quartile
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only

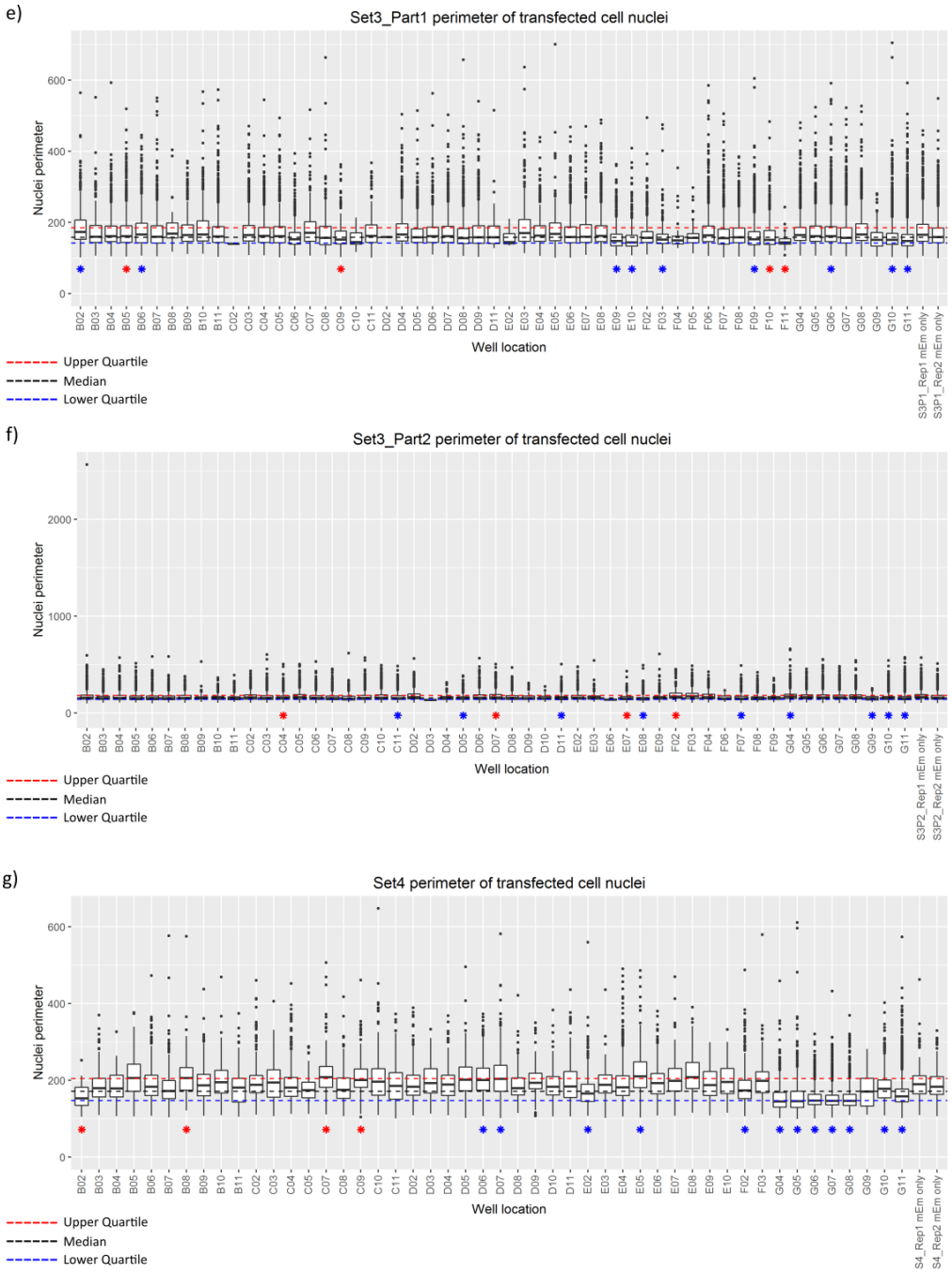
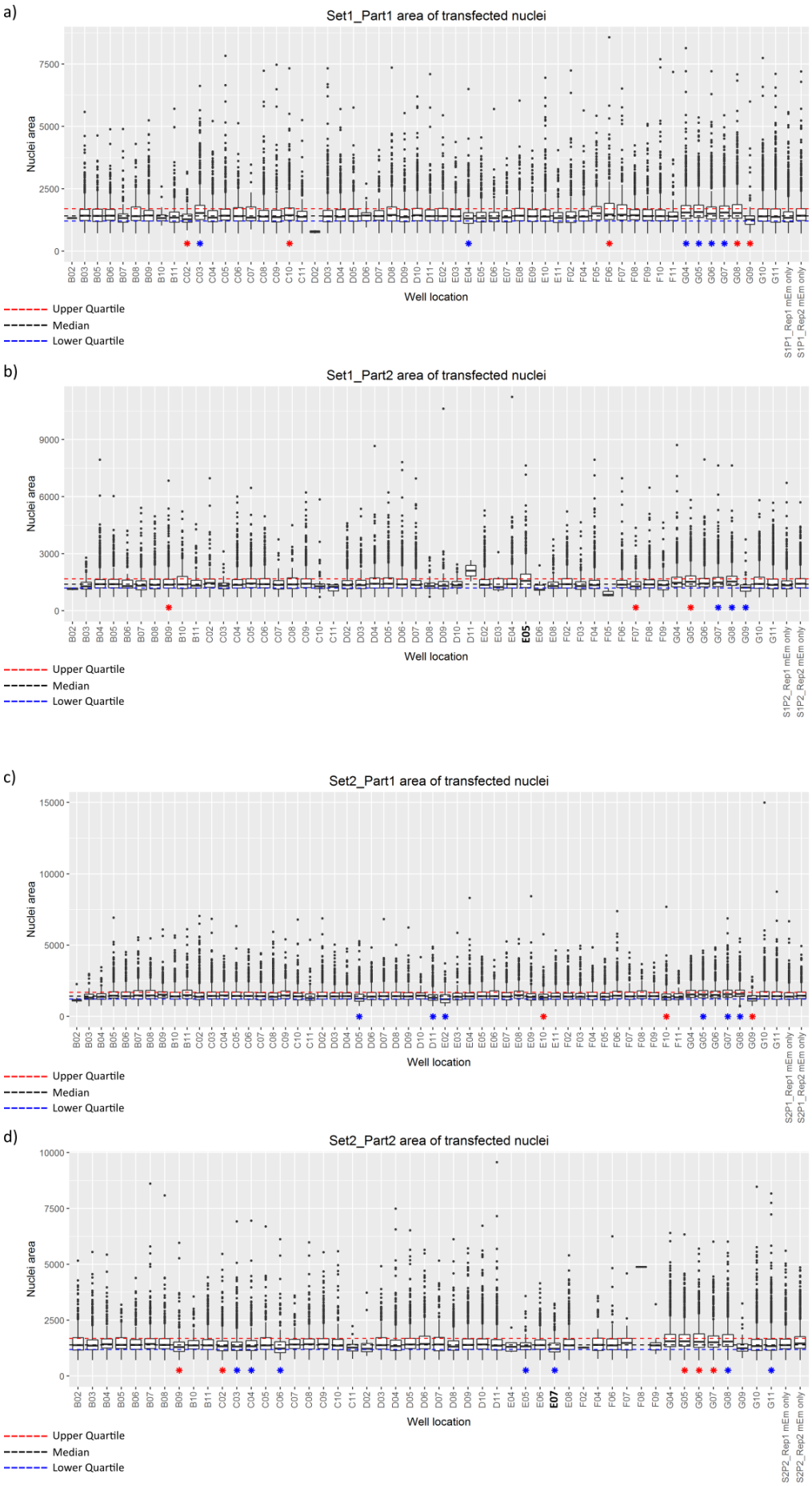


Figure 4.20. **Graphs showing the perimeter of nuclei within transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01. Graph f) can be seen to utilise a larger range on the y-axis due to the presence of an outlier in well location B02.

Table 4.7. **Top 20 hits from transfected nuclei perimeter data**. Data was normalised between plates using untreated cell data.

Top 20 hits for transfected nuclei perimeter				
Plate	Well name	Gene	Median	Standard deviation
Set2_Part2	F08	LF82_724	5.07	NA
Set4	E05	LF82_437	1.26	0.35
Set4	C07	LF82_785	1.25	0.36
Set4	E08	LF82_661	1.25	0.32
Set2_Part2	E04	LF82_261	1.24	0.40
Set4	B05	LF82_246	1.23	0.32
Set4	B08	LF82_583	1.23	0.30
Set4	D07	LF82_797	1.22	0.33
Set4	D05	LF82_781	1.21	0.34
Set4	C09	LF82_147	1.20	0.28
Set4	D06	LF82_790	1.20	0.29
Set4	E07	LF82_649	1.19	0.31
Set4	F03	LF82_314	1.19	0.30
Set2_Part1	E06	LF82_602	1.18	0.30
Set4	C10	LF82_233	1.17	0.38
Set4	E10	LF82_794	1.17	0.28
Set4	B10	LF82_689	1.16	0.28
Set4	C03	LF82_789	1.16	0.30
Set4	D09	LF82_313	1.16	0.27
Set4	E06	LF82_394	1.15	0.27


a)
Set1_Part1 area of transfected nuclei
Nuclei area
Well location
Upper Quartile
Median
Lower Quartile
b)
Set1_Part2 area of transfected nuclei
Nuclei area
Well location
Upper Quartile
Median
Lower Quartile
c)
Set2_Part1 area of transfected nuclei
Nuclei area
Well location
Upper Quartile
Median
Lower Quartile
d)
Set2_Part2 area of transfected nuclei
Nuclei area
Well location
Upper Quartile
Median
Lower Quartile

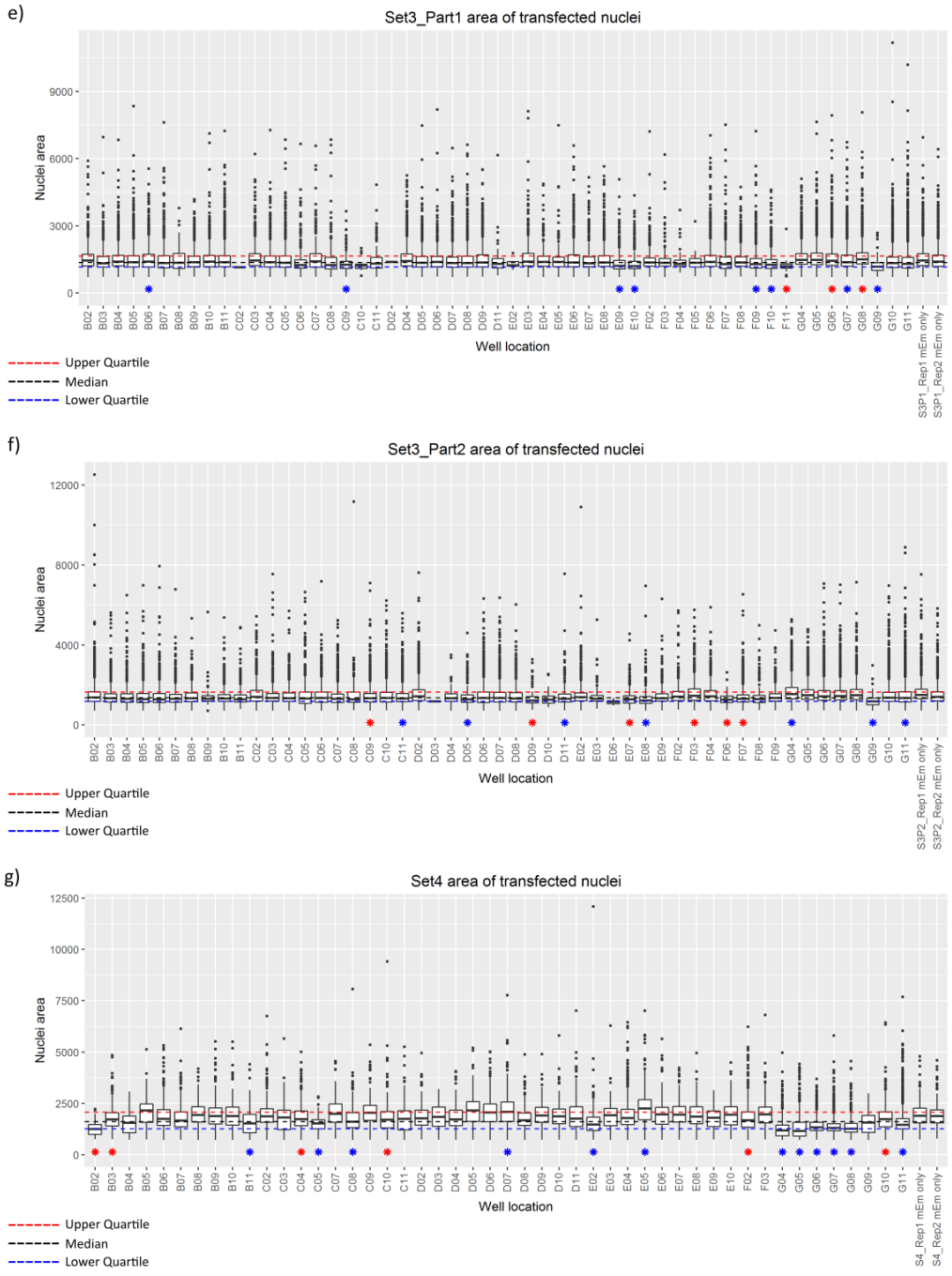


Figure 4.21. **Graphs showing the area of nuclei within transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.8. **Top 20 hits from transfected nuclei area data**. Data was normalised between plates using untreated cell data.

Top 20 hits for transfected nuclei area				
Plate	Well name	Gene	Median	Standard deviation
Set2_Part2	F08	LF82_724	3.34	NA
Set4	E05	LF82_437	1.40	0.57
Set4	D05	LF82_781	1.34	0.50
Set4	B05	LF82_246	1.34	0.45
Set4	D07	LF82_797	1.30	0.51
Set4	D06	LF82_790	1.28	0.46
Set4	C09	LF82_147	1.27	0.45
Set4	C07	LF82_785	1.24	0.46
Set4	E06	LF82_394	1.22	0.45
Set4	F03	LF82_314	1.22	0.44
Set4	E07	LF82_649	1.21	0.41
Set4	E10	LF82_794	1.21	0.41
Set4	B08	LF82_583	1.20	0.40
Set4	E03	LF82_774	1.19	0.41
Set4	D09	LF82_313	1.19	0.41
Set4	B09	LF82_662	1.17	0.45
Set4	B10	LF82_689	1.16	0.47
Set4	D10	LF82_698	1.16	0.43
Set4	C02	LF82_788	1.16	0.44
Set4	E08	LF82_661	1.15	0.42

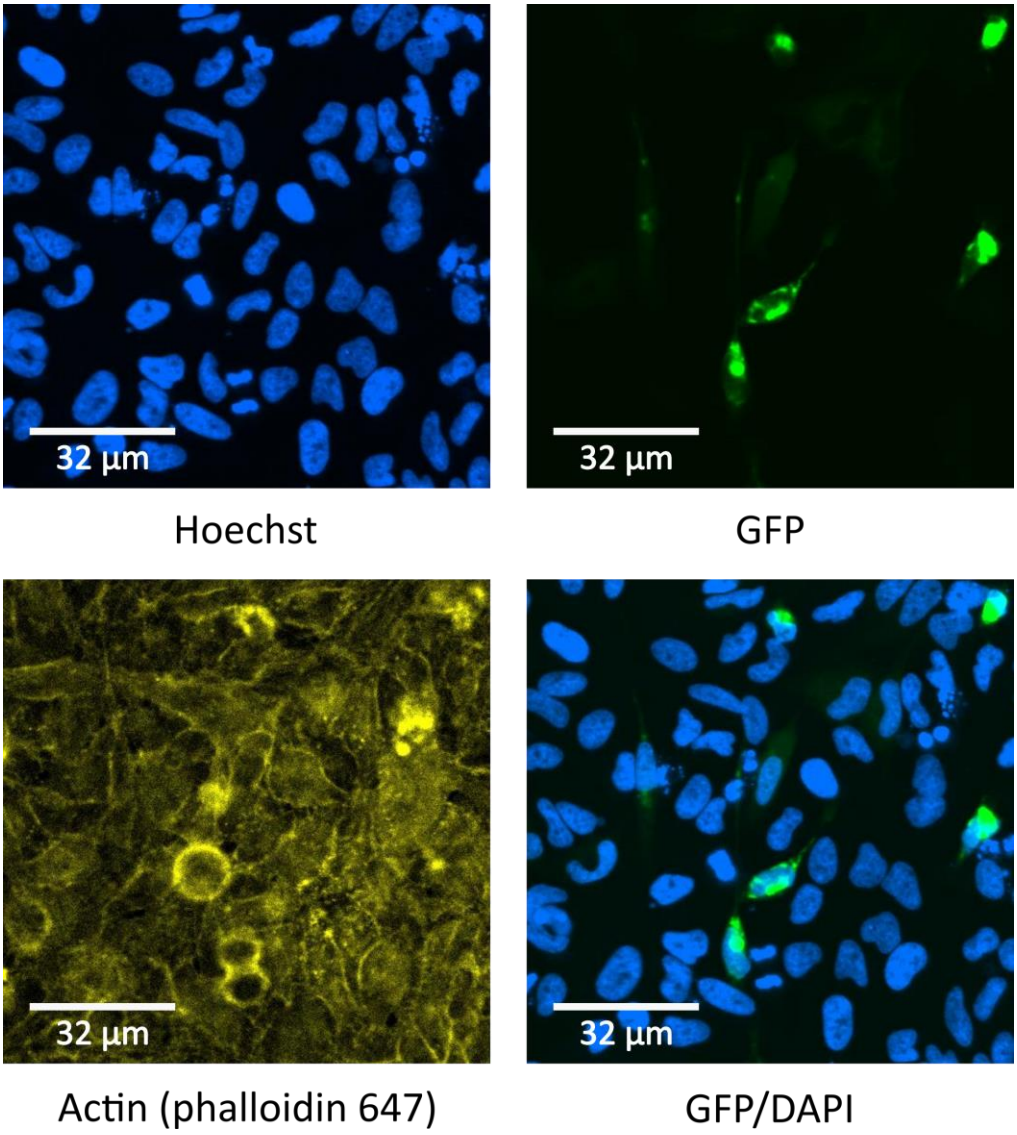


Figure 4.22. **Example well images displaying a potential effect upon nuclei size**. Well E07 from Set 2 Part 2 (site 5). Nuclei perimeter and nuclei area measurements were used to identify this well. Manual image analysis suggested that transfected nuclei may be smaller than untransfected cell nuclei.

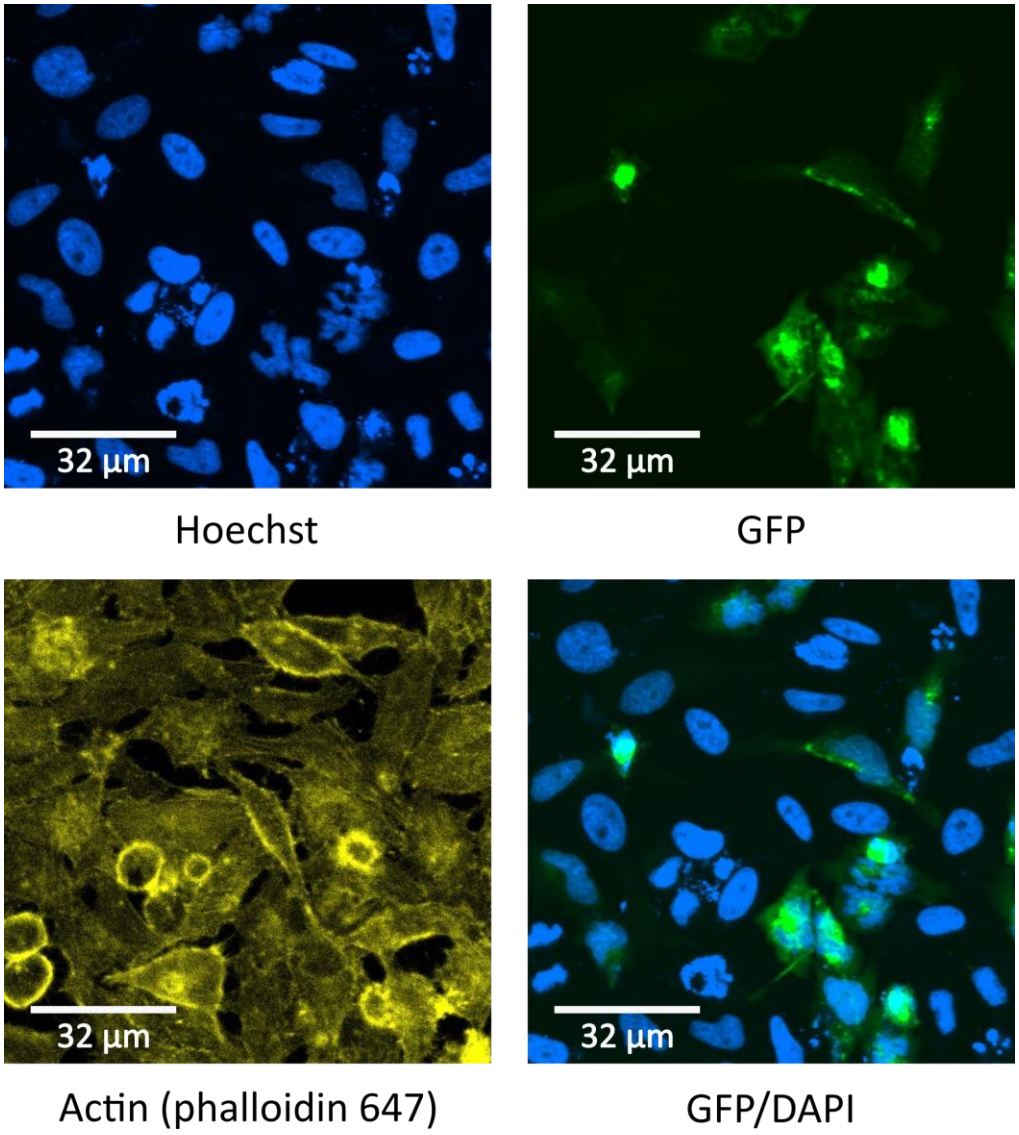


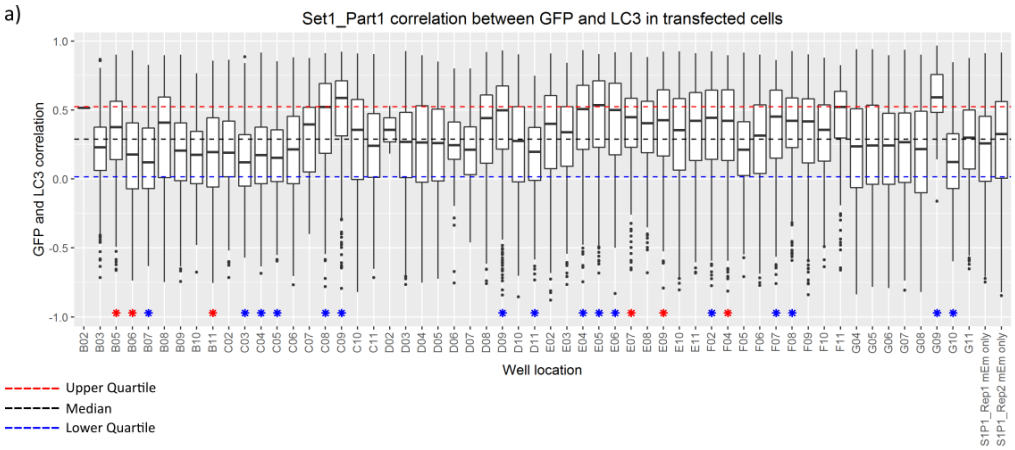
Figure 4.23. **Example well images displaying irregularly shaped nuclei**. Well E05 from Set 1 Part 2 (site 1). The nuclei perimeter measurement was used to identify this well. Manual image examination suggested that transfected nuclei displayed larger perimeters due to nuclear shape irregularity.

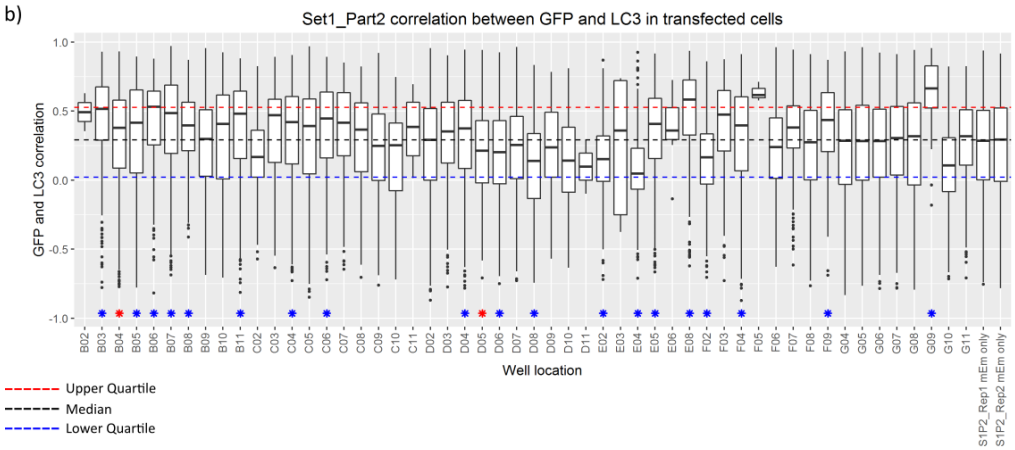
4.5.8 GFP/LC3 and GFP/Actin

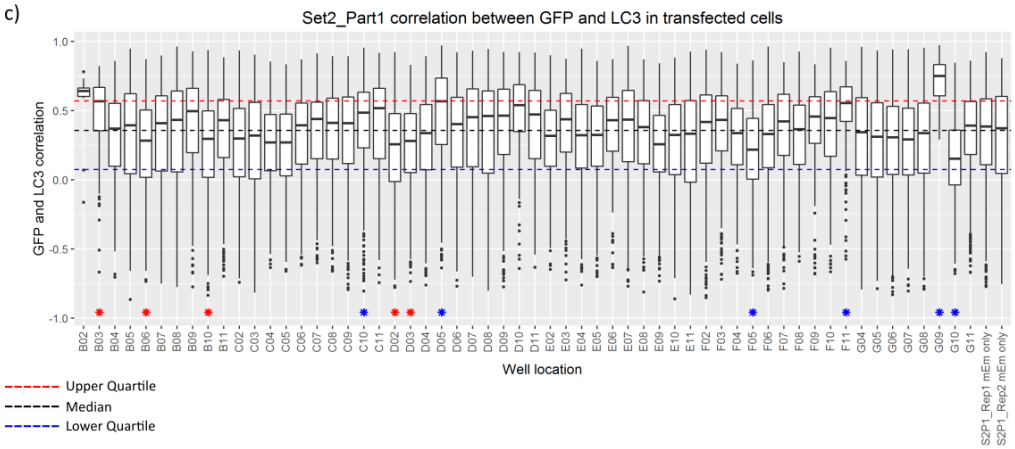
Correlation between GFP and LC3 was determined through the mCherry channel, while correlation between GFP and actin was determined via the far-red channel. Graphs for GFP/LC3 and GFP/Actin correlation can be seen in Figures 4.24 and 4.25, respectively.

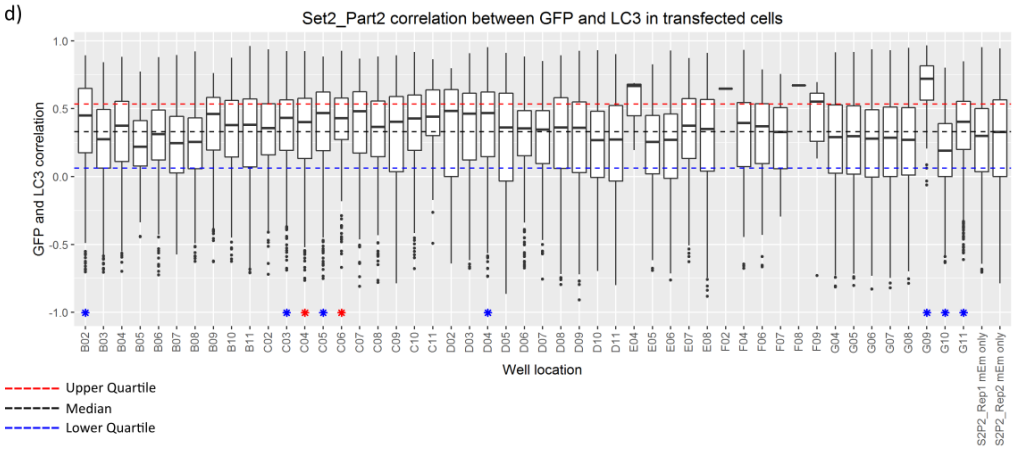
GFP/LC3 correlation was difficult to evaluate due to difficulties in identifying LC3-labelled autophagosomes (see: Limitations in the identification of autophagosomes). However, a consistently high correlation was seen in well G09 across all plates, due to overexpression of p62 (Figure 4.7, shown earlier). This suggests that, when autophagy is induced via the overexpression of p62, the GFP-labelled p62 co-localises to autophagosomes along with mCherry-LC3, and can be detected doing so. This makes the over expression of p62 a useful control for the detection of GFP localisation to autophagosomes. Despite this, there were no significant hits which showed boxplots above the upper quartile threshold in a similar way to that of p62. This suggest that either no proteins from the LF82 unique gene library resulted in an increase in autophagosome number to a similar degree as p62 (if at all), or the difficulties in identifying autophagosomes through CellProfiler made GFP/LC3 correlation difficult to determine.

Resulting graphs from GFP/Actin correlation initially looked promising, however the number of wells displaying significant difference from the control wells was suspicious. It appears that this correlation measurement struggled to detect true GFP/Actin correlation from GFP presence in the cell cytoplasm. An example of this can be seen in Figure 4.26, in which GFP localisation to the cytoplasm can be seen, with exclusion from the nuclear region. Higher resolution microscopy, along with a positive control for actin localisation, would help to resolve this issue in the future.


a)
Set1_Part1 correlation between GFP and LC3 in transfected cells
GFP and LC3 correlation
Well location
Upper Quartile
Median
Lower Quartile


b)
Set1_Part2 correlation between GFP and LC3 in transfected cells
GFP and LC3 correlation
Well location
Upper Quartile
Median
Lower Quartile


c)
Set2_Part1 correlation between GFP and LC3 in transfected cells
GFP and LC3 correlation
Well location
Upper Quartile
Median
Lower Quartile


d)
Set2_Part2 correlation between GFP and LC3 in transfected cells
GFP and LC3 correlation
Well location
Upper Quartile
Median
Lower Quartile

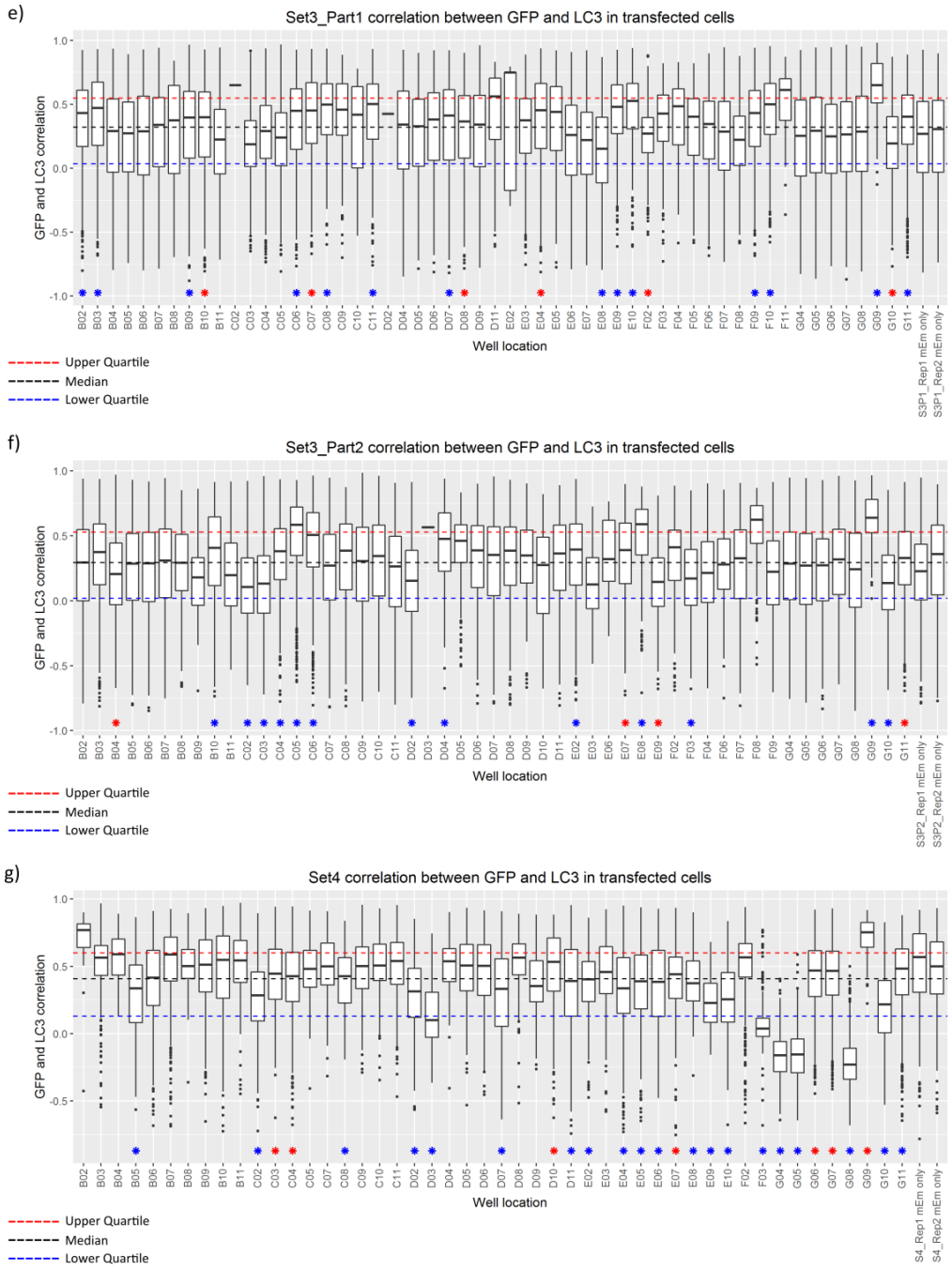
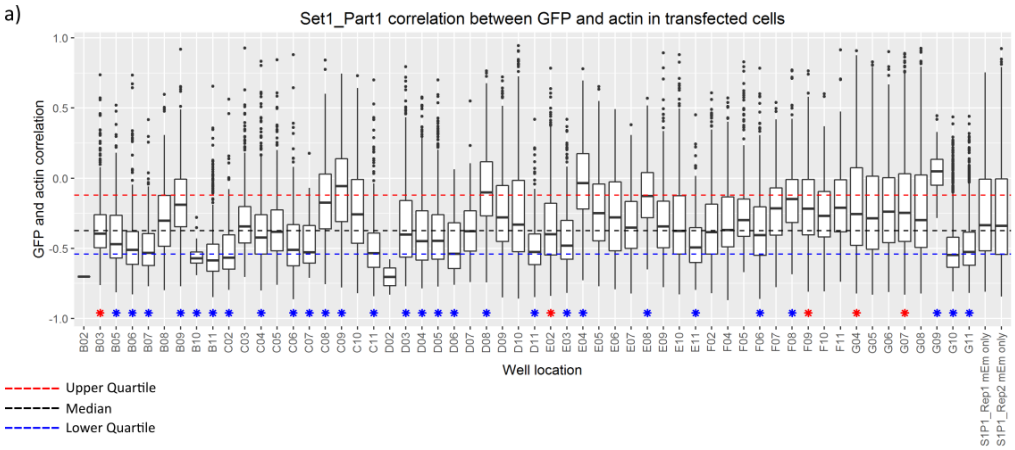
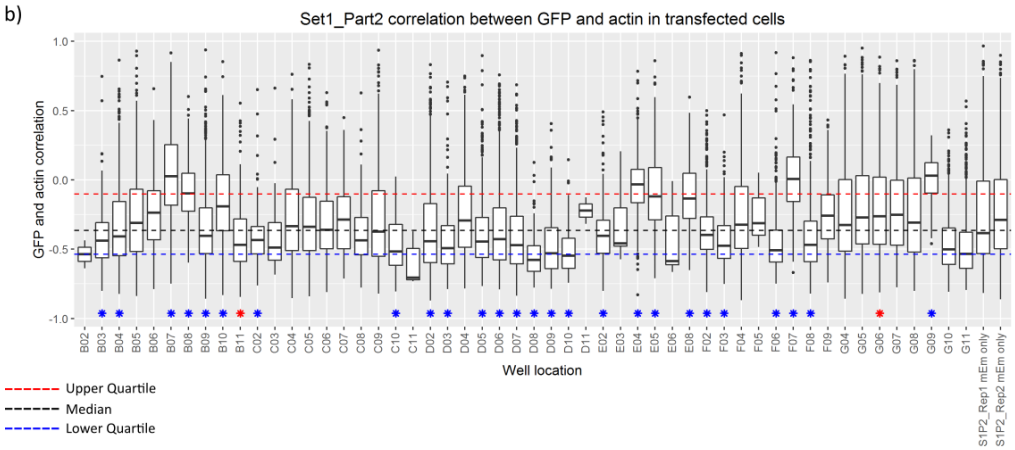


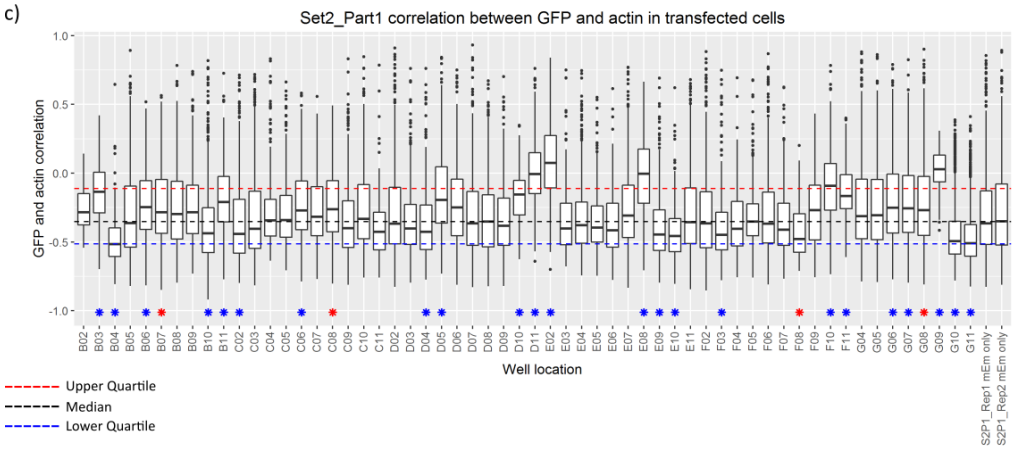
Figure 4.24. **Graphs showing the correlation between GFP and LC3 in transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

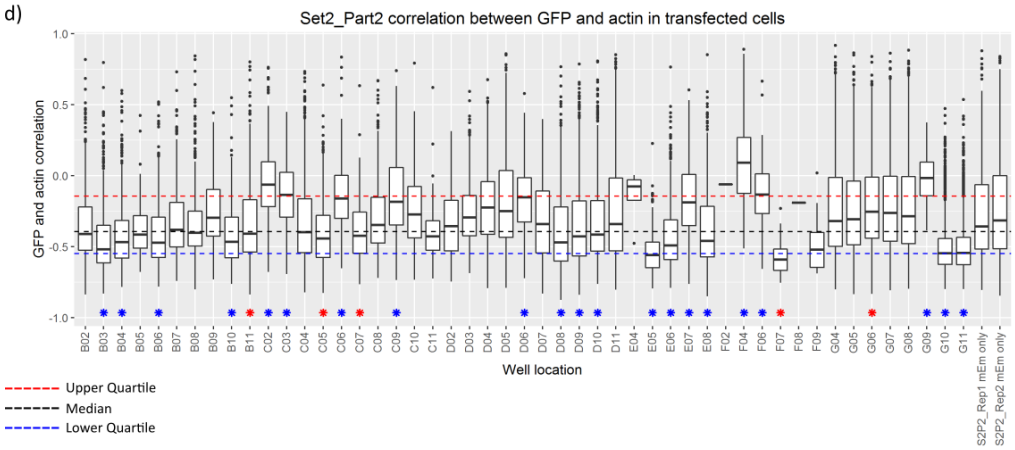
Table 4.9. **Top 20 hits from GFP and LC3 correlation data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest.

Top 20 hits for correlation between GFP and LC3 in transfected cells				
Plate	Well name	Gene	Median	Standard deviation
Set3_Part2_Rep1	*G09*	*p62*	*2.61*	*0.80*
Set1_Part1	C09	LF82_153	2.93	1.15
Set1_Part1	C08	LF82_152	2.75	1.51
Set3_Part2	F08	LF82_266	2.63	1.03
Set1_Part1	E05	LF82_326	2.57	1.44
Set3_Part2	C05	LF82_758	2.53	1.26
Set3_Part2	E08	LF82_121	2.50	1.12
Set1_Part1	E06	LF82_327	2.46	1.83
Set1_Part1	D09	LF82_250	2.44	1.61
Set1_Part1	E04	LF82_325	2.40	1.49
Set1_Part1	F11	LF82_603	2.27	1.47
Set1_Part1	B02	LF82_004	2.24	NA
Set1_Part1	B08	LF82_091	2.19	1.54
Set1_Part2	B02	LF82_660	2.18	NA
Set1_Part1	D08	LF82_248	2.13	1.47
Set3_Part1	E02	LF82_694	2.08	NA
Set1_Part1	E09	LF82_355	2.07	1.54
Set1_Part1	F07	LF82_537	2.05	1.50
Set3_Part2	C06	LF82_109	2.04	1.39
Set1_Part1	E07	LF82_350	2.01	1.34
Set1_Part2	F05	LF82_415	2.00	NA


a)
Set1_Part1 correlation between GFP and actin in transfected cells
GFP and actin correlation
Well location
Upper Quartile
Median
Lower Quartile
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only


b)
Set1_Part2 correlation between GFP and actin in transfected cells
GFP and actin correlation
Well location
Upper Quartile
Median
Lower Quartile
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only


c)
Set2_Part1 correlation between GFP and actin in transfected cells
GFP and actin correlation
Well location
Upper Quartile
Median
Lower Quartile
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only


d)
Set2_Part2 correlation between GFP and actin in transfected cells
GFP and actin correlation
Well location
Upper Quartile
Median
Lower Quartile
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only

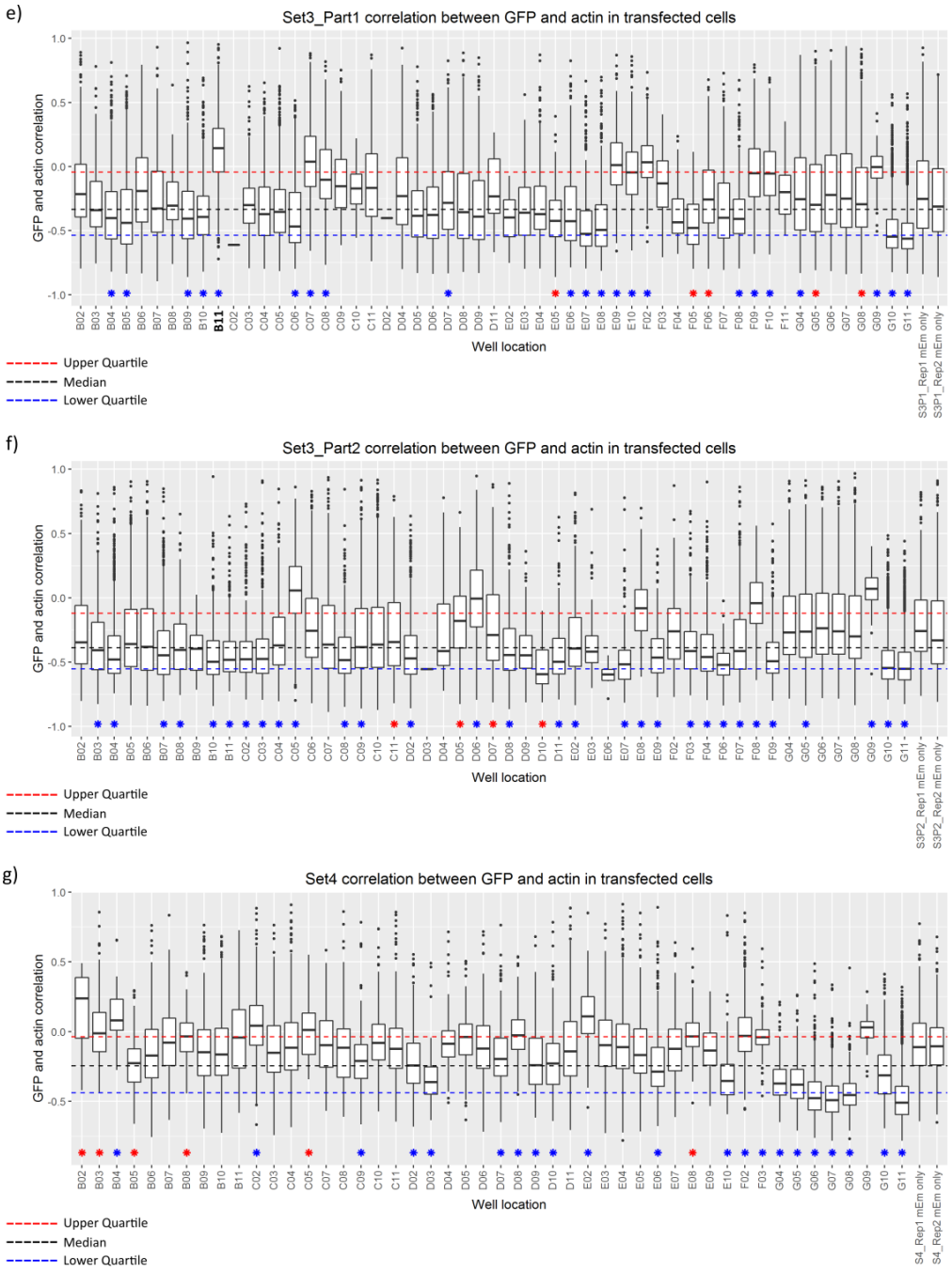


Figure 4.25. **Graphs showing the correlation between GFP and actin in transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.10. **Top 20 hits from GFP and LC3 correlation data**. Data was normalised between plates using untreated cell data.

Top 20 hits for correlation between GFP and actin in transfected cells				
Plate	Well name	Gene	Median	Standard deviation
Set1_Part1	B02	LF82_004	1.47	NA
Set1_Part1	D02	LF82_230	1.46	0.37
Set2_Part2	F09	LF82_797	1.33	0.27
Set1_Part2	B02	LF82_660	1.33	NA
Set1_Part1	D06	LF82_246	1.30	0.49
Set3_Part2	E06	LF82_3578	1.28	0.13
Set3_Part2	D10	LF82_139	1.28	0.40
Set2_Part2	F07	LF82_676	1.27	0.20
Set1_Part1	B11	LF82_101	1.26	0.47
Set2_Part2	E05	LF82_378	1.23	0.35
Set2_Part1	F08	LF82_534	1.19	0.44
Set1_Part1	C02	LF82_105	1.19	0.43
Set1_Part1	B10	LF82_093	1.19	0.13
Set1_Part1	D11	LF82_311	1.18	0.39
Set1_Part1	C11	LF82_217	1.16	0.52
Set3_Part2	E07	LF82_410	1.15	0.54
Set3_Part2	F06	LF82_0907	1.14	0.32
Set1_Part2	C11	LF82_759	1.14	0.52
Set1_Part1	B07	LF82_089	1.13	0.44
Set3_Part1	C02	LF82_146	1.12	NA

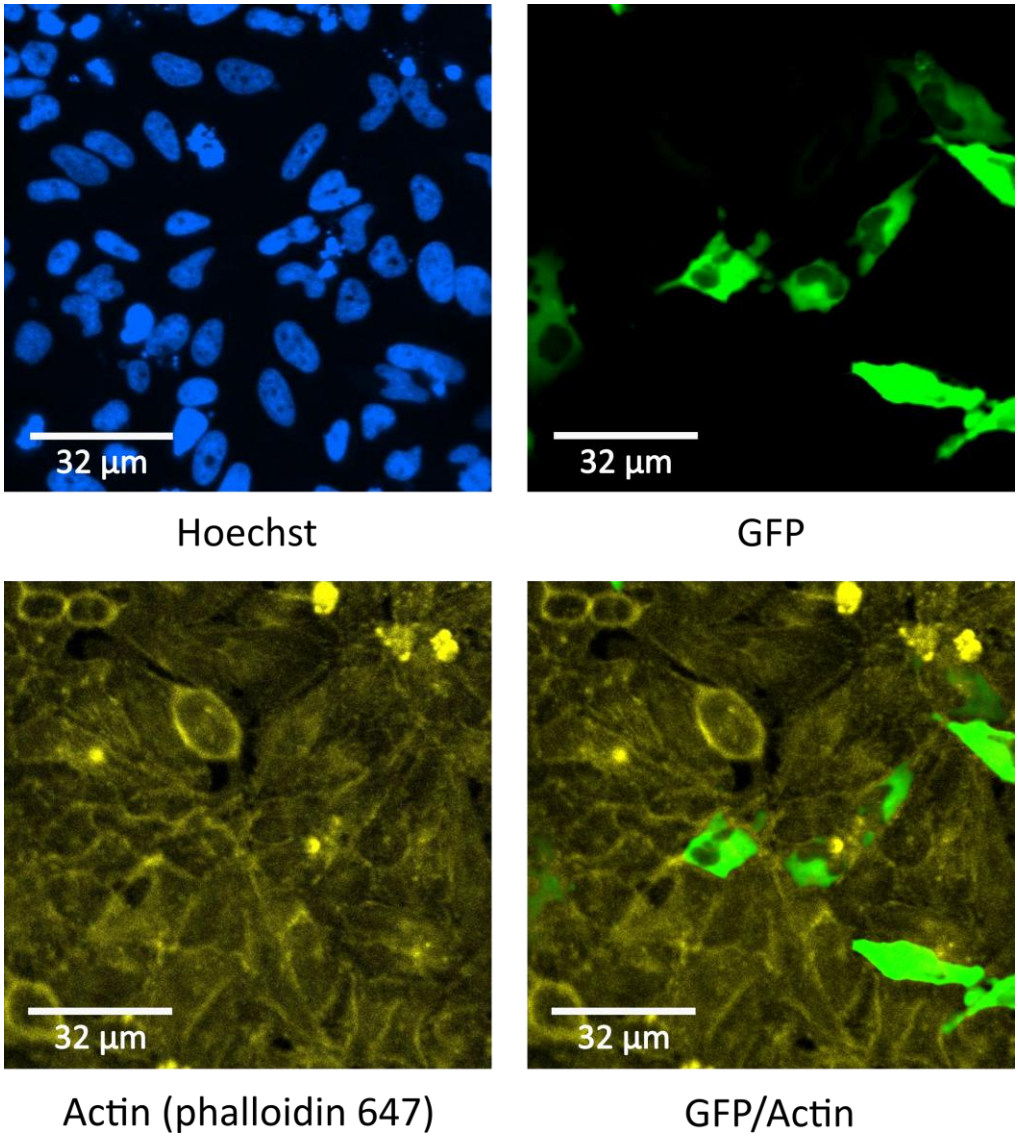


Figure 4.26. **Example well images displaying cytoplasmic localisation**. Well B11 of Set 3 Part 1 Replicate 1 (site 4). GFP/Actin correlation measurements were used to identify this well, suggesting localisation to actin. However, manual image analysis did not support this, with GFP likely localising to the cytoplasm.

4.5.9 Radial zone 4

The radial zone measurement divides any transfected cell into 4 concentric zones, within which the fraction of GFP present is determined (see: CellProfiler pipeline). This measurement was perhaps the most useful used in this screen, and resulted in the detection of several interesting phenotypes, with one in particular discussed later.

The measurement of the outer zone in the radial zone module was used to identify cells in which GFP localised to the cell cytoplasm and/or outer membrane. Graphs shown in Figure 4.27 were therefore used to identify cytoplasmic localisation of proteins, however no apparent cell membrane localisation was observed.

The radial zone 4 data was also supported by radial zone 1 data, as would be expected, with nuclear localisation on radial zone 1 graphs anti-correlating with cytoplasmic localisation on radial zone 4 graphs. An example of this can be seen with Set 3 Part 1 well B11 (Figure 4.26, above). Boxplots for this well appear as significant hits on both the radial zone 1 graph (Figure 4.11 e.) and the radial zone 4 graph (Figure 4.27 e.), with boxes appearing below and above their respective quartile thresholds. Well G09 served as a control for radial zone 4 data, with p62 known to be excluded from the cell nucleus and to show a preference for outer cytoplasmic localisation. This can be seen across all radial zone 4 graphs, with the boxplot for well G09 consistently appearing above the plate-wide upper quartile (see Figure 4.27).

For each boxplot that was significantly different from controls, and exceeded the upper quartile line on each graph, respective well images were manually analysed for confirmation of each phenotype. The results from these observations can be found in Table 4.11. Of particular interest was the well containing LF82_314 (highlighted in bold in Table 4.11), discussed in Chapter 5.

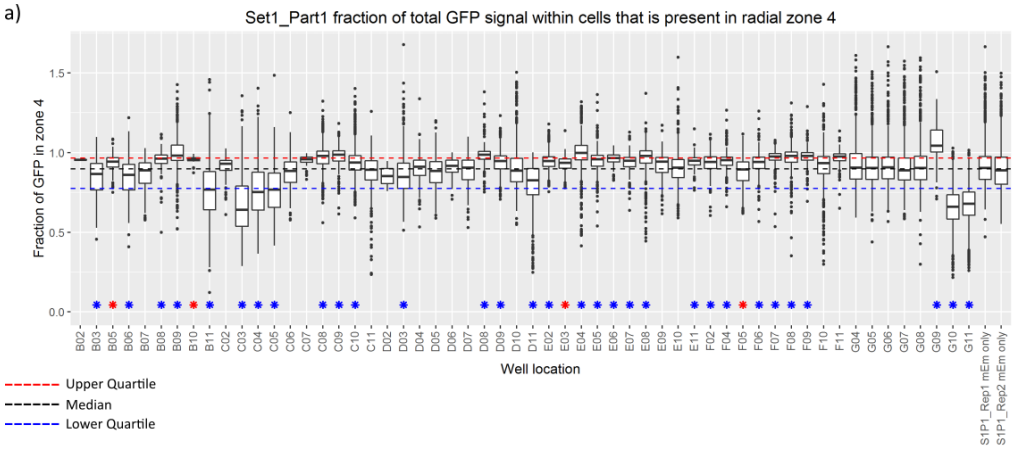
As shown in Table 4.11, well E10 from Set 3 Part 1 (designated by an asterisk) appears to have erroneously resulted in a successful transformation due to the presence of an unexpected transformation vector. This is evidenced by the presence of data for well E10 on graphs, as any wells lacking transformation vector should not contain GFP signal, should not be considered by the CellProfiler pipeline, and therefore should not result in any data output. After sequencing the contaminating vector, it was shown to contain the LF82_772 CDS of a nearby well (well F09). This contamination was likely due to a combination of human error and the necessary alteration of sample layout for microscopy. However, no obvious evidence of additional contamination was observed. Moreover, any further contamination would be detectable during downstream confirmation of phenotype and sequencing testing.

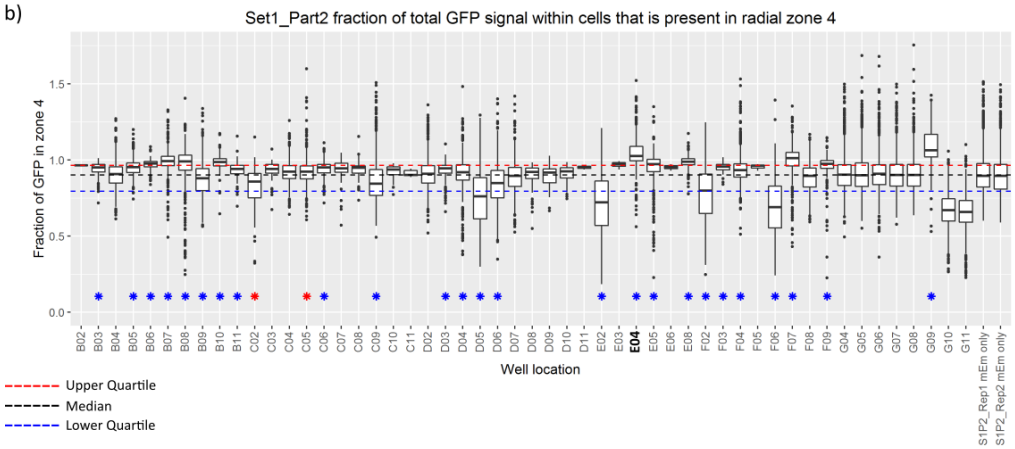
Table 4.11. **Analysis of screen hits using radial zone 1 and 4 graphs**. GFP localisation was determined for each potential hit by manually checking images acquired from each respective well. Hits were selected based upon presence of blue asterisks (representing a p value of < 0.01) on both measurement graphs, and also by comparisons between boxplot boxes and quartile lines. Highlighted in bold is LF82_314, discussed in Chapter 5. * suspected cross contamination of LF82_772 from well F09.

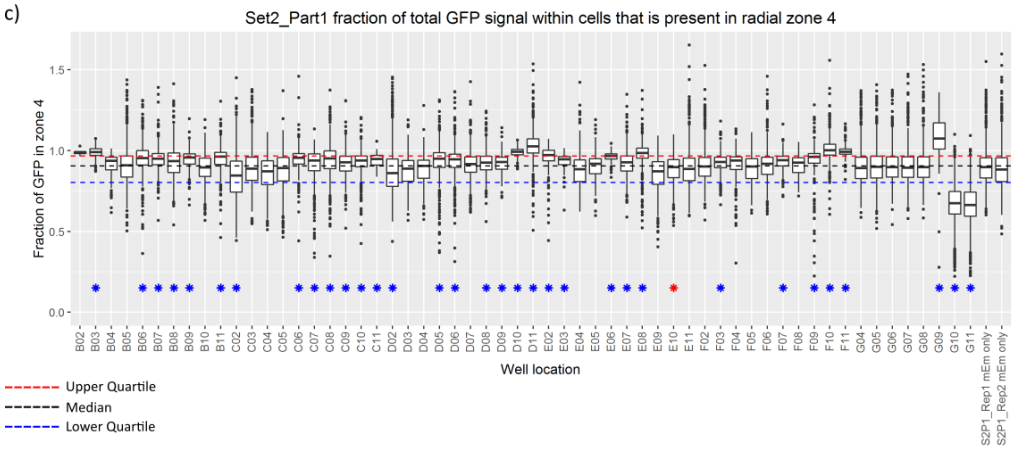
Plate	Well	GFP localisation	Respective LF82 CDS
Set 1 Part 1	B08	Multiple small foci in cytoplasm	LF82_092
Set 1 Part 1	B09	Cytoplasmic	LF82_092
Set 1 Part 1	C08	Cytoplasmic, but variable	LF82_152
Set 1 Part 1	C09	Cytoplasmic – some nuclear	LF82_153
Set 1 Part 1	D08	Cytoplasmic – some nuclear	LF82_248
Set 1 Part 1	E04	Cytoplasmic	LF82_325
Set 1 Part 1	E08	Cytoplasmic – nuclear associated	LF82_353
Set 1 Part 2	B07	Cytoplasmic – some nuclear	LF82_682
Set 1 Part 2	B08	Cytoplasmic – nuclear associated	LF82_683

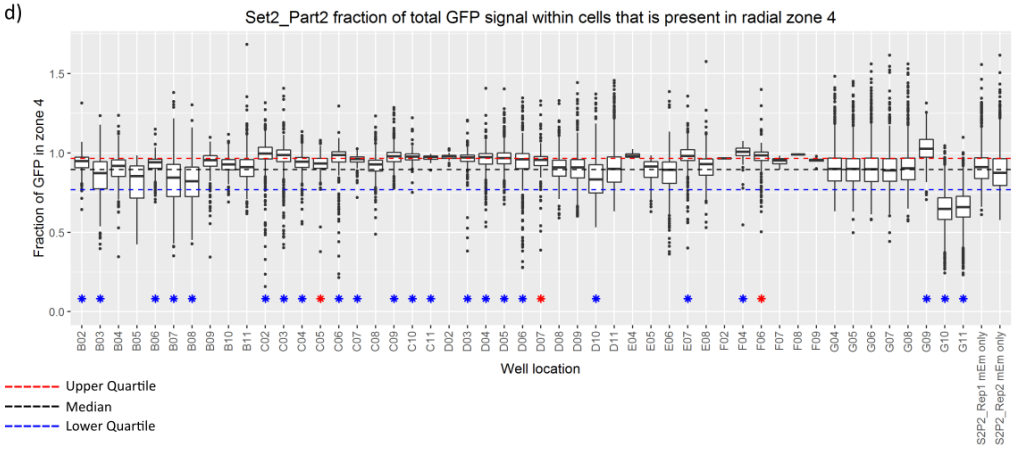
Plate	Well	GFP localisation	Respective LF82 CDS
Set 1 Part 2	B10	Cytoplasmic – some nuclear	LF82_686
Set 1 Part 2	**E04**	**Worm-like bundles in cytoplasm**	**LF82_314**
Set 1 Part 2	E05	Cytoplasmic – some nuclear	LF82_364
Set 1 Part 2	E08	Cytoplasmic	LF82_765
Set 1 Part 2	F07	Cytoplasmic	LF82_769
Set 2 Part 1	B02	Cytoplasmic – some nuclear	LF82_803
Set 2 Part 1	B03	Cytoplasmic – some nuclear	LF82_804
Set 2 Part 1	C08	Cytoplasmic	LF82_087
Set 2 Part 1	D05	Well distributed – cytoplasmic foci	LF82_219
Set 2 Part 1	D10	Cytoplasmic	LF82_256
Set 2 Part 1	D11	Cytoplasmic	LF82_259
Set 2 Part 1	E02	Cytoplasmic	LF82_282
Set 2 Part 1	E08	Multiple large foci in cytoplasm	LF82_324
Set 2 Part 1	F09	Cytoplasmic – nuclear associated	LF82_538
Set 2 Part 1	F10	Cytoplasmic	LF82_0557
Set 2 Part 1	F11	Cytoplasmic	LF82_582
Set 2 Part 2	C02	Multiple small foci in cytoplasm	LF82_678
Set 2 Part 2	C03	Cytoplasmic	LF82_684
Set 2 Part 2	C06	Cytoplasmic	LF82_695
Set 2 Part 2	D06	Cytoplasmic – multiple foci	LF82_731
Set 2 Part 2	E04	Cytoplasmic – some nuclear	LF82_260
Set 2 Part 2	E07	Multiple small foci in cytoplasm	LF82_675
Set 2 Part 2	F04	Cytoplasmic	LF82_261
Set 2 Part 2	F06	Cytoplasmic	LF82_604
Set 3 Part 1	B11	Cytoplasmic	LF82_762
Set 3 Part 1	E09	Puncta in cytoplasm	LF82_004
Set 3 Part 1	E10	Nuclear-associated puncta	LF82_772 *
Set 3 Part 1	F02	Cytoplasm – multiple small foci	LF82_1175
Set 3 Part 1	F03	Cytoplasm – potential cytoskeletal association	LF82_539
Set 3 Part 2	C05	Cytoplasm – multiple small foci	LF82_758
Set 3 Part 2	D05	Cytoplasmic – potential cytoskeletal	LF82_336

Plate	Well	GFP localisation	Respective LF82 CDS
		association	
Set 3 Part 2	D06	Cytoplasmic	LF82_347
Set 3 Part 2	E08	Cytoplasmic	LF82_121
Set 3 Part 2	F08	Cytoplasmic	LF82_266
Set 4	C02	Cytoplasmic	LF82_788
Set 4	C05	Cytoplasmic	LF82_803
Set 4	C08	Cytoplasmic – random regions of cytoplasm	LF82_130
Set 4	D04	Cytoplasmic	LF82_728
Set 4	D08	Cytoplasmic	LF82_805
Set 4	E02	Cytoplasmic	LF82_759
Set 4	E08	Cytoplasmic	LF82_661
Set 4	F02	Cytoplasmic, but variable	LF82_152
Set 4	**F03**	**Worm-like bundles in cytoplasm**	**LF82_314**


a)
Set1_Part1 fraction of total GFP signal within cells that is present in radial zone 4
Fraction of GFP in zone 4
Well location
S1P1_Rep1 mEm only
S1P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile


b)
Set1_Part2 fraction of total GFP signal within cells that is present in radial zone 4
Fraction of GFP in zone 4
Well location
S1P2_Rep1 mEm only
S1P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile


c)
Set2_Part1 fraction of total GFP signal within cells that is present in radial zone 4
Fraction of GFP in zone 4
Well location
S2P1_Rep1 mEm only
S2P1_Rep2 mEm only
Upper Quartile
Median
Lower Quartile


d)
Set2_Part2 fraction of total GFP signal within cells that is present in radial zone 4
Fraction of GFP in zone 4
Well location
S2P2_Rep1 mEm only
S2P2_Rep2 mEm only
Upper Quartile
Median
Lower Quartile

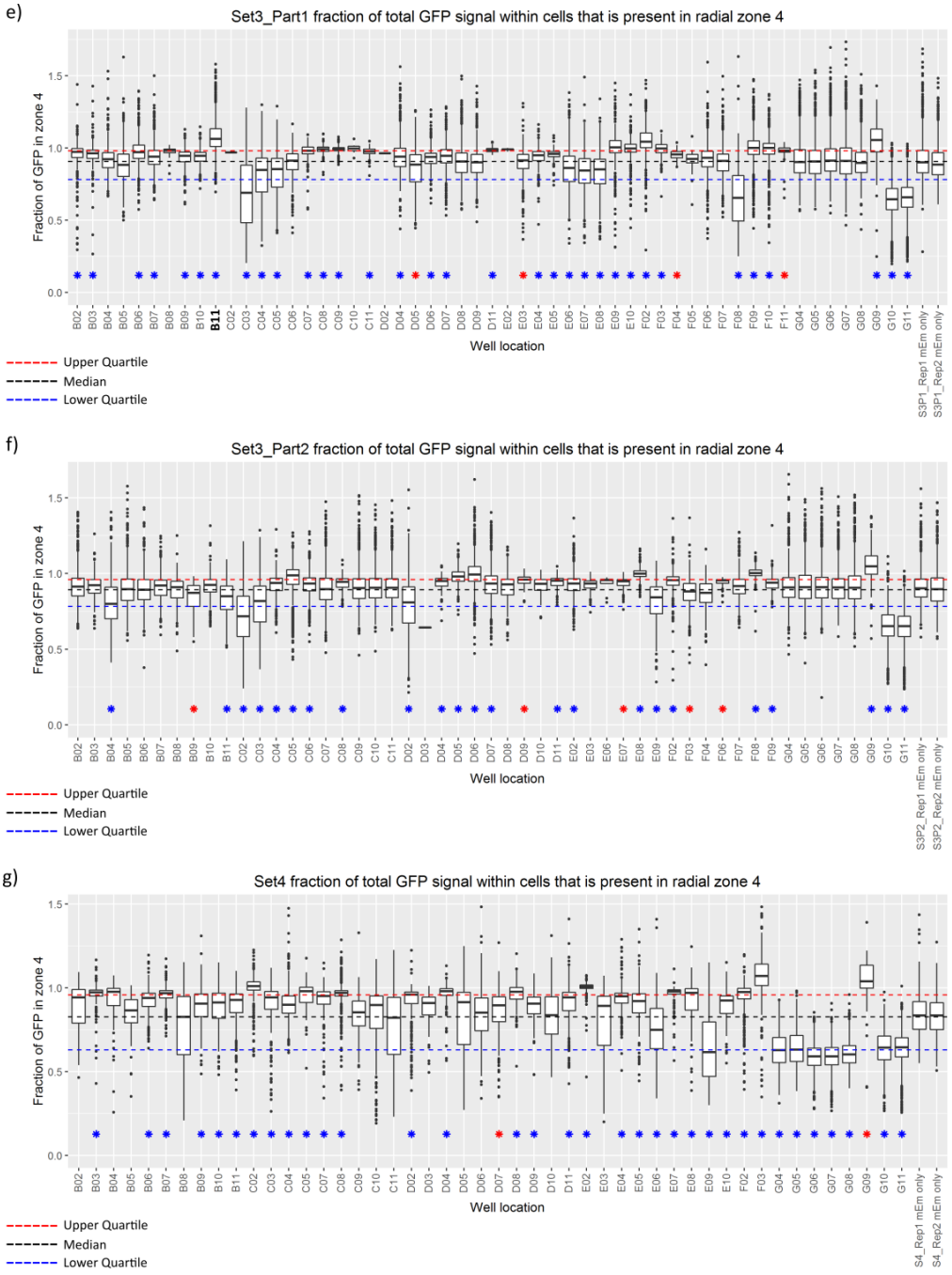


Figure 4.27. **Graphs showing the fraction of GFP signal detected within radial zone 4 of transfected cells**. Wells used to provide example images are highlighted in bold. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table 4.12. **Top 20 hits from radial zone 4 data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest. C*: data from contaminated well, discussed in Table 4.11.

Top 20 hits for the fraction of total GFP signal within cells that is present in radial zone 4				
Plate	Well name	Gene	Median	Standard deviation
Set3_Part1_Rep2	*G09*	*p62*	*1.66*	*0.27*
Set3_Part1	B11	LF82_468	1.69	0.16
Set4	F03	LF82_314	1.67	0.26
Set3_Part1	F02	LF82_1175	1.63	0.20
Set2_Part1	D11	LF82_259	1.61	0.18
Set3_Part2	F08	LF82_266	1.61	0.04
Set3_Part2	E08	LF82_121	1.60	0.05
Set3_Part1	E09	LF82_004	1.60	0.20
Set3_Part1	C10	LF82_441	1.59	0.04
Set3_Part2	D06	LF82_347	1.59	0.18
Set3_Part1	F03	LF82_539	1.59	0.09
Set3_Part1	F09	LF82_772	1.59	0.21
Set3_Part1	E10	LF82_772 C*	1.58	0.16
Set3_Part1	F10	LF82_793	1.58	0.15
Set4	C02	LF82_788	1.58	0.12
Set3_Part2	C05	LF82_758	1.58	0.17
Set3_Part1	C09	LF82_394	1.57	0.04
Set3_Part1	C08	LF82_356	1.57	0.06
Set4	E02	LF82_759	1.57	0.07
Set2_Part1	F10	LF82_0557	1.57	0.12
Set1_Part2	E04	LF82_314	1.56	0.17

4.6 Future perspectives

A screening system such as the one used here may be used in the future for the analysis of CDS expression from a wide range of bacteria. The system is open to adaptation, and could therefore be modified to search for alternative cellular features, such as *Salmonella* containing vesicles during invasion, or microtubules. However, an increase in microscopic magnification and resolution may be the modification of most interest, give the limitations reported here.

4.7 Summary

CellProfiler provides a robust and unbiased analysis of fluorescence microscopy data. While some data may be lost due to the same module parameters being applied across all images, doing this ensures that human bias does not occur during the image measurement stage. Furthermore, analysing this many images manually by eye would be tiresome, and may be impossible with larger screens. CellProfiler therefore provides a platform for high-content image analysis which can be used to identify phenotypes from a large image set.

The use of R to present data in an accessible format with an indication of statistical significance has proved to be reliable where CellProfiler was able to accurately analyse images. However, some analyses were less successful due to technical limitations, and lack of positive controls with which to compare results. CellProfiler Analyst proved to be highly useful in manually confirming any phenotypes suggested from generated graphs.

Multiple previously uncharacterised proteins were identified as potential effectors of mammalian cells. However, due to the lack of reliable detection of autophagosomes for enumeration and measurement, suspected autophagy effectors were not readily identified. Future experiments should look to improve the detection of autophagosomes, likely through the use of

higher resolution and higher magnification microscopy. However, this would require more imaging sites per well in order to screen enough cells to confirm the phenotype – increasing computing time required. Alternatively, software solutions could be found, for example, with the use of ImageJ to process mCherry-LC3 images differently and then resubmit them to the CellProfiler pipeline.

The use of CellProfiler and R to analyse microscopy data, along with CellProfiler Analyst to confirm observed phenotypes, has proven to be a lucrative method for the identification of potential effectors. This is true for nuclear and cytoplasmicly localised content, and especially so for the identification of GFP localisation using the radial zone measurement. Graphs were shown to allow the correct prediction of localisation information, as well as providing a useful base for the measurement of other cell features that would require further investigation (e.g. nuclei size). Identification of the particularly interesting worm-like phenotype of LF82_314 resulted in further investigation into the potential effector status of this gene.

Chapter 5: Analysis of candidate effectors

5.1 Background

5.1.1 Genes selected for deletion

Several AIEC LF82 genes were selected for analysis, including one primary hit from the screen, and several genes involved in the T6SS. The phenotypical experiments outlined in this chapter were performed as preliminary tests for future analyses of screen hits, and also served for initial analysis of p314.

5.1.1.1 LF82_314

From the screen described in Chapter 4, multiple 'hits' were identified via the analysis of screen data graphs. From these hits, the images taken from Set 1 Part 2 well E04 were analysed and found to display a strong phenotype – a sample from these images can be seen in figure 5.1, below.

The CDS suspected to be responsible for this phenotype was LF82_314 (hereby referred to as p314). From Figure 5.1, GFP localisation can be seen to be reminiscent of 'worms' or bundles of fibres which taper at the ends, curving to follow the nuclear or cytoplasmic membrane/s. Unlike many other phenotypes observed in the screen (nuclear dots, or cytoplasmic localisation, for example), this worm-like phenotype appeared for only one CDS in the entire screen.

While it was not initially clear which sub-cellular location p314-GFP localised to, if any, the fibrous appearance of the 'worms' suggested possible cytoskeletal component localisation, such as actin microfilaments, intermediate filaments or microtubules. Furthermore, as the adherence/invasion of LF82 has been previously shown to be mediated by actin and/or microtubules, p314 was investigated for its ability to localise to these regions and affect LF82 invasion (Boudeau et al. 1999).

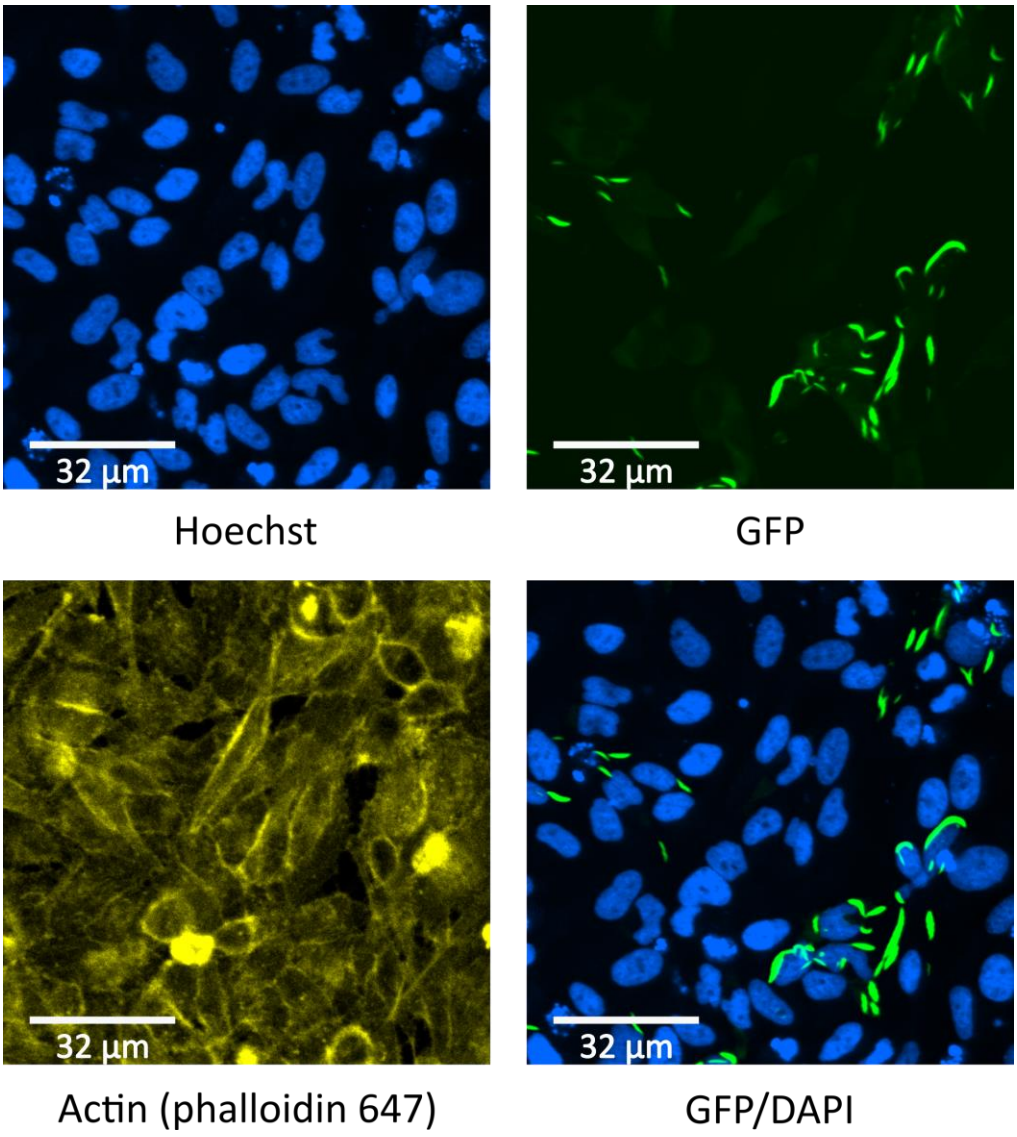


Figure 5.1. **Example images displaying the LF82_314 phenotype**. Well E04 from Set 1 Part 2 Replicate 2 (site 1), transfected using pCMV-mEm-4GS-LF82_314. A strong phenotype can be seen, with GFP localising to 'worm-like' structures in the cell cytoplasm.

While LF82_314 was identified as unique to LF82 within the limited *E. coli* genomes used for comparison, NCBI BLAST was used to show that it does bear homology to CDSs seen in other organisms. Most notably of these is *E. coli* strain NRG857C, which, as mentioned previously, is a member of the AIEC. Other organisms bearing similar CDSs include two strains of *Shigella boydii* which encode almost identical CDSs.

5.1.1.2 LF82 T6SS ClpV and TssE

The recently identified type 6 secretion system (T6SS) has been shown to be involved in the delivery of cargo to target cells, including the delivery of toxins (Hood et al. 2010). The AIEC LF82 genome encodes two T6SSs, but lacks any T3SS, suggesting that effector delivery may be handled by one or both of the encoded T6SSs (Büttner 2012; Miquel et al. 2010). A more thorough description of the T6SS can be seen in: Chapter 1: The type 6 secretion system.

Knockouts of both T6SSs were desired to test for effects of the loss of either or both systems. Several components of the T6SS were considered for deletion, including ClpV, and TssE.

ClpV is an AAA(+) ATPase whose status as an essential component of the T6SS is disputed – it is suggested that the T6SS apparatus can assemble and function to deliver one load of cargo before ClpV comes into play as a limiting factor. This is due to the ability of ClpV to recognise only the contracted form of the T6SS, which is only present after one successful cargo delivery has occurred, an example of this can be seen in Figure 1.12 (Ho et al. 2014).

TssE, on the other hand, has been shown to be essential to T6SS function. TssE is part of the secretion system base plate – a protein complex upon which the rest of the T6SS assembles – and its removal from this complex

prevents the formation of the HCP-rich needle structure which usually extends across the extracellular milieu into target cells (Brunet et al. 2015).

While other members of the T6SS structure were also considered as potential targets for deletion, TssE was selected as it appears to interact with multiple T6SS-associated proteins, thereby providing maximal disruption to the T6SS and ensuring inactivity (Brunet et al. 2015). Due to the presence of two T6SS in LF82, the PAI-1 TssE and PAI-3 TssE were both targeted for deletion. These were referred to as TssE1 and TssE3, respectively.

5.2 Aims

The characterisation of potential LF82 effectors may help to provide an insight into the organism's intracellular lifestyle. Furthermore, as the role of the T6SS in the delivery of LF82 effectors remains unconfirmed, research into the use of this secretion system in LF82 is essential. The suspected involvement of LF82 in CD, along with the elusive method with which it evades autophagy, makes this area of research highly lucrative. The aims of this chapter were to:

- Create LF82 deletion mutants for *LF82_314* (p314) – a potential effector identified from the screen.
- Create LF82 mutants with disabled T6SS/s via the creation of LF82 Δ*TssE1*, LF82 Δ*TssE3,* and LF82 Δ*TssE1*-Δ*TssE3*.
- Begin characterisation of p314 through co-localisation assays.
- Determine invasion phenotypes of Δ*p314*, Δ*TssE1*, Δ*TssE3*, and Δ*TssE1*-Δ*TssE3*.
- Show that co-localisation assays and invasion assays can be used as preliminary downstream analysis techniques for other hits from the screen.

5.3 Gene deletion

5.3.1 Deletion strategies

Several strategies were used in an attempt to remove p314, TssE1 and TssE3 from the LF82 genome. These included the one-step inactivation technique described by Datsenko & Wanner (2000), the GeneDoctoring system described by Lee et al. (2009) and the no-SCAR technique described by Reisch & Prather (2015).

The one-step technique for the inactivation of chromosomal genes was attempted first, due to the availability of required strains in house (courtesy of Dr. Rob Delahay, University of Nottingham, UK). This system relies upon the introduction of a vector which constitutively expresses λ Red recombinase in the target organism. The organism is then transformed with a linear PCR fragment with which genomic recombination can occur, thereby replacing the target gene with another desired sequence.

The one-step system is known to have several limitations relating to recombination efficiency in commensal *E. coli* strains, as well as having the potential for off-target recombination due to the sustained overexpression of λ Red recombinase (Lee et al. 2009). Furthermore, some pathogenic strains of *E. coli* have been reported to display little or no recombination using this method. These issues have been described by Lee et al. (2009) and led to the development of an alternative system: GeneDoctoring.

As with the one-step technique, GeneDoctoring utilises λ Red recombinase for recombination, however, where the one-step technique does not attempt to regulate the amount of λ Red recombinase expressed, GeneDoctoring does. The system works by first introducing two vectors, one encoding the λ Red recombinase and the I-SceI restriction enzyme, and another containing a region of homology to the sequence selected for deletion. This region of homology is bookended by I-SceI restriction sites. The simultaneous induction

of λ Red recombinase and I-SceI expression is stimulated via the addition of L-arabinose. This simultaneous expression exposes the cell to a limited 'burst' of λ Red recombinase, which is sufficient to recombine DNA linearised by the I-SceI restriction enzyme into the genome. I-SceI restriction sites on the λ Red recombinase-encoding vector result in self-cleaving, ending the burst of λ Red recombinase as well as I-SceI expression.

Both of the above systems were utilised in an attempt to remove p314, TssE1 and TssE3 CDSs from the LF82 genome. Multiple attempts were made, using different homology sequences and various durations of λ Red recombinase expression, but none were successful. The introduction of the linear fragment proved difficult in the one-step system, and the counter-selection of successful deletion mutants using sucrose and *SacB* difficult in the GeneDoctoring system.

It is not clear why either system failed; it may be that AIEC strain LF82 is even more difficult to genetically manipulate than other pathogenic bacteria, although this is not confirmed. Furthermore, as both systems failed at different stages, it is not clear whether one particular process was inefficient in LF82, or whether each system simply had its own shortcomings. An alternative, more streamlined system for gene deletion was therefore sought to limit the steps in which recombination efficiency could be reduced.

5.3.1.1 The no-SCAR (Scarless Cas9 Assisted Recombineering) gene deletion system

The Scarless Cas9 Assisted Recombineering system (no-SCAR) described by Reisch & Prather (2015) utilises the CRISPR/Cas9 system for counterselection. CRISPR/Cas9, hailed as a revolutionary system for genome editing, is utilised in this system to create double-strand breaks (DSBs) in the *E. coli* genome via the use of the Cas9 nuclease. This results in bacterial cell death due to *E.*

coli's lack of double-strand break repair systems (Reisch & Prather 2015; Doudna & Charpentier 2014; Bowater & Doherty 2006).

The lack of a DSB-repair system in *E. coli* is capitalised upon to differentiate between deletion mutants and wild type organisms. Cas9 is responsible for DNA cleavage, and is directed to its target sequence by small-guide RNA (sgRNA). sgRNA can be reprogrammed to target almost any DNA sequence with relative ease. This feature is utilised in the no-SCAR system to target Cas9 to the deletion gene of interest. By inducing Cas9 expression, the genomes of wild type cells are therefore cleaved if they contain DNA complimentary to the sequence of the Cas9 sgRNA, resulting in cell death. Meanwhile, cellular survival suggests a lack of target DNA, indicative of a successful deletion mutant.

5.3.2 Gene deletion using the no-SCAR system

The production of deletion mutants was achieved using the no-SCAR system, an overview of which can be seen in Figure 5.2, below (Reisch & Prather 2015). To use this system, two vectors were required: pKD-sgRNA and pCas9-cr4. While pCas9-cr4 did not require any modification, it was necessary to modify pKD-sgRNA to alter the target of the Cas9 DNA nuclease.

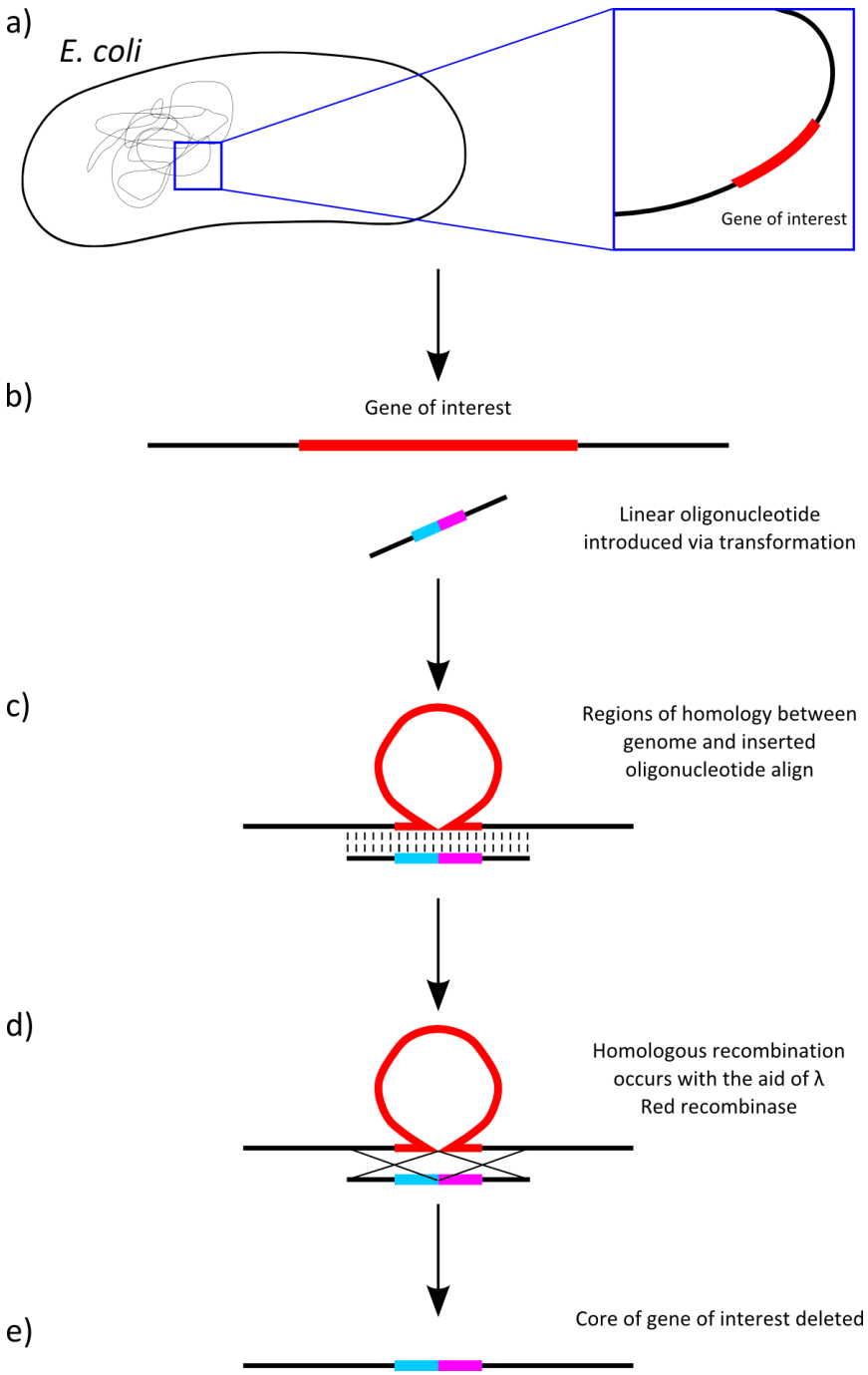


Figure 5.2. **An example of gene deletion with the no-SCAR system**. a) A gene of interest is selected for deletion from the target genome. b) A linear oligonucleotide is created which contains regions of homology to external/terminal regions of the gene of interest. c) The oligonucleotide hybridises to the target site DNA on the lagging strand. d) When λ Red recombinase is expressed, recombination occurs between the host DNA and the oligonucleotide. e) The host DNA now lacks the gene of interest. Depending on design, this may result in the removal of the majority of the gene, or its entire coding sequence. The no-SCAR system was developed by Reisch & Prather (2015).

5.3.2.1 Small guide RNA vector creation

The pKD-sgRNA-ack vector acquired from Addgene was encoded to target Cas9 to the *ack* gene in *E. coli*, encoding an acetate kinase. This vector was re-created for use as a positive control throughout the use of the no-SCAR system.

To alter the sgRNA encoded by pKD-sgRNA-ack, oligonucleotides were designed with homology to a ~20 nt region of the gene destined for deletion. The only requirement for this region was that it was preceded by the sequence nGG (the so-called protospacer adjacent motif (PAM) site) which is necessary for Cas9 recognition and cleavage of the target site. Once designed, 23 nt sequences bearing homology to the pKD-sgRNA vector were added.

Oligonucleotides were used in a PCR with two 'common' primers which contained homology to regions on the opposite side of the pKD-sgRNA vector (see Figure 2.1 in Circular polymerase extension cloning (CPEC)). This resulted in the amplification of two products of size ~4.4 Kb and ~2.8 Kb which can be seen in Figure 5.3.

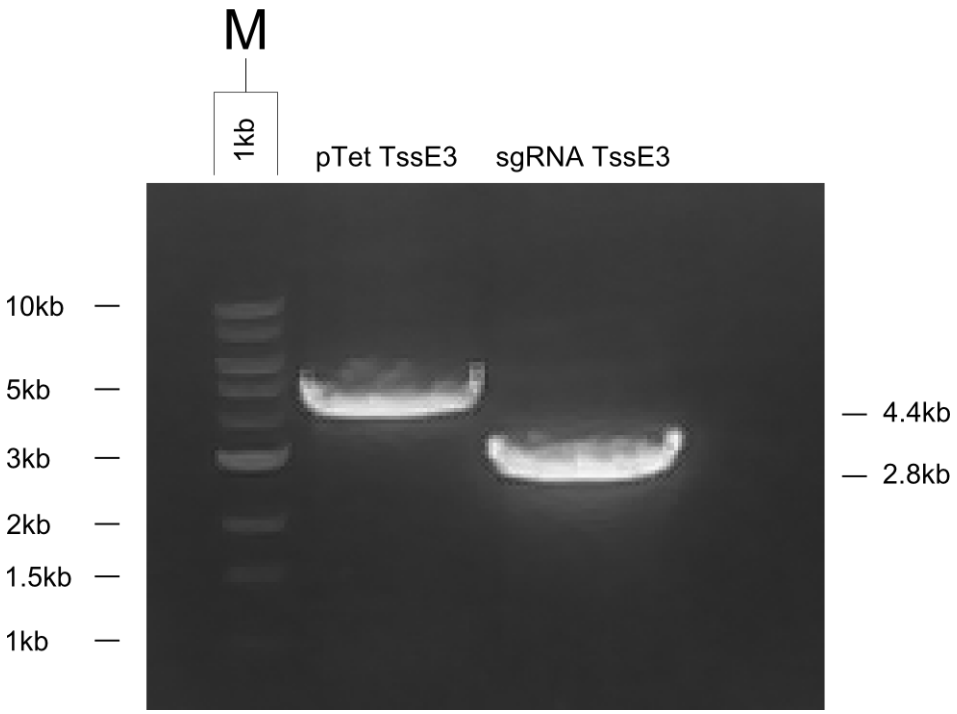


Figure 5.3. **Amplification of vector fragments for use in the CPEC reaction**. Gel electrophoresis of PCR to amplify the pKD-sgRNA-ack vector using TssE3 primers. Resulting bands were of expected size, with one fragment of size 4.4 kb, and the other of size 2.8 kb. These fragments are referred to as the pTet and sgRNA fragments, respectively.

Once successfully amplified, pKD-sgRNA fragments were subjected to a CPEC PCR. As can be seen in Figure 5.4, successfully amplified pKD-sgRNA vector was not readily identifiable from via gel electrophoresis, however, as it was still possible that vector was present, PCR products were purified and transformed into XL1-blue cells. This resulted in an average of 28 colony forming units (CFU) for test samples, compared to 2 CFU and 0 CFU for negative controls consisting of products not subjected to the CPEC reaction, and water, respectively. This suggests that successful CPEC PCR products were likely present, but may not have been visible when subjected to gel electrophoresis. Purified vectors were sequenced to confirm insertion of the new sgRNA at the desired location. pKD-sgRNA vectors were successfully created to target *LF82_314*, *TssE1*, *TssE3*, and *ack* genes.

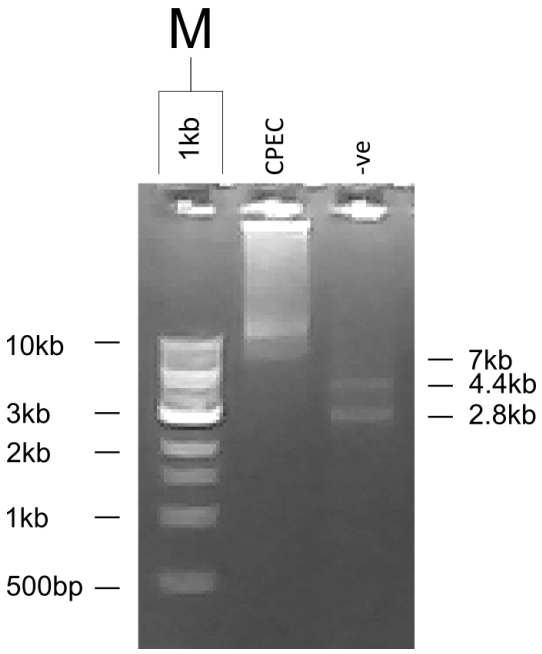


Figure 5.4. **Gel electrophoresis of the completed CPEC reaction**. The gel shows pKD-sgRNA-TssE3 CPEC PCR products, showing CPEC PCR product in Lane 2 (CPEC), and pre-CPEC product in Lane 3 (-ve). No pre-CPEC fragments are visible in Lane 2, suggesting that they were both used during the PCR, however, the expected ~7kb completed pKD-sgRNA vector is difficult to differentiate from the large smear at the top of Lane 2. Despite this, the CPEC PCR was shown to be successful using DNA sequencing.

5.3.2.2 Induction of deletion

The use of arabinose induction to control expression of λ Red recombinase was initially unsuccessful. This was later determined to likely be due to the use of super optimal broth with catabolite repression (SOC) – a modified form of super optimal broth (SOB) containing 20 mM glucose as an additional component. Glucose is known to repress *araC* promoter activity, and thus its inclusion in recovery medium post-arabinose induction likely prevented expression of λ Red recombinase (Miyada et al. 1984). SOB was therefore utilised for the remainder of the deletion attempts.

The no-SCAR protocol utilises the glycerol-mannitol method for the preparation of electrocompetent cells (see: Glycerol-mannitol electrocompetent cell preparation). Warren (2011) claimed that the use of this method allowed the production of electrocompetent *E. coli* of similar or better transformation efficiency than those prepared using traditional techniques, and with lower inter-batch variation. This method is also reportedly less laborious than traditional techniques. These claims were supported in-house when both methods were compared. The traditional method of electrocompetent cell preparation resulted in ~$7x10^7$ transformants per µg DNA, while the glycerol/mannitol method resulted in ~$2.8x10^8$ transformants per µg DNA. While these values were far below those achieved by Warren (2011) with XL1-blue (~$8.9x10^9$), the use of a wild-type *E. coli* strain (LF82) for these experiments may explain these results.

After arabinose induction and recovery in SOB, deletion was induced via the introduction of a linear ssDNA oligonucleotide. Each deletion oligonucleotide was designed with three 5' and 3' phosphorothioate bonds to reduce the likelihood of degradation once introduced to the host organism. Oligonucleotides were also designed to target the lagging strand of each respective CDS, as recombination is predicted to occur with an efficiency over two orders of magnitude higher at the lagging strand of the replication fork, than it is at the leading strand (Ellis et al. 2001).

Cells were recovered to allow cellular replication and efficient recombination at replication forks. Selection for cells having undergone recombination was carried out using Cas9 and the tailored sgRNA.

Anhydrotetracycline (aTc) was utilised to deactivate repression of Cas9 expression. aTc exhibits extremely poor association with the 30S ribosomal subunit – making it effectively bacteriocidally-null – and yet displays over 35-fold more affinity for the Tet repressor than tetracycline (Rasmussen et al. 1991). These features reduced the final concentration of aTc required, and

lowered experimental complexity in comparison to the use of tetracycline for repression.

The no-SCAR protocol outlined by Reisch & Prather (2015) provides an estimate of 1 in 16 expected Cas9 escape mutants per experiment. These mutants typically appear on agar plates alongside true successful deletion mutants due to their genomes avoiding cleavage by Cas9 and DSBs. In contrast to the reportedly low incidence of escape mutants seen with the no-SCAR system, the number of escape mutants far exceeded the number of successful deletion mutants, in our hands.

On average, only 1 in 20 deletion mutants appeared successful when colonies were picked from post-induction plates and subjected to colony PCRs. An example of this can be seen in Figure 5.5, where 21 colonies were selected and tested for potential deletion. From this, one potential deletion mutant was seen in lane 10, and was subsequently confirmed via sequencing.

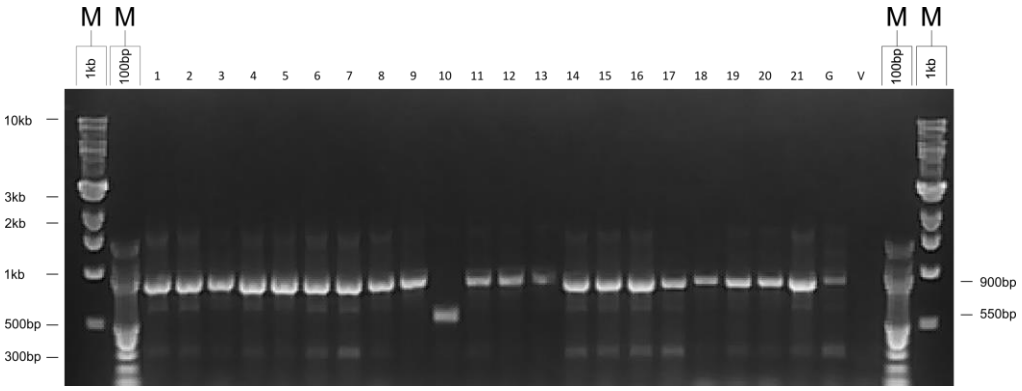


Figure 5.5. **Colony PCR of potential deletion mutants from the no-SCAR system.** 21 post-aTc-selection colonies from an attempted *TssE1* deletion were subjected to colony PCR, with expected band sizes of 651 bp for successful deletions, and 902 bp for unsuccessful deletions. The fragment in lane 10 suggests that successful deletion occurred for clone 10, with a fragment size of ~600-700 bp. 'G' refers to a colony PCR performed using genomic DNA as a template, and 'V' with pKD-sgRNA-TssE1 as a template.

While the frequency of successful deletion using the no-SCAR method was not ideal, the testing of ~40 clones for the production of each deletion mutant ensured at least one successful deletion event. This process was repeated for *TssE3*, *p314*, and *ack* and resulted in the production of successful deletion mutants, which were confirmed via sequencing across the scar of the deletion region.

While success was seen via the use of the no-SCAR system, it is not clear how the overall efficiency of this method can be improved. It may be that the lower efficiency seen here, in comparison to that reported by Reisch & Prather (2015), is strain dependent and is nearing the maximum efficiency possible using LF82.

5.4 p314 phenotype analysis

The filamentous 'worm-like' phenotype observed with p314 expression was unlike that seen for any other member of the screen. Despite this distinct phenotype, however, other examples of similar phenotypes were difficult to find in the literature. The opinions of several academics were obtained regarding this phenotype, although this did not shed any light on potential p314 localisation.

Despite a lack of evidence for the target of p314 localisation, co-localisation assays were performed (see below). Not only were these conducted to ascertain information regarding p314 localisation to actin, intermediate filaments and/or microtubules, but also to begin preliminary optimisation of co-localisation assays for the future analysis of other screen hits. While these were being performed, however, evidence of a similar phenotype to that seen with p314 was found.

Work by other groups regarding the disruption of microtubules appears to result in a similar phenotype as p314 expression. For example, Rohena &

Mooberry (2014) have shown that the stabilisation of tubulin by paclitaxel leads to increased levels of insoluble tubulin, resulting in bundles of microtubules in the cytoplasm (Figure 5.6 a and b). Furthermore, Mollinari et al. (2002) have shown that the overexpression of a mitotic spindle-associated protein called PRC1 results in the bundling of microtubules (Figure 5.6 c). Although these examples suggest the involvement of microtubules in the phenotype observed during p314 expression, confirming this proved difficult.

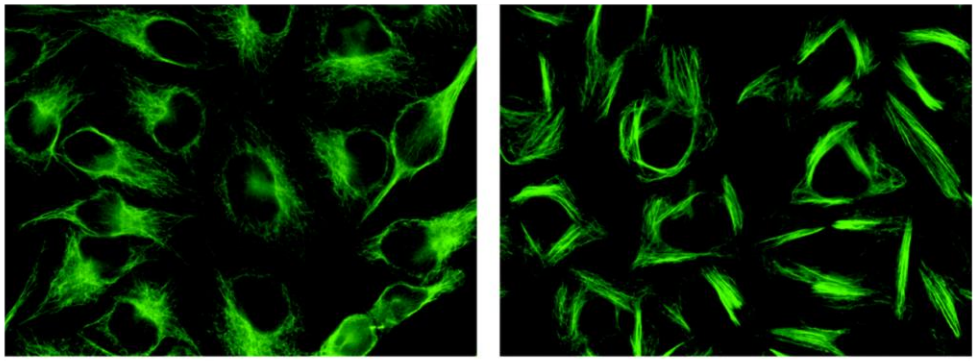
5.4.1 Phenotype confirmation

Prior to any co-localisation assays, it was first confirmed that p314 was responsible for the 'worm-like' filamentous phenotype. Transfections were performed using pCMV-mEm-4GS-p314, and once again resulted in the 'worm-like' phenotype seen in the screen (Figure 5.1). The pCMV-mEm-4GS vector for the transfection and expression of the p314-mEm fusion was then sequenced, confirming the presence of the p314.

5.4.2 Co-localisation assays

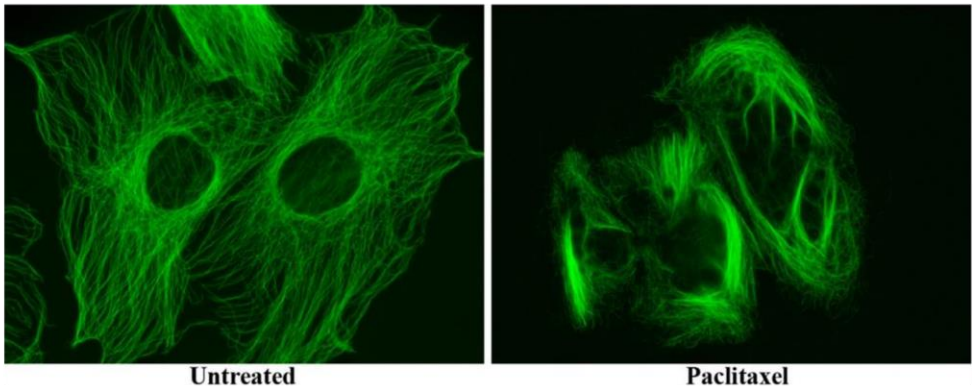
The screen reported in Chapter 4 attempted to identify some co-localisation information through the use of phalloidin 647, Hoechst, mCherry-LC3, and GFP. However, as described in the Graph inconsistencies section, these attempts were not as effective as desired – particularly when considering the GFP/Actin correlation. Due to this, downstream analysis was performed to help determine localisation information. This analysis included co-expression assays using a range of fluorescent-tagged proteins from the Michael Davidson collection (Addgene), the use of phalloidin 647 and Hoechst staining, and immunofluorescence. Initial experiments were performed to test a range of fluorescent-tagged reporters, seen in Table 5.1, below.

a)



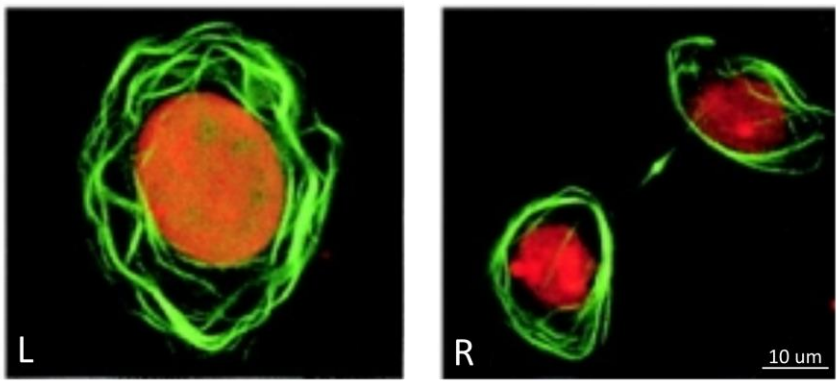
Untreated Paclitaxel

b)

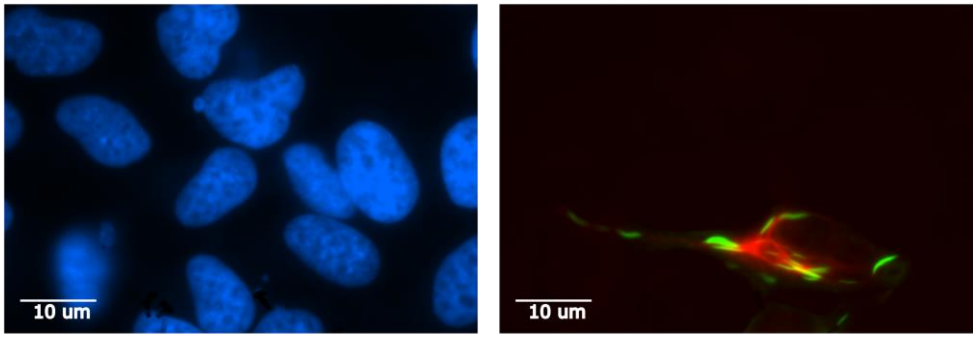


Untreated Paclitaxel

c)



d)



Hoechst GFP/mCherry merge

Figure 5.6. **Examples of microtubule stabilisation, resembling the p314 phenotype**. a) Untreated HeLa and b) A10 cells, cells compared to cells treated with paclitaxel – known to stabilise microtubules. Adapted from Rohena & Mooberry (2014). c) HeLa cells overexpressing a mitotic spindle-associated protein-EGFP fusion (PRC1-EGFP) (green) with nuclei stained with Hoechst (red). L displays perinuclear microtubule filaments at interphase, and R displays a similar phenotype at the end of telophase. d) An example of the phenotype displayed during p314-GFP expression for comparison, see Figure 5.7 for more information. Adapted from Mollinari et al. (2002).

Table 5.1. **Fluorescent-tagged protein reporters acquired from the Michael Davidson collection (Addgene)**. All reporters are cloned into plasmids bearing CMV promoters and SV40 polyA signals for mammalian expression. Fluorescent fusions are described positioned in relation to localisation specific insert, e.g. C-terminal fusions are fused to fluorescent proteins at the C-terminus of the localisation-generating sequence.

Reporter name	Fluorescent tag	Localisation	Localisation signal
Vimentin 7	mCherry (C-terminal fusion)	Intermediate filaments	Full-length human vimentin
Farnesyl-5	mCherry (N-terminal fusion)	Plasma membrane	C-terminal 20 aa of H-Ras (human)
Fibrillarin-7	FusionRed (mKate2.7) (N-terminal fusion)	Nucleoli	Full-length human fibrillarin
Endo-14	mCherry (N-terminal fusion)	Endosomes	192 aa of human RhoB
Cathepsin B-6	mCherry (C-terminal fusion)	Lysosomes	333 aa of rat Cathepsin-B

5.4.2.1 Cell cytoskeleton

There are three main components of the cell cytoskeleton: microtubules, intermediate filaments (IFs), and actin microfilaments. The visualisation of these components was attempted using vimentin-mCherry expression, phalloidin 647 staining, and immunofluorescence, respectively. Transfections were performed using vectors from the Michael Davidson collection (addgene).

Vimentin is an IF protein which, when expressed as an mCherry fusion, allows visualisation of intermediate filaments (IFs). IFs are a component of the cell

cytoskeleton, and, as their name suggests, have an intermediate size in comparison to microtubules and actin microfilaments (typically 8-15 nm diameter) (Margiotta & Bucci 2016). Co-transfections were performed using Lipofectamine 3000 to introduce both the GFP-p314 and vimentin-mcherry vectors to HeLa cells. From these, fluorescence microscopy was performed to analyse cells for co-localisation.

Figure 5.7 shows an example of cells co-expressing GFP-p314 and vimentin-mcherry. From these images, there did not appear to be any evidence for the co-localisation of p314 with IFs. Control cells were transfected with empty pCMV-mEm-4GS vector and vimentin-mcherry, and can be seen in Figure 5.8. Phalloidin 647 was not used for these cells, but can be seen in Figures 5.9 and 5.10, where it was used during mCherry-farnesyl expression. As with intermediate filaments, no co-localisation was apparent with actin either.

The visualisation of microtubules was attempted via the detection of β-tubulin via immunofluorescence. However, despite the use of two different primary antibodies, these experiments were unsuccessful. It remains unclear why this was the case.

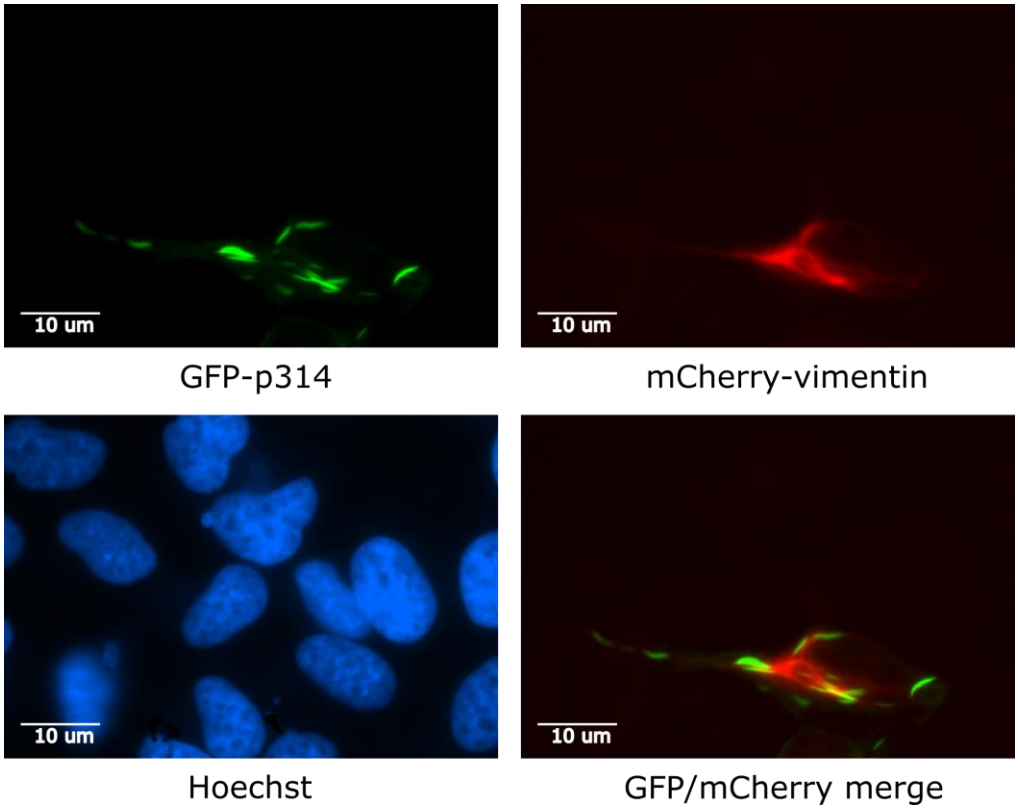


Figure 5.7. **HeLa cell co-localisation assay with p314 and intermediate filaments**. HeLa cells expressing GFP-p314 (green) and vimentin-mcherry (red) do not appear to display co-localisation between p314 and IFs.

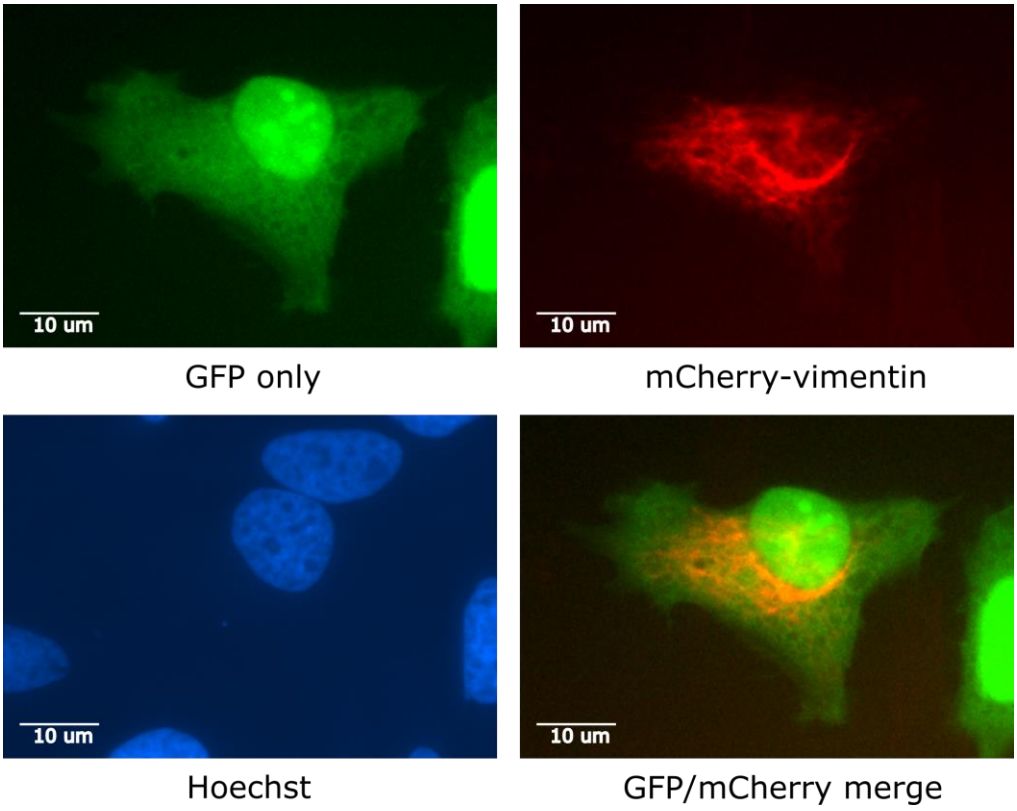


Figure 5.8. **HeLa cell co-localisation assay control with GFP only, and intermediate filaments**. HeLa cells expressing GFP (green) and vimentin-mcherry (red) do not appear to display co-localisation, as GFP is dispersed throughout the cell.

5.4.2.2 Other cellular components

While it is believed that p314 localises to a component of the cell cytoskeleton, other co-localisation assays were performed to confirm the visualisation of alternative cell components. This was performed to aid future research into other hits from the screen. Cellular components visualised included: the cytoplasmic membrane, nucleoli, endosomes, and lysosomes. An example of images seen from the expression of mCherry-farnesyl to visualise the cytoplasmic membrane can be seen in Figures 5.9 and 5.10.

Other vectors for the detection of a wide range of cell components can be found in the Michael Davidson collection (Addgene), and will likely be used in the future to help determine the localisation of other hits from the screen.

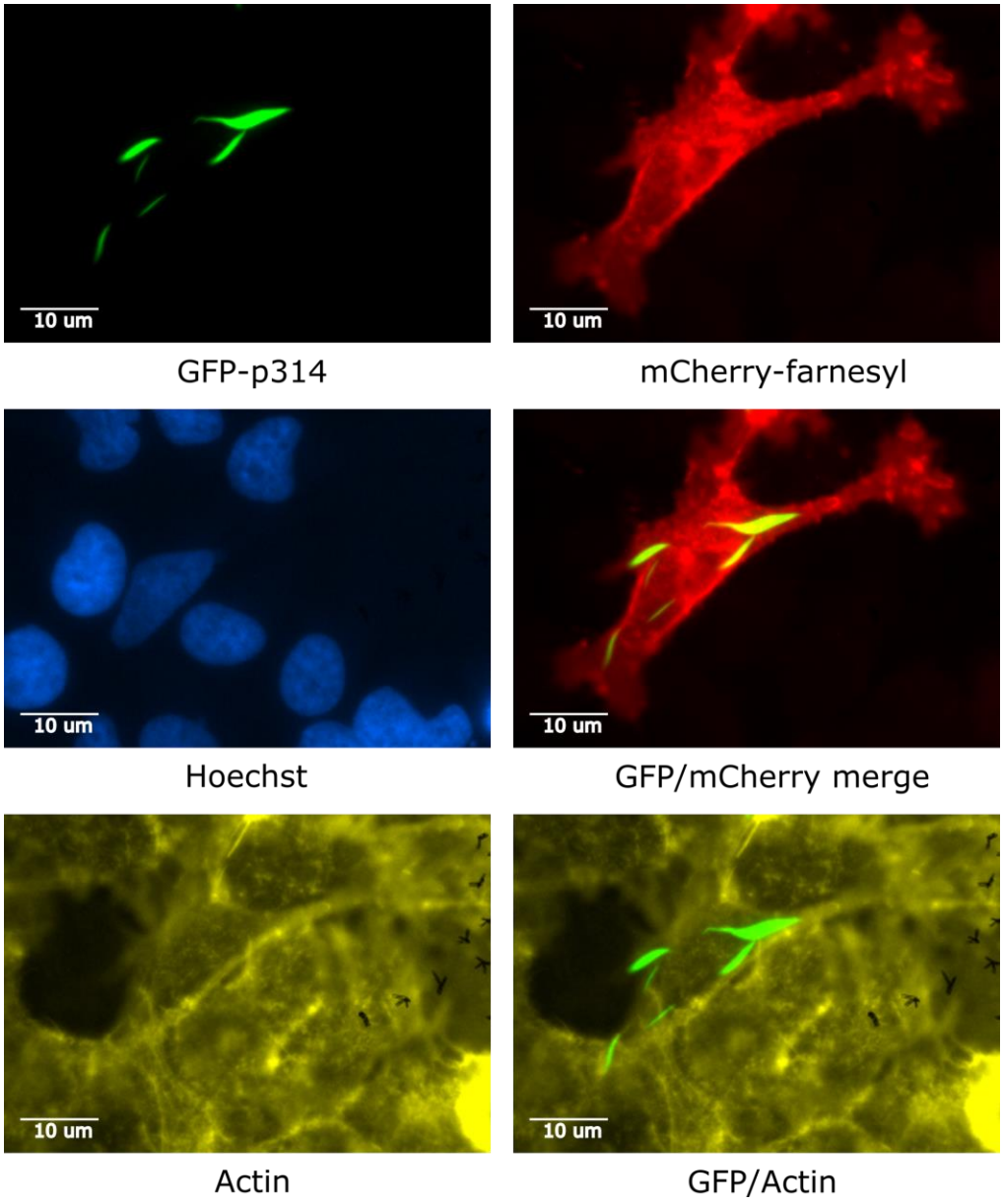


Figure 5.9. **HeLa cell co-localisation assay with p314, the cell membrane, and actin microfilaments**. HeLa cells expressing GFP-p314 (green) and mCherry-farnesyl (red) do not appear to display co-localisation between p314 and the cell membrane. Likewise, no obvious co-localisation is apparent with p314 and actin.

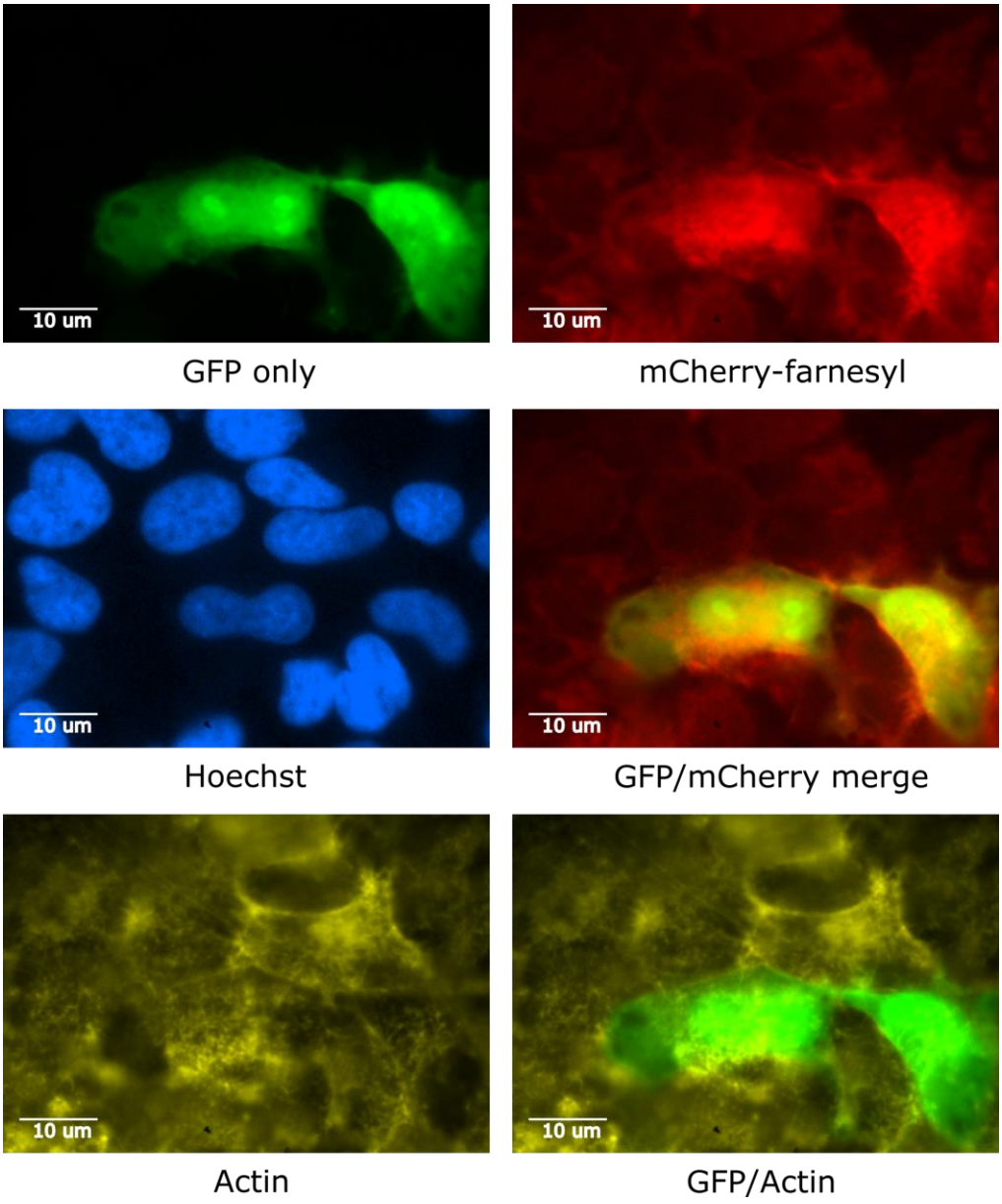


Figure 5.10. **HeLa cell co-localisation assay with GFP, the cell membrane, and actin microfilaments**. HeLa cells expressing empty pCMV-mEm-4GS (green) and mCherry-farnesyl (red) do not appear to display any specific co-localisation between GFP and the cell membrane, as GFP is dispersed throughout the cell. Likewise, no obvious co-localisation is apparent with GFP and actin.

5.4.3 Detection of secretion

Multiple hits were observed during screen analysis, and the classification of their resulting phenotypes was begun using the methods described above. However, as these CDSs were directly expressed in eukaryotic cells, it is not clear whether they are natively translocated to this site or not. *In vivo* p314, for example, may or may not be secreted into the target eukaryotic cell. It is therefore important to consider the secretion status of any putative effectors, to gain insight into the likelihood of them being secreted into the host cell during infection.

A method was therefore sought to determine the secretion status of any screen hits identified. This was initially attempted via the use of a GSK tag (see below), however, recent progress has been continued by other lab members using a split-GFP-based strategy.

5.4.3.1 Detection of protein secretion using a glycogen synthase kinase (GSK) tag

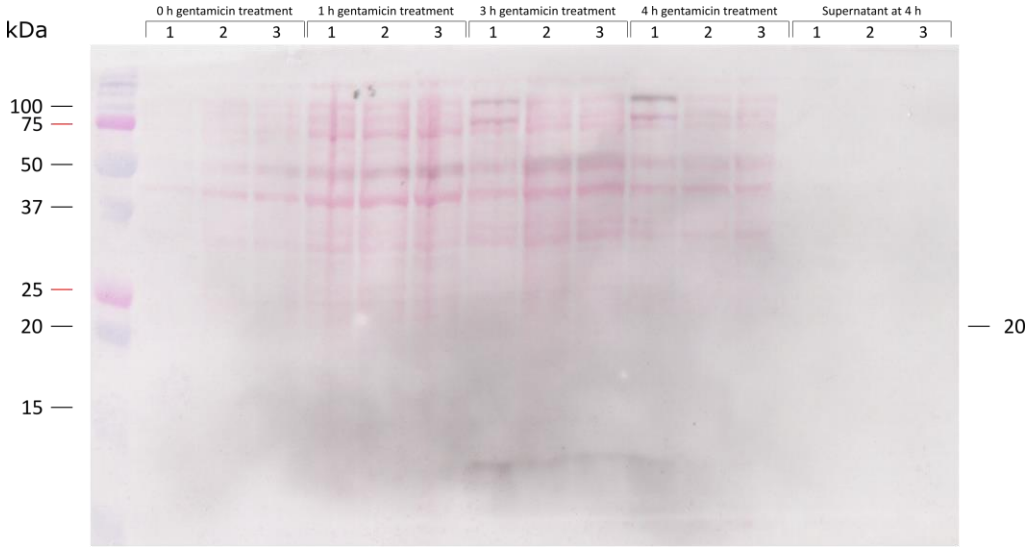
The use of a C-terminal, 13 amino acid GSK tag has been described by Garcia et al. (2006) for the detection of proteins delivered using the T3SS and T4SS; it has also been used for the detection of proteins secreted by the T6SS (Bröms et al. 2012). The glycogen synthase kinase (GSK) tag is phosphorylated by host cell machinery upon entry into a eukaryotic cell. In contrast, the GSK tag remains unphosphorylated when it remains within the bacterium. Due to the small size of this tag, the disruption of protein function and secretion signalling is minimised, especially when compared to alternative techniques (Garcia et al. 2006). Once expressed, GSK-tagged proteins delivered to the eukaryotic cell cytoplasm are detectable via the use of phosphospecific anti-GSK antibodies.

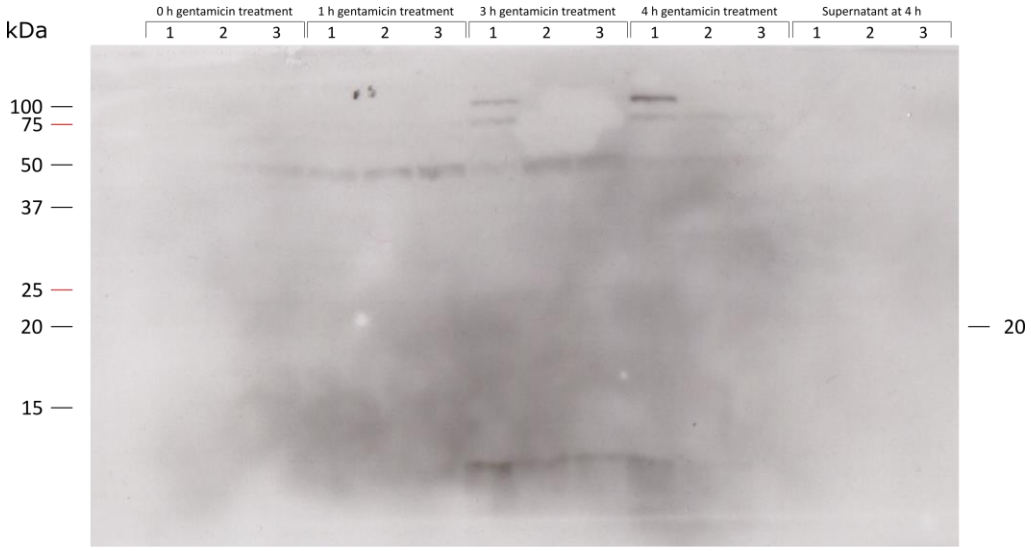
Here the GSK tag was trialled pre-screen analysis, and was therefore not tested on any hits identified from the screen; it was instead tested using

LF82_152 (referred to as p152 – mentioned previously in: Successful detection of autophagosomes). A vector for the arabinose inducible expression of p152 with a C-terminal GSK tag was created. This was transformed into LF82 and three bacterial samples prepared: LF82 wild-type ('1' in Figure 5.11), LF82 pBAD18-p152-GSK (uninduced) ('2' in Figure 5.11), LF82 pBAD18-p152-GSK (induced) ('3' in Figure 5.11). From these, an invasion assay was performed, with bacteria allowed to invade HeLa cells for 2 h before samples were taken. After this initial invasion time, samples were taken at 0 h (no gentamicin treatment), 1 h, 3 h, and 4 h (1, 3, and 4 h of gentamicin treatment, respectively). HeLa cells were pelleted and prepared for SDS-PAGE, followed by a Western blot, and ponceau staining of the membrane. The PVDF membrane was then probed using anti-GSK antibodies (for the detection of both phosphorylated and unphosphorylated GSK).

The ponceau staining showed that protein was present and was successfully transferred from the SDS-PAGE gel onto the membrane (See Figure 5.11). However, Western blot analysis did not reveal any of the expected bands at ~20 kDa for the p152-GSK fusion, with both unphosphorylated and phosphorylated GSK tags remaining undetected. The probed and ponceau stained Western blot membrane can be seen in Figure 5.11.

To account for potential low expression levels of the p152-GSK fusion, the GSK assay was repeated using 5x more protein for SDS-PAGE. However, once again, no presence of p152-GSK was detected. Finally, lysates of induced/uninduced bacteria were included, but again no signal was detected. This strategy for the determination of protein secretion status was abandoned in favour of a split-GFP-based approach, performed by other lab members.


0 h gentamicin treatment
1 h gentamicin treatment
3 h gentamicin treatment
4 h gentamicin treatment
Supernatant at 4 h
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
kDa
100
75
50
37
25
20
15
20


0 h gentamicin treatment
1 h gentamicin treatment
3 h gentamicin treatment
4 h gentamicin treatment
Supernatant at 4 h
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
kDa
100
75
50
37
25
20
15
20

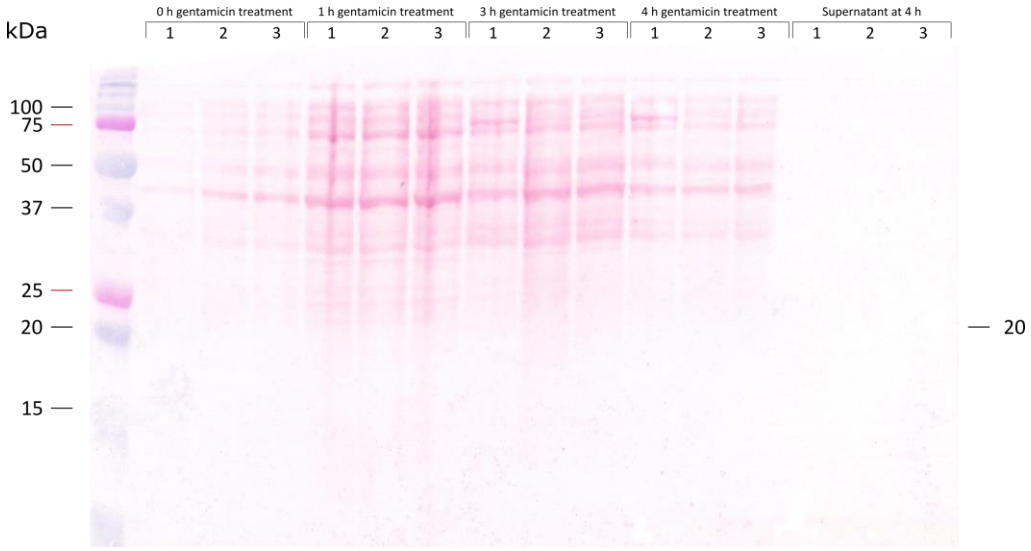

0 h gentamicin treatment
1 h gentamicin treatment
3 h gentamicin treatment
4 h gentamicin treatment
Supernatant at 4 h
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
kDa
100
75
50
37
25
20
15
20

Figure 5.11. **A Western blot from an unsuccessful GSK secretion assay**. Membranes were stained using ponceau stain to visualise protein (bottom panel). Western blots (middle panel) were performed to probe for p152-GSK (~20 kDa). The top panel shows both detection methods overlaid. As can be seen, no bands were detected within the expected range. Protein standards are in the leftmost lane, and each band is labelled with its associated kDa. To the right of each image, there is a label for the 20 kDa region in which p152-GSK bands were expected.

5.5 Invasion assays

The ability of wild-type LF82 to invade HeLa cells was compared to that of deletion mutants using invasion assays based upon the well-established gentamicin protection assay (GPA) (Mandell 1973). The observation that bacteria can be protected from antibiotic challenge by their residence inside a host cell, such as a leukocyte or other mammalian cell led to the development of the GPA (Mandell 1973). The assay depends upon bacteria invading a host cell (e.g. HeLa or Caco-2), and their survival when the cell population is treated with gentamicin. The GPA concept was therefore utilised for invasion assays here, in which bacteria that do not invade eukaryotic cells are killed via gentamicin, allowing enumeration of invaded bacteria by gentle host cell lysis and bacterial cell culture.

5.5.1 Invasion assay optimisation

Assays were initially tested and optimised using wild-type LF82 as a positive (invasive) control, and XL1-blue as a negative (non-invasive) control. The assay was based upon that described by Boudeau et al. (1999), and steps were optimised as much as possible. Optimised steps included: bacterial growth phase, multiplicity of infection (MOI), HeLa cell density, time allowed for infection to occur, and the inclusion of a centrifugation step directly after bacterial application. Each alteration that resulted in an increase in LF82 invasion was noted and used to generate an invasion assay protocol (see: Invasion assays using HeLa CCL2). The addition of a 15 minute centrifugation step directly after the addition of bacterial cells to HeLa cells appeared to have the greatest impact upon LF82 invasion.

5.5.1.1 LF82 invasion after LF82_314 deletion and knockout of the T6SSs

Invasion assays were performed with Δp314 and ΔTssE mutants, using the optimised protocol. Results from these assays are expressed as a proportion of bacteria invaded (PI), which represents the number of bacteria invaded out of the total number of bacteria applied to each well. However, despite

the invasion assay optimisation performed, levels of invasion displayed by positive controls were far below typical values. For example, where Lapaquette et al. (2010) see a PI of ~4% when using wild-type LF82, an average PI of ~0.0005% was seen with wild-type LF82 here. Furthermore, Boudeau et al. (2003) describe similar percentages of invasion using wild-type LF82, with a PI of ~4.6-6%, when using Intestine-407, Caco-2 and human colon carcinoma (HCT-8) cell lines.

Due to the low invasion efficiency seen with wild-type LF82, more rounds of optimisation were attempted alongside invasion assays using the previously 'optimised' technique. Despite these attempts at optimisation, little improvement was seen in invasion. Results of these attempts are therefore presented in Figure 5.12.

From Figure 5.12, XL1-blue can be seen to have a lower PI than wild-type LF82. Perhaps more interestingly is the apparent lower PI of LF82 Δp314, suggesting that an invasion defect may be the result of p314 deletion. However, given the inefficiency of invasion seen with positive controls, and the variance seen in error bars, no conclusions can be made from this data.

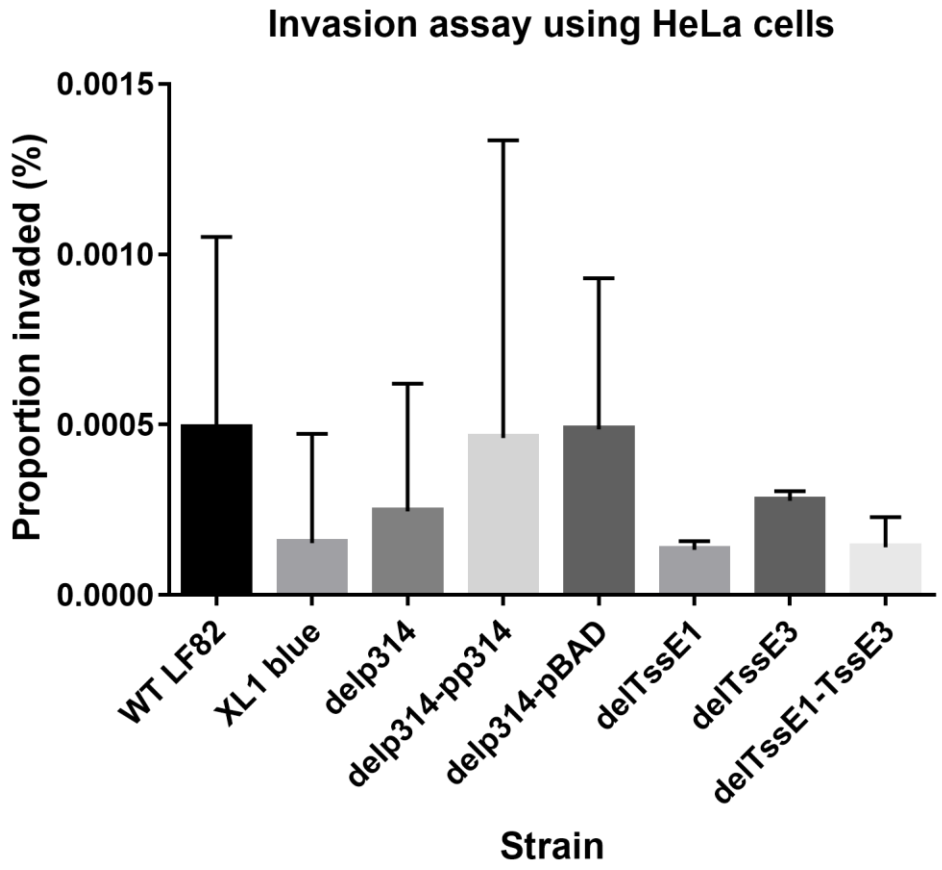


Figure 5.12. **Invasion assay for the analysis of p314 and TssE variants using HeLa cells**. Proportion of invasion can be seen to be far below that observed by others in the field (4-6%), making this data difficult to interpret with confidence. However, XL1-blue does appear to invade at a lower rate, as would be expected from this non-invasive strain.

The reason for such low invasion rates remains unknown, however, given the source of HeLa cells (the cervix), and the lack of their use in some other studies, it was thought that the eukaryotic cell line used may have been at fault. For example, Lapaquette et al. (2010) utilise mouse embryonic fibroblasts (MEFs) for their invasion assays, while Elliott et al. (2013) and Boudeau et al. (2003) utilise Caco-2 cells. Furthermore, it was later found that HeLa cells do not express the CEACAM6 receptor, which LF82 is known to utilise for invasion of host cells (Berger et al. 2004). Therefore Caco-2 cells were obtained from Dr. Chris Hale (Wellcome Trust Sanger Institute, Cambridge, UK) and invasion assays re-attempted.

5.5.1.2 Caco-2 invasion assay

Caco-2 cells are derived from epithelial tissue from colorectal adenocarcinomas and therefore are more representative of the tissues experienced by LF82 during CD-related activities. Furthermore, LF82 have been shown to induce CEACAM6 expression in Caco-2 cells, suggesting that they are superior to HeLa cells for invasion assays (Negroni et al. 2012).

Caco-2 cells can be treated similarly to HeLa cells, but require longer to grow, exhibit a lower tolerance to genetic manipulation, and are less evenly dispersed throughout the well/plate if seeded at too low a density. The invasion assay protocol was modified slightly, with bacterial growth performed in broth at 37 °C, but with no shaking. This was due to recent research within the lab which suggested that LF82 invasion is marginally increased when bacterial cultures are grown this way.

Invasion assays for LF82 Δp314 and all three LF82 ΔTssE variants were re-attempted using Caco-2 cells. However, due to time restrictions, this was only attempted once. Despite the use of a more suitable cell line, and a slightly more optimised protocol, the PI values obtained were still not comparable to those by others in the field, with wild-type LF82 and LF82 Δp314 infections resulting in a PI of ~0.07% and 0.02%, respectively. These results can be seen in Figure 5.13.

Despite the poor invasiveness of positive controls, LF82 Δp314 does appear to show lower levels of invasion than positive controls. However, until the issues surrounding positive control invasion rates are solved, little can be determined from this experiment.

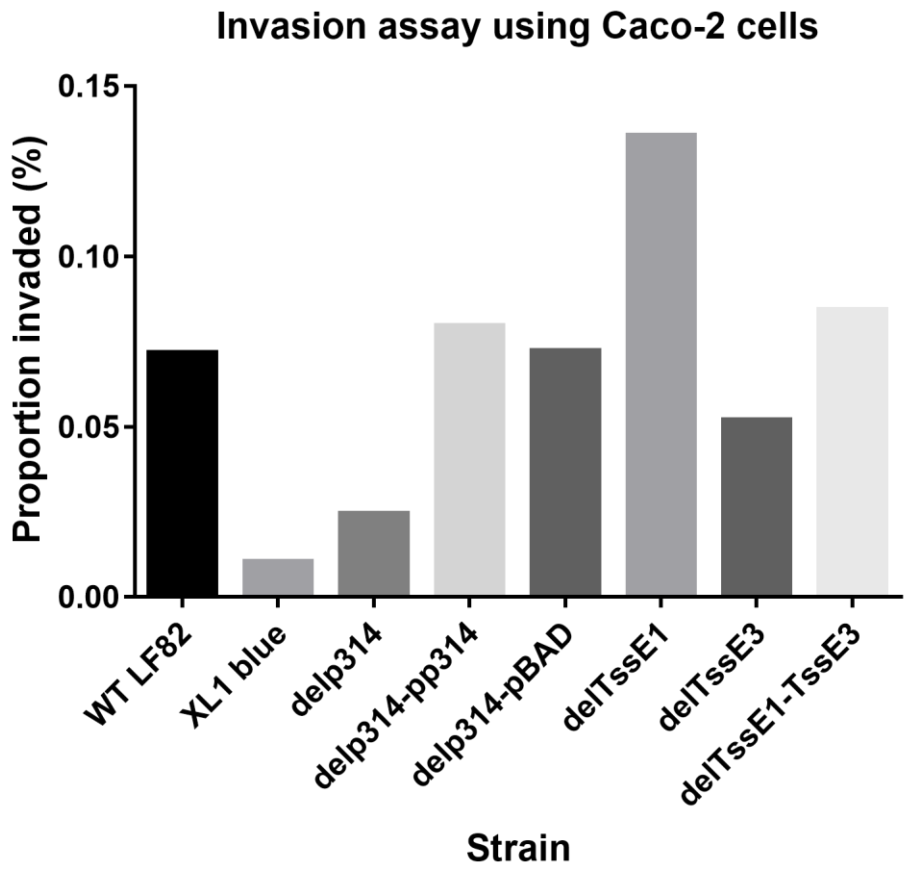


Figure 5.13. **Invasion assay for the analysis of p314 and TssE variants using Caco-2 cells**. Proportion of invasion, while an improvement over that seen with HeLa cells, is still far below that observed by others in the field (4-6%). As seen with HeLa cells, XL1-blue appears to have difficulty invading, as expected, with p314 displaying a similarly low level of invasion, although conclusions cannot be made due to low levels of invasion and no repeat experiments. This experiment was only performed once due to time restrictions.

5.5.1.3 Invasion assay summary

While results of both the HeLa and Caco-2 invasion assays may suggest an invasion phenotype related to p314 deletion, these results cannot be used to confirm this. Furthermore, contrasting results were shown by TssE1 deletion mutants, in which LF82 ΔTssE1 appears to reduce invasion for HeLa cells, but increase it for Caco-2 cells. This also suggests that the invasion assays are not performing as desired. Further optimisation of the invasion assay protocol is therefore essential before any conclusions can be made about the effects of p314 and/or TssE1/2 deletion upon LF82 invasion.

5.5.2 Reverse transcriptase PCR (RT-PCR)

RT-PCR was performed with the assistance of Dr. Carey Lambert (University of Nottingham, UK) in an attempt to identify whether or not p314 was expressed during a particular bacterial growth phase. If, for example, p314 expression was found to be downregulated during exponential phase, an insight into gene function may be possible. The involvement of gene regulation during a particular growth phase and the ability of a bacterium to invade mammalian cells has been documented previously, for example, with *Salmonella enterica* serovar Typhimurium (*S. enterica*), in which genes located on the *Salmonella* pathogenicity island (SPI) 1 are upregulated during late exponential phase. The upregulation of this PAI results in a significant increase in *S. enterica* invasion of eukaryotic cells (Ibarra et al. 2010).

5.5.2.1 RT-PCR of p314 expression under different growth conditions

Wild-type LF82 was cultured in several different conditions to identify any change in expression depending on several factors, including: culture technique, culture medium, and growth phase. The differing growth conditions consisted of LB broth at exponential and stationary phase, DMEM cell culture medium at exponential phase, and resuspended bacteria from an overnight culture on LB agar.

RNA was purified from each sample, and its integrity checked using the Agilent bioanalyser. Primers for the test RT-PCR were designed to amplify a 80 bp region of the p314 CDS (p314_RT_F and p314_RT_R), while a positive control RT-PCR was performed using primers designed to amplify a 140 bp region of the β-subunit of the bacterial RNA polymerase CDS, RpoB (rpoB_RT_F and rpoB_RT_R). Primers used can be seen in Table 2.4 (Oligonucleotides).

Results for this RT-PCR can be seen in Figure 5.14, in which PCRs are separated based upon which primer was used (p314 or RpoB). Unfortunately,

the RT-PCR did not reveal any information about p314 expression, with no specific bands detected in any of the test samples. While genomic positive controls did result in a product of the correct size, amplification was not as successful as desired, with even these bands appearing faint. The RpoB RT-PCR was also unsuccessful. These results suggest that optimisation of the RT-PCR is required, possibly via the adjustment of the Tm values used. However, due to time constraints, this RT-PCR was not re-attempted.

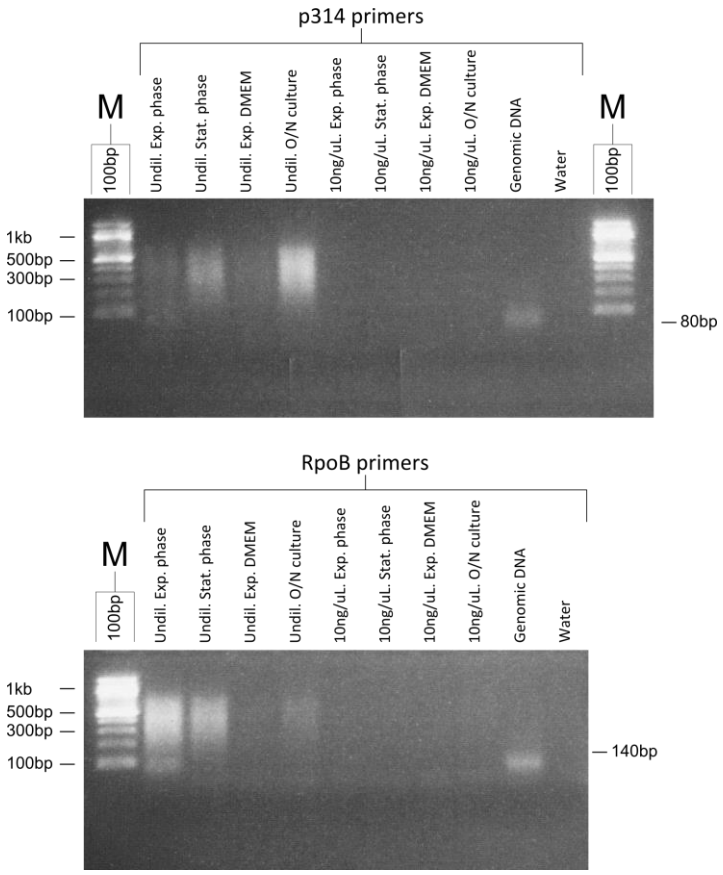


Figure 5.14. **Results from RT-PCR**. An RT-PCR performed using p314 primers was unsuccessful (top panel), with only the genomic positive control resulting in a band at ~80 bp which represents p314 amplification. RpoB primers were used as positive controls (bottom panel), with bands expected to be present at ~140 bp. However, bands were not seen at the expected size, suggesting that this RT-PCR also failed. Each lane is labelled with its contents: Undil.: Undiluted RNA sample, Exp.: Exponential, Stat.: Stationary, O/N: Overnight.

5.6 Future perspectives

Further characterisation of *LF82_314* is necessary to ascertain its role, if any, in LF82 adherence/invasion. Initial experiments were performed here to begin determining localisation of p314, but were not completed due to time restrictions and issues with microtubule detection. In order to establish the role (or lack thereof) of p314 and the removal of the T6SSs/s, invasion assays were utilised. However, due the inefficiency of wild-type control strains at invading, it was difficult to determine the invasion efficiency of LF82 Δ*TssE1,* LF82 Δ*TssE3,* LF82 Δ*TssE1*-Δ*TssE3* or LF82 Δp314 in comparison to wild-type LF82. Subsequent testing of Caco-2 cells for invasion assays did not rectify these issues, and will be continued by other researchers in the lab in the future.

Future experiments should look to replicate the proportion of bacterial invasion of ~4-6%, as reported by others in the field (Lapaquette et al. 2010; Boudeau et al. 2003). This will then either confirm or deny the role of p314 in LF82 invasion. Furthermore, adhesion assays may also be considered to determine whether p314 of the T6SS/s have any effect upon LF82 adherence to target cells. It was anecdotally observed that LF82 Δp314 mutants appeared to drop out of suspension when grown in broth, perhaps suggesting a role of p314 in bacteria-bacteria communication and possibly biofilm formation. It may therefore be of interest to characterise this phenotype fully.

For the determination of p314 localisation, a method for the detection of microtubules is desirable. While the reasoning for this is currently speculative, the similarity in phenotypes observed between p314 expression and microtubule effectors has led to the hypothesis that microtubules may be intimately involved in the otherwise uncommon phenotype observed during p314 expression. A suitable anti-tubulin antibody is therefore desired, in particular one which permits cell fixing using 4% formalin to align with current lab practices and techniques. The use of reagents such as Tubulin

Tracker (ThermoFisher) may also be considered, and may provide a suitable alternative assay for co-localisation with p314.

The analysis of GFP-p314 via a GFP pulldown experiment should be considered for both determination of interacting proteins and/or the identification of microtubules as interactors. Systems such as the GFP-Trap (ChromoTek) may be of use for such an assay. Evaluation of p314 secretion may also be of high interest, given the lack of any T3SS in LF82, and initial stages of secretion analysis via the split-GFP system have shown positive results (see Detection of p314 secretion using a split-GFP system). Following on from this, expression information obtained via RT-PCR may provide further clues as to what role p314 plays in LF82 invasion or intracellular lifestyle, if any.

While it is hypothesised that wild-type LF82 experienced difficulty invading HeLa cells due to a lack of HeLa CEACAM6 expression, it is not clear why Caco-2 cells also displayed such poor invasion. Two independent researchers within the lab attempted invasion assays using Caco-2 cells and also saw variable invasion rates, with PI values ranging from <1% to ~4%. It is currently believed that Caco-2 cell differentiation may play a role in bacterial invasion, with cells possibly requiring confluence in order to polarise and be efficiently invaded. The growth phase of bacteria during invasion may also play a role, as is seen with *Salmonella* in which certain growth phases favour invasion. Caco-2 cell differentiation and LF82 growth phase should therefore be considered for the future development of this system. However, despite the issues seen with efficiency of bacterial invasion in these experiments, it is still notable that LF82 Δp314 shows a decreased ability to invade eukaryotic cells.

5.7 Summary

This chapter has shown that hits from a screen such as that described in Chapter 4 can be analysed using a multitude of downstream applications. Although no conclusive evidence was displayed for p314 localisation, or for the effects of p314 or T6SS mutants on LF82 adherence/invasion, the scene has been set for this to be determined in the future. Furthermore, many hits from the screen remain unexplored and may also help to determine the methods via which LF82 adheres to and invades host cells.

Chapter 6: Conclusions and future perspectives

6.1 Research overview and discussion

Crohn's disease (CD) is a form of inflammatory bowel disease which is of increasing concern due to its rise in incidence in Western countries, and its unknown aetiology. Many factors have been associated with CD, and can broadly be divided into three categories: environmental (diet, smoking etc.), genetic (mutations in genes associated with autophagy, or cell surface receptors etc.), or microbial. The latter originally gained much attention due to similarities drawn between Johne's disease (JD) and CD, with *Mycobacterium avium* subspecies *paratuberculosis* (MAP) shown to be responsible for JD. However, where MAP has shown an inconsistent and controversial association with CD, AIEC has been strongly implicated in disease progression (Manuc et al. 2016; Dalton et al. 2014). AIEC type-stain LF82 was therefore utilised here as a representative of AIEC for the creation of a high-content screen to identify putative effectors. The findings from this research will lead to a greater understanding of AIEC, and may shed light on the role of this organism in CD.

6.1.1.1 The identification of numerous putative effectors in LF82

AIEC type-strain LF82 was compared with several other *E. coli* strains (pathogenic and commensal) using InParanoid. Through this, CDSs were identified which did not appear to be orthologous to CDSs found in comparison strains. This method showed that LF82 contained 235 CDSs that could be considered 'unique' from other members of the comparison set.

LF82 'unique' genes were mapped onto the LF82 genome, with a majority appearing to be present in regions of genomic plasticity. As these regions are often associated with the presence of virulence factors (in so-called pathogenicity islands), this may suggest that a number of these LF82 genes play a role in bacterial virulence. The CDSs of these 235 'unique' genes were therefore selected for analysis via the development of a high-content screen.

6.1.1.2 The use of the Seamless Ligation in Cell Extract (SLiCE) technique in a high content setting

Seamless Ligation in Cell Extract (SLiCE) was used to clone each LF82 'unique' gene into a destination vector. The combination of this approach with high-content bacterial transformation resulted in the production of a library of mammalian expression vectors for the expression of each CDS fused to GFP. Although this method did not result in 100% of the CDSs being cloned into expression vectors, it was successful for a majority of CDSs (89%).

Optimisation for the use of these techniques in a high-content setting was time consuming. However, once optimised, this system proved itself as a quick and cheap method for the creation of a library of customised expression vectors. The major costs associated with this system were due to oligonucleotide synthesis, although via the use of a two-step PCR strategy, oligonucleotides were comparable in both length and expense to those used in commercial systems such as Gateway cloning. Overall SLiCE was substantially cheaper than alternatives due to the lack of proprietary reagents or enzymes, and omitted the entry vector step associated with most Gateway systems.

6.1.1.3 High-content screening for the analysis of a range of cell phenotypes

The expression vector library was transfected and expressed in a HeLa cell line, designed for the analysis of the effects of CDS expression upon autophagy via an mCherry-LC3 reporter. Transfections were performed using a high-content approach, and cells were subsequently stained for visualisation of the nucleus and actin cytoskeleton. Cells were then subjected to automated microscopy, allowing the capture of numerous sites within each well, with little-to-no human interaction.

A CellProfiler pipeline was developed for the analysis of microscopy images. While this pipeline required over 24 h to run to competition (depending on

available hardware, and the size of the input data), it allowed for the unbiased analysis of hundreds of thousands of images in a relatively short period of time. This pipeline may be easily modified in the future for use in alternative experiments and screens.

Screen data was exported as a database containing measurement data from thousands of samples and millions of individual cells. To analyse this data, measurements of interest were selected relating to GFP localisation, nuclei shape and size, actin correlation etc. and plotted using boxplots. Statistical analysis for each measurement was performed, and was used to plot a marker on each graph representing significance. Boxplots were the chosen method of data presentation as they allowed an 'at a glance' overview of the data, and, alongside indication of significance, the subsequent identification of potential screen hits.

Despite the use of high-content analysis software, manual scrutiny was still employed to confirm phenotypes of suspected hits. The screen was therefore useful for indicating the presence of a phenotype of interest, but was not capable of fully characterising each hit. Future modifications to the image analysis pipeline and software may increase this ability. For example, the implementation of strategies used in a recent publication by Bray et al. (2016) (the developers of CellProfiler) during their so-called 'cell painting' technique, may allow for a more detailed morphological analysis of the existing image dataset.

Due to the fact that most hit identification was performed via the use of GFP distribution within an individual cell or nucleus, it could be argued that manual analysis of screen data may have been just as successful at identifying hits. However, software-based analysis of screen data has been shown to outperform humans in the ability to identify some patterns (Huang & Murphy 2004). An example of this in the screen reported here was the measurement of cell size, or nuclear circumference, which were difficult to scrutinise by

eye. Overall, future development of the image processing pipeline will be required, but a solid foundation has been created here which may be used for larger screens, different cell types, and different CDSs.

6.1.1.4 Autophagosomes detection and measurement

The measurement of autophagosomes – as well as their enumeration – was initially intended to be a phenotype of high interest for analysis due to the association of autophagy with CD, and the apparent ability of LF82 to circumvent autophagy. However, it appeared that CellProfiler was not able to differentiate foci of autophagosomes from background noise, meaning that autophagosomes were often not detected, despite their presence. While information regarding LF82 interactions with autophagosomes was therefore not available, the vast amount of data obtained from other cellular objects was sufficient for the identification of putative effectors. Future techniques using higher magnifications, higher resolution microscopy, or alternative image processing techniques may provide an alternative method for the successful high-content screening of autophagic effectors.

6.1.1.5 The development of a working mutagenesis system in AIEC

Three systems for the genetic manipulation of AIEC strain LF82 were tested, resulting in the development of a method for the mutagenesis of LF82. This method utilised the no-SCAR system developed by Reisch & Prather (2015) and resulted in successful deletion of targeted sites within the LF82 genome. However, due to the nature of the no-SCAR system, this mutagenesis method is not limited to gene deletion and can also be utilised for the introduction of point mutations or new coding regions, or can be used for the deletion of larger regions of the genome (e.g. the deletion of entire PAIs). Through this method, future research may, for example, introduce GFP under control of a promoter of either of the T6SSs to monitor its activity during invasion.

6.1.1.6 The creation of an LF82 T6SS mutant

The AIEC mutagenesis system was used to create LF82 mutants lacking each T6SS encoded on PAI-1 and PAI-3 (LF82 Δ*TssE1* and LF82 Δ*TssE3*, respectively), and also to create a double T6SS mutant lacking both T6SSs (LF82 Δ*TssE1*-Δ*TssE3*). These mutants were created to help determine the role of the T6SS in LF82 pathogenicity, invasion, and effector delivery. They were used in preliminary experiments to analyse the effects of individual or double T6SS knockouts on LF82 invasion. Future experiments may involve the use of these mutants to determine the status of putative T6SS-secreted effectors, or as control strains in assays determining the function or roles of Hcp, VgrG or other components of the T6SS.

6.1.1.7 LF82_314 may be an effector used by LF82

A number of potential effectors were identified via the screen, with LF82_314 selected as a strong candidate for effector activity and taken forward for the preliminary stages of effector characterisation. p314 displayed a phenotype similar to that seen with microtubule stabilising agents (see Figure 5.6, earlier), which may be of interest due to the known enrichment of some autophagy-related proteins at microtubule sites – including LC3 – and the known role of microtubules in autophagy (Mackeh et al. 2013). Furthermore, research by Domingues et al. (2014) has shown that microtubules may play a role in the regulation and control of *Salmonella* containing vacuoles (SCVs) in *S.* Typhimurium. Attempts were therefore made to characterise p314 and determine its status as an effector using a variety of methods.

Confirmation of the phenotype associated with p314 expression in the screen was achieved via both the transfection of pCMV-mEm-4GS-p314 into HeLa cells, and the sequencing of this vector. Subsequent experiments were then performed to begin to determine the localisation of p314 via co-localisation assays, the expression profile of p314 during various growth phases, and the role of p314, if any, in LF82 invasion of mammalian cells.

While considerable effort was put into these characterisation assays, it was not possible to confirm or deny whether p314 is indeed an LF82 effector. However, these assays – along with a split-GFP assay developed by other lab members – will provide a base for the characterisation of p314 and other putative effectors identified from the screen during future lab research.

6.1.1.8 Detection of p314 secretion using a split-GFP system

Research into the secretion of p314 was continued by Dr. Alan Huett and Rob Ward via the implementation of the split-GFP method described by Kaddoum et al. (2010) and Cabantous et al. (2005).

This method utilised the expression of a protein of interest fused to a GFP-11 tag (15 amino acids), and a lentiviral-transduced HeLa cell line expressing GFP 1-10. Both GFP-11 and GFP1-10 are non-fluorescent in isolation, but are able to auto-assemble to form fluorescent GFP when in the presence of one another. The invasion of HeLa cells expressing GFP1-10, by LF82 expressing a protein fused to GFP-11, therefore results in GFP fluorescence if the candidate protein is secreted into a HeLa-GFP1-10 cell. An overview of this process can be seen in Figure 6.1, below.

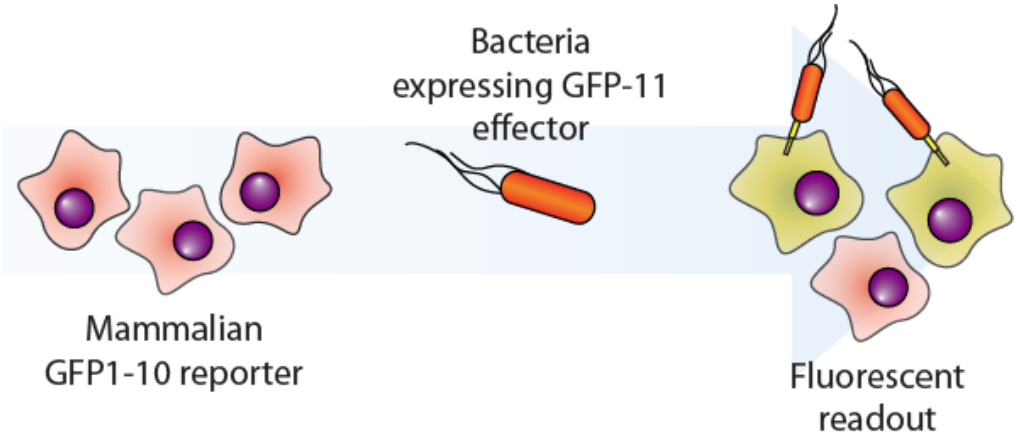


Figure 6.1. **An overview of the split-GFP secretion assay**. Mammalian cells expressing GFP1-10 are non-fluorescent. Bacterial cells expressing putative effectors fused to GFP-11 are introduced to the GFP1-10 reporter cell line. If effectors are secreted into the reporter cells, fluorescence will result due to auto-assembly GFP. Figure courtesy of Dr. Alan Huett (University of Nottingham, UK).

The use of this system allowed the investigation of p314, Hcp (a component of the T6SS needle structure), and VgrG (a structure on the tip of the T6SS needle). Figure 6.2 shows microscopy images from initial split-GFP experiments. Uninduced Hcp-GFP-11 can be seen to result in no GFP fluorescence, although some bleed-through from the Hoechst channel is evident. Induced Hcp-GFP-11 results in a focus within the HeLa cells, suggesting that Hcp is successfully secreted into target cells. Likewise, VgrG-GFP-11 induction results in foci within cells, indicative of successful secretion. The bottom panel of Figure 6.2 displays the p314-GFP-11 secretion assay, where p314 expression appears to result in the familiar 'worm-like' phenotype associated with its expression. This suggests that p314 is secreted, although further evidence is required to support this claim, and further research necessary to determine which secretion system is utilised for p314 secretion.

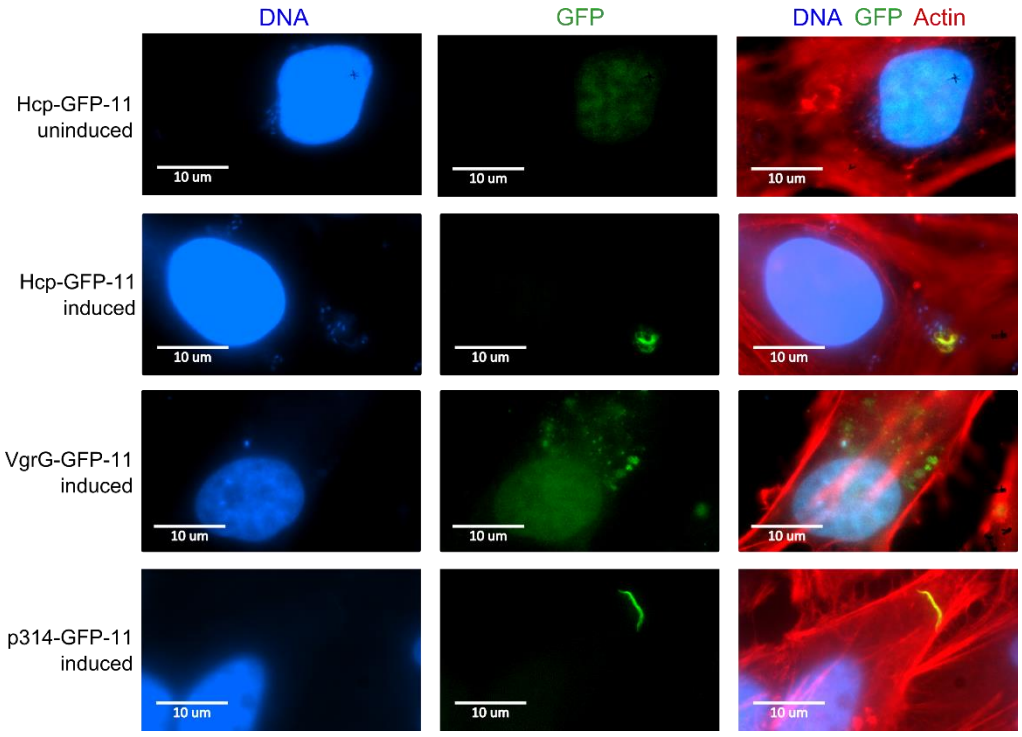


Figure 6.2. **Split-GFP assays, performed by Dr. Alan Huett.** LF82 expressing C-terminal GFP-11 fused proteins were used to infect HeLa cells expressing GFP1-10. Top panel: Negative control in which LF82 Hcp-GFP-11 expression was uninduced and therefore did not result in GFP fluorescence in HeLa cells. Panels two and three: Positive controls in which LF82 Hcp-GFP-11, and LF82 VgrG-GFP-11 expression was induced, respectively. Expression resulted in GFP fluorescence in HeLa cells. Bottom panel: LF82 p314-GFP-11 expression resulted in GFP expression in HeLa cells. Fluorescence took the form of the 'worm-like' phenotype commonly observed with p314 expression. This suggests that p314 is secreted into HeLa cells by LF82. Figure and unpublished data courtesy of Dr. Alan Huett (University of Nottingham, UK).

6.1.2 Future work

One of the most obvious questions to ask after this work is whether or not p314 is an effector. While efforts were focussed on answering this question, answers were not forthcoming. Optimisation of invasion assays would therefore be of primary interest, as an effect upon LF82 invasion of eukaryotic cells may suggest a role for p314 in invasion, or a role in the evasion of host cell autophagy. Alternatively, if p314 is later shown not to be secreted into eukaryotic cells, any effect upon invasion may help to understand the mechanics of LF82 invasion anyway. Information regarding

the translocation of p314 into eukaryotic cells is therefore also desired, and initial experiments to determine this are underway using split-GFP.

Co-localisation experiments are of high interest, especially given the strength of phenotype displayed by p314 and the potential for p314 association with microtubules. A GFP-trap based experiment, to capture interacting proteins and identify them via mass-spectrometry (similar to that described by Pearson et al. (2013)) was considered for p314 towards the end of this research, but was deemed to require more time than was available. This experiment would allow an overview of proteins which p314 directly interacts with, and may be considered as an alternative to co-localisation assays in the future.

The characterisation of putative effectors from the screen, other than p314, may be of high interest. For example, LF82_1175, which was identified as the only CDS in the screen to potentially affect autophagosomes, may play a role in LF82 manipulation of autophagy. Alternatively, the CDSs appearing to localise to nuclear dots may be investigated, given the apparent linkage of DNA damage to effector delivery via OMVs (Tyrer et al. 2014). For the investigation of these CDSs, more assays may require development. For example, a more sensitive assay for the detection and/or measurement of autophagosomes may be useful in characterising LF82_1175.

It would be interesting to see how the screen as a whole performed using different datasets. For example, alternative organisms may be subjected to similar InParanoid comparisons, and their resulting 'unique' gene set screened for the presence of putative effectors. An increased number of repeats may also be interesting to analyse using this system, as this may result in clearer results on data analysis graphs, making hit selection easier.

The LF82 T6SS mutant strains created here could be utilised for the future assessment of each LF82-encoded T6SS in bacteria-bacteria competition, or

for the investigation of the activity of each T6SS in different stages of bacterial infection. Similar research has been conducted in *S.* Typhimurium, which has been shown to employ a SPI-6 encoded T6SS in inter-bacterial competition by Sana et al. (2016), and has been shown to utilise the same T6SS in later stages of macrophage invasion by Mulder et al. (2012). Furthermore, the LF82 T6SS mutants developed in this study may be suited to an antibacterial competition assay to determine the use of the T6SS in inter-bacterial survival. These assays, described by Murdoch et al. (2011), involve the mixing of a bacterial strain of interest with a bacterial challenger. Cells are then allowed to mix and interact and are later cultured, before the enumeration of recovered colonies. This assay would allow the evaluation of each (or both) LF82 T6SS for its role in antibacterial activity against specific bacteria.

6.1.3 Summary

This project resulted in the identification of 235 CDSs of interest from the CD-associated bacterium, LF82. These CDSs were expressed in eukaryotic cells and, through the use of a high-content screen, their resulting phenotypes were analysed. From this screen, multiple CDSs were identified which appeared to result in interesting phenotypes, although the identification of CDSs affecting autophagy was not possible. LF82_314 was selected as a CDS of particular interest, and initial characterisation assays were performed with the aim of determining its status as an effector of LF82. In the future, LF82_314 will be subjected to in-depth analysis by other lab members to fully characterise its role.

This study sets the scene for future research of a similar nature using high-content screening, as well as for the selection and analysis of putative effectors of LF82. The development of a system for the mutagenic manipulation of LF82 also provides a method for the alteration of LF82 for uses in future experimental studies. The discovery of novel virulence factors

from LF82 will aid understanding of this organism in CD, and may provide clues for its ability to manipulate host cell biology.

References

Abounit, K., Scarabelli, T.M. & McCauley, R.B., 2012. Autophagy in mammalian cells. World journal of biological chemistry, 3(1), pp.1–6.

Abraham, V.C., Taylor, D.L. & Haskins, J.R., 2004. High content screening applied to large-scale cell biology. Trends in Biotechnology, 22(1), pp.15–22.

Adolph, T.E. et al., 2013. Paneth cells as a site of origin for intestinal inflammation. Nature, 503(7475), pp.272–6.

Agus, A. et al., 2014. Understanding Host-Adherent-Invasive *Escherichia coli* Interaction in Crohn's Disease: Opening Up New Therapeutic Strategies. BioMed Research International, 2014, pp.1–16.

Akopyan, K. et al., 2011. Translocation of surface-localized effectors in type III secretion. Proceedings of the National Academy of Sciences of the United States of America, 108(4), pp.1639–1644.

Alexis, B. et al., 2016. Activation of the EIF2AK4-EIF2A/eIF2α-ATF4 pathway triggers autophagy response to Crohn disease-associated adherent-invasive *Escherichia coli* infection. Autophagy, 12(5), pp.770-83.

Altindis, E. et al., 2015. Secretome analysis of *Vibrio cholerae* type VI secretion system reveals a new effector-immunity pair. mBio, 6(2), p.e00075.

Amezaga, A.J. & Van Assche, G., 2016. Practical approaches to "top–down" therapies for Crohn's disease. Current Gastroenterology Reports, 18(7), p.35.

Amos, W.B., 1993. Light microscopy. Trends in Cell Biology, 3(1), p.28.

Ananthakrishnan, A.N., 2015. Epidemiology and risk factors for IBD. Nature reviews. Gastroenterology & hepatology, 12(April), pp.205–17.

van Assche, G. et al., 2010. The second European evidence-based consensus on the diagnosis and management of Crohn's disease: Special situations. Journal of Crohn's and Colitis, 4(1), pp.63–101.

Barnich, N. et al., 2007. CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. The Journal of clinical investigation, 117(6), pp.1566–74.

Barnich, N. & Darfeuille-Michaud, A., 2007. Adherent-invasive *Escherichia coli* and Crohn's disease. Current opinion in gastroenterology, 23, pp.16–20.

Basler, M. et al., 2012. Type VI secretion requires a dynamic contractile phage tail-like structure. Nature, 483(7388), pp.182–6.

Bauckman, K.A., Owusu-Boaitey, N. & Mysorekar, I.U., 2015. Selective autophagy: Xenophagy. Methods, 75, pp.120–7.

Behrends, C. et al., 2010. Network organization of the human autophagy system. Nature, 466(7302), pp.68–76.

Bendtsen, J.D. et al., 2005. Non-classical protein secretion in bacteria. BMC microbiology, 5, p.58.

Benz, J. et al., 2012. Structural insights into the effector - immunity system Tse1/Tsi1 from *Pseudomonas aeruginosa*. PLoS ONE, 7(7), p.e40453.

Berger, C.N. et al., 2004. Differential recognition of members of the carcinoembryonic antigen family by Afa/Dr adhesins of diffusely adhering *Escherichia coli* (Afa/Dr DAEC). Molecular Microbiology, 52(4), pp.963–83.

Bhullar, M. et al., 2014. Prediction of Crohn's disease aggression through NOD2 /CARD15 gene sequencing in an Australian cohort. World Journal of Gastroenterology, 20(17), pp.5008–16.

Bjørkøy, G. et al., 2005. p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. The Journal of cell biology, 171(4), pp.603–14.

Bomberger, J.M. et al., 2009. Long-distance delivery of bacterial virulence factors by *Pseudomonas aeruginosa* outer membrane vesicles. PLoS Pathogens, 5(4), p.e1000382.

Bomberger, J.M. et al., 2014. *Pseudomonas aeruginosa* Cif protein enhances the ubiquitination and proteasomal degradation of the transporter associated with antigen processing (TAP) and reduces major histocompatibility complex (MHC) class I antigen presentation. The Journal of biological chemistry, 289(1), pp.152–62.

Bosca-Watts, M.M., 2015. Pathogenesis of Crohn's disease: Bug or no bug. World Journal of Gastrointestinal Pathophysiology, 6(1), p.1.

Boudeau, J. et al., 2003. Inhibitory effect of probiotic *Escherichia coli* strain Nissle 1917 on adhesion to and invasion of intestinal epithelial cells by adherent-invasive *E. coli* strains isolated from patients with Crohn's disease. Alimentary pharmacology & therapeutics, 18(1), pp.45–56.

Boudeau, J. et al., 1999. Invasive Ability of an *Escherichia coli* Strain Isolated from the Ileal Mucosa of a Patient with Crohn's Disease. Infection and Immunity, 67(9), pp.4499–509.

Bowater, R. & Doherty, A.J., 2006. Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genetics, 2(2), pp.93–9.

Brass, A.L. et al., 2008. Identification of host proteins required for HIV infection through a functional genomic screen. Science, 319(5865), pp.921–26.

Bray, M.A. et al., 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nature protocols, 11(6), pp.1757–74.

Bray, M.A. & Carpenter, A., 2012. Advanced Assay Development Guidelines for Image-Based High Content Screening and Analysis. Assay Guidance Manual, pp.1–33.

Breslauer, K.J. et al., 1986. Predicting DNA duplex stability from the base sequence. Proceedings of the National Academy of Sciences of the United States of America, 83(11), pp.3746–50.

Bringer, M.A. et al., 2012. Replication of Crohn's disease-associated AIEC within macrophages is dependent on TNF-α secretion. Laboratory investigation; a journal of technical methods and pathology, 92(3), pp.411–9.

Brito, I.L. & Alm, E.J., 2016. Tracking strains in the microbiome: insights from metagenomics and models. Frontiers in Microbiology, 7(May), pp.1–8.

Brodin, P. et al., 2010. High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathogens, 6(9), p.e1001100.

Bröms, J.E. et al., 2012. DotU and VgrG, core components of type VI secretion systems, are essential for *Francisella* LVS pathogenicity. PloS one, 7(4), p.e34639.

Brunet, Y.R. et al., 2015. The Type VI Secretion TssEFGK-VgrG phage-like baseplate is recruited to the TssJLM membrane complex via multiple contacts and serves as assembly platform for tail tube/sheath polymerization. PLoS Genetics, 11(10), pp.1–21.

Büttner, D., 2012. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and

animal-pathogenic bacteria. Microbiology and molecular biology reviews, 76(2), pp.262–310.

Cabantous, S., Terwilliger, T.C. & Waldo, G.S., 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature biotechnology, 23(1), pp.102–7.

Carlsson, S.R. & Simonsen, A., 2015. Membrane dynamics in autophagosome biogenesis. Journal of cell science, 128(2), pp.1–13.

Carpenter, A.E. et al., 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome biology, 7(10), p.R100.

Carvalho, F. et al., 2009. Crohn's disease adherent-invasive *Escherichia coli* colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. The Journal of experimental medicine, 206(10), pp.2179–89.

Carver, T. et al., 2012. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics, 28(4), pp.464–69.

Cernada, M. et al., 2016. Sepsis in preterm infants causes alterations in mucosal gene expression and microbiota profiles compared to non-septic twins. Scientific Reports, 6(April), p.25497.

Chargui, A. et al., 2012. Subversion of autophagy in adherent invasive *Escherichia coli*-infected neutrophils induces inflammation and cell death. PloS one, 7(12), p.e51727.

Chassaing, B. et al., 2011. Crohn disease–associated adherent-invasive *E. coli* bacteria target mouse and human Peyer's patches via long polar fimbriae. Journal of clinical investigation, 121(3) pp.966–75.

Chen, W.P. & Kuo, T.T., 1993. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic acids research, 21(9), p.2260.

Chen, X., Zaro, J.L. & Shen, W.C., 2013. Fusion protein linkers: Property, design and functionality. Advanced Drug Delivery Reviews, 65(10), pp.1357–69.

Christophe, T. et al., 2009. High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. PLoS pathogens, 5(10), p.e1000645.

Cianfanelli, F.R., Monlezun, L. & Coulthurst, S.J., 2015. Aim, load, fire: The Type VI Secretion System, a bacterial nanoweapon. Trends in Microbiology, 24(1), pp.51–62.

Cieza, R.J. et al., 2015. The IbeA invasin of adherent-invasive *Escherichia coli* mediates interaction with intestinal epithelia and macrophages. Infection and immunity, 83(5), pp.1904–18.

Cornelis, G.R. et al., 1998. The virulence plasmid of *Yersinia*, an antihost genome. Microbiology and molecular biology reviews, 62(4), pp.1315–52.

Costa, T.R.D. et al., 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nature Reviews Microbiology, 13(6), pp.343–59.

Crohn, B., Ginzburg, L. & Oppenheimer, G.D., 1952. Regional ileitis; a pathologic and clinical entity. The American journal of medicine, 13(5), pp.583–90.

Croxen, M. a & Finlay, B.B., 2010. Molecular mechanisms of *Escherichia coli* pathogenicity. Nature reviews: Microbiology, 8(1), pp.26–38.

Cuervo, A.M. & Macian, F., 2012. Autophagy, nutrition and immunology. Molecular Aspects of Medicine, 33(1), pp.2–13.

Dalton, J.P. et al., 2014. Detection of *Mycobacterium avium* subspecies paratuberculosis in patients with Crohn's disease is unrelated to the presence of single nucleotide polymorphisms rs2241880 (ATG16L1) and rs10045431 (IL12B). Medical microbiology and immunology, 203(3), pp.195–205.

Darfeuille-Michaud, A. et al., 2004. High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. Gastroenterology, 127(2), pp.412–21.

Datsenko, K. a & Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America, 97(12), pp.6640–5.

van Dijk, A.P. et al., 1998. Transdermal nicotine inhibits interleukin 2 synthesis by mononuclear cells derived from healthy volunteers. Eur J Clin Invest, 28(8), pp.664–71.

Domingues, L., Holden, D.W. & Mota, L.J., 2014. The *Salmonella* effector SteA contributes to the control of membrane dynamics of *Salmonella*-containing vacuoles. Infection and Immunity, 82(7), pp.2923–34.

Dooley, H.C. et al., 2014. WIPI2 links LC3 conjugation with PI3P, autophagosome formation, and pathogen clearance by recruiting ATG12-5-16L1. Molecular Cell, 55(2), pp.238–52.

Doudna, J.A. & Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213), pp.1258096–96.

Dubytska, L.P., Rogge, M.L. & Thune, R.L., 2016. Identification and characterization of putative translocated effector proteins of the *Edwardsiella ictaluri* Type III Secretion System. mSphere, 1(3), pp.e00039-16.

Eaves-Pyles, T. et al., 2008. *Escherichia coli* isolated from a Crohn's disease patient adheres, invades, and induces inflammatory responses in polarized intestinal epithelial cells. International Journal of Medical Microbiology, 298(5–6), pp.397–409.

Edgren, T. et al., 2012. Type III secretion in *Yersinia*: Injectisome or not? PLoS Pathogens, 8(5), pp.2–4.

Ek, W.E., D'Amato, M. & Halfvarson, J., 2014. The history of genetics in inflammatory bowel disease. Annals of gastroenterology : quarterly publication of the Hellenic Society of Gastroenterology, 27(4), pp.294–303.

Elliott, T.R. et al., 2015. Defective macrophage handling of *Escherichia coli* in Crohn's disease. Journal of gastroenterology and hepatology, 30(8), pp.1265–74.

Elliott, T.R. et al., 2013. Quantification and characterization of mucosa-associated and intracellular *Escherichia coli* in inflammatory bowel disease. Inflammatory bowel diseases, 19(11), pp.2326–38.

Ellis, H.M. et al., 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proceedings of the National Academy of Sciences of the United States of America, 98(12), pp.6742–6.

Ellis, T.N. & Kuehn, M.J., 2010. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. Microbiology and Molecular Biology Reviews, 74(1), pp.81–94.

Fakhoury, M. et al., 2014. Inflammatory bowel disease: clinical aspects and treatments. Journal of inflammation research, 7, pp.113–20.

Fengsrud, M. et al., 2000. Ultrastructural characterization of the delimiting membranes of isolated autophagosomes and amphisomes by freeze-

fracture electron microscopy. European journal of cell biology, 79(12), pp.871–82.

Figueira, R. & Holden, D.W., 2012. Functions of the *Salmonella* pathogenicity island 2 (SPI-2) type III secretion system effectors. Microbiology, 158(5), pp.1147–61.

Filloux, A., Hachani, A. & Bleves, S., 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. Microbiolog, 154(Pt 6), pp.1570–83.

Fox, J., 2005. The R Commander: A Basic-Statistics Graphical User Interface to R. Journal of Statistical Software, 14(9), pp.1–42.

Francis, M.S., Wolf-Watz, H. & Forsberg, Å., 2002. Regulation of type III secretion systems. Current Opinion in Microbiology, 5(2), pp.166–72.

Friedman, A. & Perrimon, N., 2006. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature, 444(7116), pp.230–4.

Fujita, N. et al., 2008. The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. Molecular biology of the cell, 19(5), pp.2092–100.

Galán, J.E. & Collmer, A., 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. Science, 284(5418), pp.1322–8.

Garcia, J.T. et al., 2006. Measurement of effector protein injection by type III and type IV secretion systems by using a 13-residue phosphorylatable glycogen synthase kinase tag. Infection and Immunity, 74(10), pp.5645–57.

Gardet, A. & Xavier, R.J., 2012. Common alleles that influence autophagy and the risk for inflammatory bowel disease. Current Opinion in Immunology, 24(5), pp.522–9.

Geng, J. et al., 2010. Post-Golgi sec proteins are required for autophagy in *Saccharomyces cerevisiae*. Molecular biology of the cell, 21, pp.2257–69.

Girardi, J.P., Pereira, L. & Bakovic, M., 2011. De novo synthesis of phospholipids is coupled with autophagosome formation. Medical Hypotheses, 77(6), pp.1083–7.

Girardin, S.E. et al., 2003. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. Journal of Biological Chemistry, 278(11), pp.8869–72.

Glasser, A.L. et al., 2001. Adherent invasive *Escherichia coli* strains from patients with Crohn's disease survive and replicate within macrophages

without inducing host cell death. Infection and immunity, 69(9), pp.5529–37.

Glick, D., Barth, S. & Macleod, K.F., 2010. Autophagy: cellular and molecular mechanisms. The Journal of pathology, 221(1), pp.3–12.

Goldfine, H. & Wadsworth, S.J., 2002. Macrophage intracellular signalling induced by *Listeria monocytogenes*. Microbes and Infection, 4(13), pp.1335–43.

Gould, S.J., Booth, A.M. & Hildreth, J.E.K., 2003. The Trojan exosome hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 100(19), pp.10592–7.

Govan, J.R. & Deretic, V., 1996. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. Microbiological reviews, 60(3), pp.539–74.

Grant, J.R. & Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic acids research, 36, pp.W181–4.

Green, E.R. & Mecsas, J., 2016. Bacterial secretion systems: an overview. Microbiology Spectrum, 4(1), pp.1–32.

Hachani, A., Wood, T.E. & Filloux, A., 2016. Type VI secretion and anti-host effectors. Current Opinion in Microbiology, 29, pp.81–93.

Hacker, J. et al., 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extra intestinal *Escherichia coli* isolates. Microbial Pathogenesis, 8(3), pp.213–25.

Hacker, J. & Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annual Review of Microbiology, 54(1), pp.641–79.

Hailey, D.W. et al., 2010. Mitochondria supply membranes for autophagosome biogenesis during starvation. Cell, 141(4), pp.656–67.

Hamasaki, M. et al., 2013. Autophagosomes form at ER-mitochondria contact sites. Nature, 495(7441), pp.389–93.

Hamilton, C.M. et al., 2002. Multichannel plating unit for high-throughput plating of cell cultures. BioTechniques, 33(2), pp.420–3.

Hampe, J. et al., 2001. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet, 357(9272), pp.1925–8.

Haney, S.A. et al., 2006. High-content screening moves to the front of the line. Drug Discovery Today, 11(19–20), pp.889–94.

Haraga, A., Ohlson, M.B. & Miller, S.I., 2008. *Salmonellae* interplay with host cells. Nature Reviews Microbiology, 6(1), pp.53–66.

Hartwell, K.A. et al., 2013. Niche-based screening identifies small-molecule inhibitors of leukemia stem cells. Nature Chemical Biology, 9(12), pp.840–8.

Hayee, B.H. et al., 2010. Crohn's disease as an immunodeficiency. Expert Review of Clinical Immunology, 6(4), pp.585–96.

Ho, B.T., Dong, T.G. & Mekalanos, J.J., 2014. A view to a kill: The bacterial type VI secretion system. Cell Host and Microbe, 15(1), pp.9–21.

Ho Sui, S.J. et al., 2009. The association of virulence factors with genomic Islands. PLoS ONE, 4(12) p.e8094.

Hood, R.D. et al., 2010. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. Cell host & microbe, 7(1), pp.25–37.

Huang, K. & Murphy, R.F., 2004. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. BMC bioinformatics, 5, p.78.

Huang, S., Bjornsti, M.A. & Houghton, P.J., 2003. Rapamycins: Mechanism of action and cellular resistance. Cancer Biology and Therapy, 2(3), pp.222–32.

Huett, A. & Xavier, R.J., 2010. Autophagy at the gut interface: Mucosal responses to stress and the consequences for inflammatory bowel diseases. Inflammatory Bowel Diseases, 16(1), pp.152–74.

Hugot, J. et al., 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature, 411(6837), pp.599–603.

Ibarra, J.A. et al., 2010. Induction of *Salmonella* pathogenicity island 1 under different growth conditions can affect *Salmonella*-host cell interactions in vitro. Microbiology, 156(Pt 4), pp.1120–33.

Ingrassia, I., Leplingard, A. & Darfeuille-michaud, A., 2005. *Lactobacillus casei* DN-114 001 inhibits the ability of Adherent-Invasive *Escherichia coli* isolated from Crohn's Disease patients to adhere to and to invade intestinal epithelial cells. Applied and environmental microbiology, 71(6), pp.2880–7.

Inohara, N. et al., 2005. NOD-LRR PROTEINS: role in host-microbial interactions and inflammatory disease. Annual Review of Biochemistry, 74(1), pp.355–83.

Ito, K. & Murphy, D., 2013. Application of ggplot2 to Pharmacometric Graphics. CPT: pharmacometrics & systems pharmacology, 2(August), p.e79.

Jani, A.J. & Cotter, P. a, 2010. Type VI secretion: not just for pathogenesis anymore. Cell host & microbe, 8(1), pp.2–6.

Jess, T., Frisch, M. & Simonsen, J., 2013. Trends in overall and cause-specific mortality among patients with inflammatory bowel disease from 1982 to 2010. Clinical Gastroenterology and Hepatology, 11(1), pp.43–8.

Jiang, F. et al., 2014. A *Pseudomonas aeruginosa* type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. Cell Host and Microbe, 15(5), pp.600–10.

Jin, M. et al., 2014. Transcriptional regulation by Pho23 modulates the frequency of autophagosome formation. Current Biology, 24(12), pp.1314–22.

Jin, M. & Klionsky, D.J., 2014. Regulation of autophagy: Modulation of the size and number of autophagosomes. FEBS Letters, 588(15), pp.2457–63.

Jones, T.R. et al., 2008. CellProfiler Analyst: data exploration and analysis software for complex image-based screens. BMC bioinformatics, 9, p.482.

Jones, T.R. et al., 2006. Methods for high-content, high-throughput image-based cell screening. Proceedings of the First MICCAI Workshop on Microscopic Image Analysis with Applications in Biology, pp.65–72.

Jostins, L. et al., 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 491(7422), pp.119–24.

Joung, I., Strominger, J.L. & Shin, J., 1996. Molecular cloning of a phosphotyrosine-independent ligand of the p56lck SH2 domain. Proceedings of the National Academy of Sciences of the United States of America, 93(12), pp.5991–5.

Jung, C., Hugot, J.-P. & Barreau, F., 2010. Peyer's patches: The immune sensors of the intestine. International journal of inflammation, 2010, p.823710.

Jung, C.H. et al., 2010. mTOR regulation of autophagy. FEBS Letters, 584(7), pp.1287–95.

Kabeya, Y. et al., 2004. LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. Journal of cell science, 117(Pt 13), pp.2805–12.

Kaddoum, L. et al., 2010. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. BioTechniques, 49(4), p.727–36.

Kang, J. et al., 2016. Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. Nat Biotech, 34(1), pp.70–7.

Kassam, Z. et al., 2014. Inflammatory Bowel Disease Cause-specific Mortality. Inflammatory Bowel Diseases, 20(12), pp.2483–92.

Kaufmann, A. et al., 2014. Molecular mechanism of autophagic membrane-scaffold assembly and disassembly. Cell, 156(3), pp.469–81.

Kaushik, S. & Cuervo, A.M., 2012. Chaperones in autophagy. Pharmacological Research, 66(6), pp.484–93.

Kayal, S. & Charbit, A., 2006. Listeriolysin O: A key protein of *Listeria monocytogenes* with multiple functions. FEMS Microbiology Reviews, 30(4), pp.514–29.

Keenan, J.I. et al., 2014. Influences of enteral nutrition upon CEACAM6 expression by intestinal epithelial cells. Innate immunity, 20(8), pp.848–56.

Kiel, J. a K.W., 2010. Autophagy in unicellular eukaryotes. Philosophical transactions of the Royal Society of London, 365(1541), pp.819–30.

Kim, P.K. et al., 2008. Ubiquitin signals autophagic degradation of cytosolic proteins and peroxisomes. Proceedings of the National Academy of Sciences of the United States of America, 105(52), pp.20567–74.

Klein, J.R. & Jones, B.D., 2001. *Salmonella* pathogenicity island 2-encoded proteins SseC and SseD are essential for virulence and are substrates of the type III secretion system. Infection and Immunity, 69(2), pp.737–43.

Klionsky, D.J., 2009. Crohn's disease, autophagy, and the Paneth cell. The New England journal of medicine, 360(17), pp.1785–6.

Klionsky, D.J., Eskelinen, E.L. & Deretic, V., 2014. Autophagosomes, phagosomes, autolysosomes, phagolysosomes, autophagolysosomes... Wait, I'm confused. Autophagy, 10(4), pp.549–51.

Knævelsrud, H. et al., 2013. Membrane remodeling by the PX-BAR protein SNX18 promotes autophagosome formation. The Journal of cell biology, 202(2), pp.331–49.

Koleva, P.T., Bridgman, S.L. & Kozyrskyj, A.L., 2015. The infant gut microbiome: Evidence for obesity risk and dietary intervention. Nutrients, 7(4), pp.2237–60.

Komatsu, M. et al., 2007. Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. Cell, 131(6), pp.1149–63.

Kraft, C. et al., 2008. Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. Nature Cell Biology, 10(5), pp.602–10.

Kuballa, P. et al., 2008. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated ATG16L1 variant. PloS one, 3(10), p.e3391.

Kunsmann, L. et al., 2015. Virulence from vesicles: Novel mechanisms of host cell injury by *Escherichia coli* O104:H4 outbreak strain. Scientific reports, 5(August), p.13252.

Lala, S. et al., 2003. Crohn's disease and the NOD2 gene: A role for paneth cells. Gastroenterology, 125(1), pp.47–57.

Lapaquette, P. et al., 2010. Crohn's disease-associated adherent-invasive *E. coli* are selectively favoured by impaired autophagy to replicate intracellularly. Cellular microbiology, 12(1), pp.99–113.

Lawrence, J.G. & Ochman, H., 1997. Amelioration of bacterial genomes: Rates of change and exchange. Journal of Molecular Evolution, 44(4), pp.383–97.

Lee, D.J. et al., 2009. Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. BMC microbiology, 9, p.252.

Lee, J.E. et al., 2007. Phylogenetic analysis of Shiga toxin 1 and Shiga toxin 2 genes associated with disease outbreaks. BMC microbiology, 7, p.109.

Levine, B., 2005. Eating oneself and uninvited guests. Cell, 120(2), pp.159–62.

Levine, B. & Kroemer, G., 2008. Autophagy in the pathogenesis of disease. Cell, 132(1), pp.27–42.

Lin, L. & Baehrecke, E.H., 2015. Autophagy, cell death, and cancer. Molecular & Cellular Oncology, 2(3), p.e985913.

Lippai, M. & Lőw, P., 2014. The role of the selective adaptor p62 and ubiquitin-like proteins in autophagy. BioMed research international, 2014, p.832704.

Liu, T., Sims, D. & Baum, B., 2009. Parallel RNAi screens across different cell lines identify generic and cell type-specific regulators of actin organization and cell morphology. Genome biology, 10(3), p.R26.

Loftus, E. V, 2004a. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology, 126(6), pp.1504–17.

Loftus, E. V, 2004b. Management of extraintestinal manifestations and other complications of inflammatory bowel disease. Current gastroenterology reports, 6(6), pp.506–13.

Lorkowski, M. et al., 2014. *Salmonella enterica* invasion of polarized epithelial cells is a highly cooperative effort. Infection and Immunity, 82(6), pp.2657–67.

Low, D. et al., 1984. Gene clusters governing the production of gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype 04 and 06 isolates from urinary tract infections. Infection and Immunity, 43(1), pp.353–8.

Luna, L. et al., 2005. Dynamic relocalization of hOGG1 during the cell cycle is disrupted in cells harbouring the hOGG1-Cys326 polymorphic variant. Nucleic Acids Research, 33(6), pp.1813–24.

Ma, A.T. & Mekalanos, J.J., 2010. In vivo actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. Proceedings of the National Academy of Sciences of the United States of America, 107(9), pp.4365–70.

Macarron, R. et al., 2011. Impact of high-throughput screening. Nature Reviews Drug Discovery, 10(3), pp.188-95.

MacIntyre, D.L. et al., 2010. The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. Proceedings of the National Academy of Sciences of the United States of America, 107(45), pp.19520–4.

Mackeh, R. et al., 2013. Autophagy and microtubules - new story, old players. Journal of cell science, 126(Pt 5), pp.1071–80.

Maekawa, S. et al., 2016. Crystal structure of NOD2 and its implications in human disease. Nature communications, 7(May), p.11813.

Mahid, S.S. et al., 2006. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clinic proceedings, 81(11), pp.1462–71.

Maier, H.J. & Britton, P., 2012. Involvement of autophagy in coronavirus replication. Viruses, 4(12), pp.3440–51.

Mandell, G.L., 1973. Interaction of intraleukocytic bacteria and antibiotics. Journal of Clinical Investigation, 52(7), pp.1673–9.

Manichanh, C. et al., 2006. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut, 55(2), pp.205–11.

Manuc, T.-E.M., Manuc, M.M. & Diculescu, M.M., 2016. Recent insights into the molecular pathogenesis of Crohn's disease: a review of emerging therapeutic targets. Clinical and experimental gastroenterology, 9, pp.59–70.

Margiotta, A. & Bucci, C., 2016. Role of intermediate filaments in vesicular traffic. Cells, 5(2), p.20.

Marks, D.J.B. et al., 2010. Crohn's Disease: an Immune Deficiency State. Clinical Reviews in Allergy & Immunology, 38(1), pp.20–31.

Marques, A.C. et al., 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biology, 9(3), p.R54.

Martinek, L. et al., 2015. NOD2/CARD15 mutations and the risk of reoperation in patients with Crohn's disease. Rozhl Chir, 94(6), pp.242–6.

Martinez-Medina, M. et al., 2009. Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli* (AIEC). BMC microbiology, 9, p.202.

Martinez-Vicente, M. & Cuervo, A.M., 2007. Autophagy and neurodegeneration: when the cleaning crew goes on strike. Lancet Neurology, 6(4), pp.352–61.

Martinez, E. et al., 2014. Identification of OmpA, a *Coxiella burnetii* protein involved in host cell invasion, by multi-phenotypic high-content screening. PLoS Pathogens, 10(3), p.e1004013.

Martinez, E., Cantet, F. & Bonazzi, M., 2015. Generation and multi-phenotypic high-content screening of *Coxiella burnetii* transposon mutants. Journal of visualized experiments : JoVE, (99), pp.1–11.

Massier, S. et al., 2015. DOP077. Involvement of type VI secretion systems in virulence of adherent-invasive *Escherichia coli* isolated from patients with Crohn's disease. Journal of Crohn's and Colitis, 9(suppl 1), pp.S67–S68.

McCarroll, S.A. et al., 2008. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. Nature Genetics, 40(9), pp.1107–12.

Meijer, W.H. et al., 2007. ATG genes involved in non-selective autophagy are conserved from yeast to man, but the selective Cvt and pexophagy pathways also require organism-specific genes. Autophagy, 3(2), pp.106–16.

Méresse, S. et al., 2001. Remodelling of the actin cytoskeleton is essential for replication of intravacuolar *Salmonella*. Cellular microbiology, 3(8), pp.567–77.

Michaud, A.D. et al., 1998. Presence of adherent *Escherichia coli* strains in ileal mucosa of patients with Crohn's disease. Gastroenterology, 115(6), pp.1405–13.

Michiels, F. et al., 2002. Arrayed adenoviral expression libraries for functional screening. Nature biotechnology, 20(11), pp.1154–7.

Mimouna, S. et al., 2014. HIF1A regulates xenophagic degradation of adherent and invasive *Escherichia coli* (AIEC). Autophagy, (December), pp.37–41.

Miquel, S. et al., 2010. Complete genome sequence of Crohn's disease-associated adherent-invasive *E. coli* strain LF82. PloS one, 5(9).

Miyada, C.G., Stoltzfus, L. & Wilcox, G., 1984. Regulation of the araC gene of *Escherichia coli:* catabolite repression, autoregulation, and effect on araBAD expression. Proceedings of the National Academy of Sciences of the United States of America, 81(13), pp.4120–4.

Mizushima, N. et al., 1998. A protein conjugation system essential for autophagy. Nature, 395(6700), pp.395–8.

Mizushima, N. & Komatsu, M., 2011. Autophagy: Renovation of cells and tissues. Cell, 147(4), pp.728–41.

Mizushima, N., Yoshimori, T. & Levine, B., 2010. Methods in mammalian autophagy research. Cell, 140(3), pp.313–26.

Mohawk, K.L. & O'Brien, A.D., 2011. Mouse models of *Escherichia coli* O157:H7 infection and Shiga toxin injection. Journal of Biomedicine and Biotechnology, 2011, pp.1–17.

Mollinari, C. et al., 2002. PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone. Journal of Cell Biology, 157(7), pp.1175–86.

Molodecky, N.A. et al., 2012. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology, 142(1), p.46–54.

Morabito, S. et al., 2003. A mosaic pathogenicity island made up of the locus of enterocyte effacement and a pathogenicity island of *Escherichia coli* O157:H7 is frequently present in attaching and effacing *E. coli*. Infection and Immunity, 71(6), pp.3343–8.

Morgan, X.C. et al., 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome biology, 13(9), p.R79.

Mougous, J.D. et al., 2006. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. Science, 312(5779), pp.1526–30.

Mulder, D.J. et al., 2014. A tale of two diseases: The history of inflammatory bowel disease. Journal of Crohn's and Colitis, 8(5), pp.341–8.

Mulder, D.T., Cooper, C.A. & Coombes, B.K., 2012. Type VI Secretion System-associated gene clusters contribute to pathogenesis of *Salmonella enterica* serovar Typhimurium. Infection and Immunity, 80(6), pp.1996–2007.

Munera, D. et al., 2010. N-terminal type III secretion signal of enteropathogenic *Escherichia coli* translocator proteins. Journal of Bacteriology, 192(13), pp.3534–9.

Murdoch, S.L. et al., 2011. The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors. Journal of bacteriology, 193(21), pp.6057–69.

Murphy, R.F., Velliste, M. & Porreca, G., 2003. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, 35(3), pp.311–21.

Nash, J.H. et al., 2010. Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. BMC genomics, 11(1), p.667.

Nataro, J.P. & Kaper, J.B., 1998. Diarrheagenic *Escherichia coli* strains. Clinical Microbiology Reviews, 11(1), pp.142–201.

Navarre, W.W. & Zychlinsky, A., 2000. Pathogen-induced apoptosis of macrophages: A common end for different pathogenic strategies. Cellular Microbiology, 2(4), pp.265–73.

Negroni, A. et al., 2012. Characterization of adherent-invasive *Escherichia coli* isolated from pediatric patients with inflammatory bowel disease. Inflammatory bowel diseases, 18(5), pp.913–24.

Neurath, M.F., 2014. Cytokines in inflammatory bowel disease. Nature reviews. Immunology, 14(5), pp.329–42.

Newton, H.J. et al., 2014. A screen of *Coxiella burnetii* mutants reveals important roles for dot/icm effectors and host autophagy in vacuole biogenesis. PLoS Pathogens, 10(7), p.e1004286.

Nguyen, H.T.T. et al., 2014. Crohn's disease-associated adherent invasive *Escherichia coli* modulate levels of microRNAs in intestinal epithelial cells to reduce autophagy. Gastroenterology, 146(2), pp.508–19.

Nishida, Y. et al., 2009. Discovery of Atg5/Atg7-independent alternative macroautophagy. Nature, 461(7264), pp.654–8.

O'Brien, K.P., 2004. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Research, 33(Database issue), pp.D476–D480.

Odes, S., 2008. How expensive is inflammatory bowel disease? A critical analysis. World Journal of Gastroenterology, 14(43), pp.6641–7.

Ogier, J.-C. et al., 2010. Units of plasticity in bacterial genomes: new insight from the comparative genomics of two bacteria interacting with invertebrates, Photorhabdus and Xenorhabdus. BMC genomics, 11(1), p.568.

Ogura, Y. et al., 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature, 411(6837), pp.603–6.

Ogura, Y. et al., 2001. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. The Journal of biological chemistry, 276(7), pp.4812–8.

Ostlund, G. et al., 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic acids research, 38(Database issue), pp.D196-203.

Pareja, M.E.M. & Colombo, M.I., 2013. Autophagic clearance of bacterial pathogens: molecular recognition of intracellular microorganisms. Frontiers in cellular and infection microbiology, 3(September), p.54.

Paton, A. & Paton, J., 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for stx1, stx2, eaeA, enterohemorrhagic *E. coli* hlyA, rfbO111, and rfbO157. Journal of Clinical Microbiology, 36(2), pp.598–602.

Pearson, J.S. et al., 2013. A type III effector antagonizes death receptor signalling during bacterial gut infection. Nature, 501(7466), pp.247–51.

Pegoraro, G. et al., 2014. A high-content imaging assay for the quantification of the *Burkholderia pseudomallei* induced multinucleated giant cell (MNGC) phenotype in murine macrophages. BMC microbiology, 14(1), p.98.

Petryszyn, P.W. & Paradowski, L., 2015. Economic implications of biological therapy for Crohn's disease. Gastroenterology Review, 4(4), pp.197–202.

Petryszyn, P.W. & Witczak, I., 2016. Costs in inflammatory bowel diseases. Przeglad Gastroenterologiczny, 11(1), pp.6–13.

Pugsley, H.R., 2016. Quantifying autophagy: Measuring LC3 puncta and autolysosome formation in cells using multispectral imaging flow cytometry. Methods, 112, pp.1–10.

Pujol, C. & Bliska, J.B., 2005. Turning *Yersinia* pathogenesis outside in: Subversion of macrophage function by intracellular yersiniae. Clinical Immunology, 114(3 SPEC. ISS.), pp.216–26.

Pukatzki, S. et al., 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. Proceedings of the National Academy of Sciences, 103(5), pp.1528–33.

Pukatzki, S., McAuley, S.B. & Miyata, S.T., 2009. The type VI secretion system: translocation of effectors and effector-domains. Current opinion in microbiology, 12(1), pp.11–7.

Pyo, J.-O., Nah, J. & Jung, Y.-K., 2012. Molecules and their functions in autophagy. Experimental & Molecular Medicine, 44(2), p.73.

Qin, J. et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 464(7285), pp.59–65.

Qiu, J. et al., 2016. Ubiquitination independent of E1 and E2 enzymes by bacterial effectors. Nature, 533(7601), pp.120–4.

Quan, J. & Tian, J., 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PloS one, 4(7), p.e6441.

R Development Core Team, R., 2011. R: A language and environment for statistical computing, Available at: http://www.r-project.org.

Ramos-Morales, F. & Ramos-Morales, F., 2012. Impact of *Salmonella enterica* Type III Secretion System effectors on the eukaryotic host cell. ISRN Cell Biology, 2012, pp.1–36.

Rasmussen, B. et al., 1991. Molecular basis of tetracycline action: Identification of analogs whose primary target is not the bacterial ribosome. Antimicrobial Agents and Chemotherapy, 35(11), pp.2306–11.

Ravikumar, B. et al., 2010. Plasma membrane contributes to the formation of pre-autophagosomal structures. Nature cell biology, 12(8), pp.747–57.

Reisch, C.R. & Prather, K.L.J., 2015. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. Scientific Reports, 5(October), p.15096.

Remm, M., Storm, C.E. V & Sonnhammer, E.L.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. Journal of molecular biology, 314(5), pp.1041–52.

Rencz, F. et al., 2015. Biological therapy in inflammatory bowel diseases: access in Central and Eastern Europe. World journal of gastroenterology : WJG, 21(6), pp.1728–37.

Ricciardi, R. et al., 2009. Epidemiology of *Clostridium difficile* colitis in hospitalized patients with inflammatory bowel diseases. Diseases of the Colon and Rectum, 52(1), pp.40–5.

Rogge, M.L. & Thune, R.L., 2011. Regulation of the *Edwardsiella ictaluri* type III secretion system by pH and phosphate concentration through EsrA, EsrB, and EsrC. Applied and Environmental Microbiology, 77(13), pp.4293–302.

Rohena, C.C. & Mooberry, S.L., 2014. Recent progress with microtubule stabilizers: new compounds, binding modes and cellular activities. Natural product reports, 31(3), pp.335–55.

Rohn, J.L. et al., 2011. Comparative RNAi screening identifies a conserved core metazoan actinome by phenotype. Journal of Cell Biology, 194(5), pp.789–805.

Roier, S. et al., 2016. A novel mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. Nature communications, 7, p.10515.

Rolhion, N. et al., 2005. Strong decrease in invasive ability and outer membrane vesicle release in Crohn's disease-associated adherent-invasive *Escherichia coli* strain LF82 with the yfgL gene deleted. Journal of bacteriology, 187(7), pp.2286–96.

Russell, A. et al., 2012. A widespread bacterial type VI secretion effector superfamily identified using a heuristic approach. Cell host & microbe, 11(5), pp.538–49.

Russell, A.B. et al., 2011. Type VI secretion delivers bacteriolytic effectors to target cells. Nature, 475(7356), pp.343–7.

Sakurai, Y. et al., 2015. Ebola virus. Two-pore channels control Ebola virus host cell entry and are drug targets for disease treatment. Science, 347(6225), pp.995–8.

Salcedo, S.P. et al., 2001. Intracellular replication of *Salmonella typhimurium* strains in specific subsets of splenic macrophages in vivo. Cellular Microbiology, 3(9), pp.587–97.

Salmond, G.P.C. & Reeves, P.J., 1993. Membrance traffic wardens and protein secretion in Gram-negative bacteria. Trends in Biochemical Sciences, 18(1), pp.7–12.

Salomon, D. et al., 2014. Marker for type VI secretion system effectors. Proceedings of the National Academy of Sciences of the United States of America, 111(15), pp.9271–6.

Sana, T.G. et al., 2016. *Salmonella Typhimurium* utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. Proceedings of the National Academy of Sciences, 113(34), pp.e5044–51.

SantaLucia, J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences of the United States of America, 95(4), pp.1460–5.

Sartor, R.B., 2005. Does *Mycobacterium avium* subspecies paratuberculosis cause Crohn’s disease? Gut, 54(7), pp.896–8.

Sato, Y. et al., 2003. Localization of influenza virus proteins to nuclear dot 10 structures in influenza virus-infected cells. Virology, 310(1), pp.29–40.

Schmidt, H. & Hensel, M., 2004. pathogenicity islands in bacterial pathogenesis pathogenicity islands in bacterial pathogenesis. Clinical Microbiology Reviews, 17(1), pp.14–56.

Schorey, J.S. & Bhatnagar, S., 2008. Exosome function: From tumor immunology to pathogen biology. Traffic, 9(6), pp.871–81.

Schwarz, S., Hood, R.D. & Mougous, J.D., 2010. What is type VI secretion doing in all those bugs? Trends in microbiology, 18(12), pp.531–7.

Schwechheimer, C. & Kuehn, M.J., 2015. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. Nature reviews. Microbiology, 13(10), pp.605–19.

Scribano, M.L. & Prantera, C., 2013. Use of antibiotics in the treatment of Crohn's disease. World Journal of Gastroenterology, 19(5), pp.648–53.

Serbati, N. et al., 2015. NOD2/CARD15 gene influences disease behaviour but not IBD susceptibility in a Moroccan population. The Turkish Journal of Gastroenterology, 25(1), pp.122–8.

Shalom, G., Shaw, J.G. & Thomas, M.S., 2007. In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. Microbiology, 153(8), pp.2689–99.

Shao, Q., Hawkins, A. & Zeng, L., 2015. Phage DNA dynamics in cells with different fates. Biophysical Journal, 108(8), pp.2048–60.

Shaw, S.Y., Blanchard, J.F. & Bernstein, C.N., 2010. Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. The American Journal of Gastroenterology, 105(12), pp.2687–92.

Shi, C. & Pamer, E.G., 2011. Monocyte recruitment during infection and inflammation. Nature reviews. Immunology, 11(11), pp.762–74.

Shneider, M.M. et al., 2013. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. Nature, 500(7462), pp.350–3.

Slagowski, N.L. et al., 2008. A functional genomic yeast screen to identify pathogenic bacterial proteins. PLoS pathogens, 4(1), p.e9.

Small, C.-L.N. et al., 2013. Persistent infection with Crohn's disease-associated adherent-invasive *Escherichia coli* leads to chronic inflammation and intestinal fibrosis. Nature Communications, 4, p.1957.

Snyder, L.A.S. & Saunders, N.J., 2006. The majority of genes in the pathogenic Neisseria species are present in non-pathogenic Neisseria lactamica, including those designated as “virulence genes”. BMC genomics, 7, p.128.

Soderborg, T.K. et al., 2016. Microbial transmission from mothers with obesity or diabetes to infants: an innovative opportunity to interrupt a vicious cycle. Diabetologia, 59(5), pp.895–906.

Spinner, J.L., Cundiff, J.A. & Kobayashi, S.D., 2008. *Yersinia pestis* type III secretion system-dependent inhibition of human polymorphonuclear leukocyte function. Infection and Immunity, 76(8), pp.3754–60.

Stanley, S.A. et al., 2014. Identification of host-targeted small molecules that restrict intracellular *Mycobacterium tuberculosis* growth. PLoS Pathogens, 10(2), p.e1003946.

Stone, C.D., 2012. The economic burden of inflammatory bowel disease: Clear problem, unclear solution. Digestive Diseases and Sciences, 57(12), pp.3042–4.

Strober, W., 2011. Adherent-invasive *E. coli* in Crohn disease: bacterial “agent provocateur”. The Journal of clinical investigation, 121(3), pp.841–4.

Su, H. & Wang, X., 2011. P62 stages an interplay between the ubiquitin-proteasome system and autophagy in the heart of defense against proteotoxic stress. Trends in Cardiovascular Medicine, 21(8), pp.224–8.

Sun, R. et al., 2015. Ammonium chloride inhibits autophagy of hepatocellular carcinoma cells through SMAD2 signaling. Tumor Biology, 36(2), pp.1173–7.

Tawfik, A., Flanagan, P.K. & Campbell, B.J., 2014. *Escherichia coli*-host macrophage interactions in the pathogenesis of inflammatory bowel disease. World Journal of Gastroenterology, 20(27), pp.8751–63.

Taylor, D.L., 2007. Past, present, and future of high content screening and the field of cellomics. Methods in molecular biology, 356, pp.3–18.

Thomas, G.A. et al., 2000. Role of smoking in inflammatory bowel disease: implications for therapy. Postgraduate medical journal, 76(895), pp.273–9.

van Tonder, A.J. et al., 2014. Defining the estimated core genome of bacterial populations using a bayesian decision model. PLoS Computational Biology, 10(8), p.e1003788.

Tooze, S.A., 2013. Current views on the source of the autophagosome membrane. Essays Biochem, 55(1), pp.29–38.

Tooze, S. a & Yoshimori, T., 2010. The origin of the autophagosomal membrane. Nature cell biology, 12(9), pp.831–5.

Travassos, L.H. et al., 2010. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. Nature immunology, 11(1), pp.55–62.

Trifan, A. et al., 2014. Impact of *Clostridium difficile* infection on inflammatory bowel disease outcome: A review. World Journal of Gastroenterology, 20(33), pp.11736–42.

Troisfontaines, P. & Cornelis, G., 2005. Type III Secretion: More systems than you think. Physiology, 20, p.326.

Tyrer, P., Frizelle, F. & Keenan, J., 2014. *Escherichia coli*-derived outer membrane vesicles are genotoxic to human enterocyte-like cells. Infectious agents and cancer, 9(1), p.2.

Umekawa, M. & Klionsky, D.J., 2012. The cytoplasm-to-vacuole targeting pathway: A historical perspective. International Journal of Cell Biology, 2012, pp.1–9.

Vance, J.E. & Tasseva, G., 2013. Formation and function of phosphatidylserine and phosphatidylethanolamine in mammalian cells. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids, 1831(3), pp.543–54.

Vavassori, P. et al., 2004. CARD15 mutation analysis in an Italian population: Leu1007fsinsC but neither Arg702Trp nor Gly908Arg mutations are associated with Crohn's disease. Inflammatory Bowel Diseases, 10(2), pp.116–21.

Vazeille, E. et al., 2015. Monocyte-derived macrophages from Crohn's disease patients are impaired in the ability to control intracellular adherent-invasive *Escherichia coli* and exhibit disordered cytokine secretion profile. Journal of Crohn's and Colitis, 9, pp.410–20.

Wallace, K.L. et al., 2014. Immunopathology of inflammatory bowel disease. World journal of gastroenterology : WJG, 20(1), pp.6–21.

Wang, G. et al., 2010. Prediction of pathogenicity islands in Enterohemorrhagic *Escherichia coli* O157:H7 using genomic barcodes. FEBS Letters, 584(1), pp.194–198.

Wang, H. et al., 2003. Nicotinic acetylcholine receptor alpha7 subunit is an essential regulator of inflammation. Nature, 421(6921), pp.384–388.

Wang, H.H. & Church, G.M., 2011. Multiplexed genome engineering and genotyping methods: Applications for synthetic biology and metabolic engineering. Methods in Enzymology, 498, pp.409–26.

Warren, D.J., 2011. Preparation of highly efficient electrocompetent *Escherichia coli* using glycerol/mannitol density step centrifugation. Analytical Biochemistry, 413(2), pp.206–7.

Weidberg, H., Shvets, E. & Elazar, Z., 2011. Biogenesis and cargo selectivity of autophagosomes. Annual review of biochemistry, 80, pp.125–56.

Weiner, A. et al., 2016. Macropinosomes are key players in Early *Shigella* invasion and vacuolar escape in epithelial cells. PLOS Pathogens, 12(5), p.e1005602.

Wen, H., Miao, E.A. & Ting, J.P.Y., 2013. Mechanisms of NOD-like receptor-associated inflammasome activation. Immunity, 39(3), pp.432–41.

Wen, Q. et al., 2012. Identification of regulators of polyploidization presents therapeutic targets for treatment of AMKL. Cell, 150(3), pp.575–89.

Wetterstrand, K., 2015. DNA Sequencing Costs. National Human Genome Research Institute, 20(June 2015), pp.2–4.

White, E. & DiPaola, R.S., 2009. The double-edged sword of autophagy modulation in cancer. Clinical Cancer Research, 15(17), pp.5308–16.

Wickham, H., 2009. ggplot2, Available at: http://link.springer.com/10.1007/978-0-387-98141-3\npapers3://publication/doi/10.1007/978-0-387-98141-3.

Wickham, H. et al., 2015. Package " RSQLite ." CRAN Reference manual. Available at: http://cran.r-project.org/web/packages/RSQLite/.

Wine, E. et al., 2010. Adherent-invasive *Escherichia coli* target the epithelial barrier. Gut Microbes, 1(2), pp.80–4.

Wollman, R. & Stuurman, N., 2007. High throughput microscopy: from raw images to discoveries. Journal of Cell Science, 120(21), pp.3715–22.

Xie, Z., Nair, U. & Klionsky, D.J., 2008. atg8 controls phagophore expansion during autophagosome formation. Molecular Biology of the Cell, 19(8), pp.3290–8.

Yoon, H. et al., 2011. Discovery of *Salmonella* virulence factors translocated via outer membrane vesicles to murine macrophages. Infection and Immunity, 79(6), pp.2182–92.

Youle, R.J. & Narendra, D.P., 2011. Mechanisms of mitophagy. Nature Reviews Molecular Cell Biology, 12(1), pp.9–14.

Zanella, F., Lorens, J.B. & Link, W., 2010. High content screening: Seeing is believing. Trends in Biotechnology, 28(5), pp.237–45.

Zhang, L. et al., 2009. Detection and isolation of *Campylobacter* species other than *C. jejuni* from children with Crohn's disease. Journal of Clinical Microbiology, 47(2), pp.453–5.

Zhang, Y. et al., 2015. Identification of candidate Adherent-Invasive *E. coli* signature transcripts by genomic/transcriptomic analysis. Plos One, 10(6), p.e0130902.

Zhang, Y., Werling, U. & Edelmann, W., 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic acids research, 40(8), p.e55.

van Zutphen, T. et al., 2014. Lipid droplet autophagy in the yeast *Saccharomyces cerevisiae*. Molecular Biology of the Cell, 25(2), pp.290–301.

Appendix I: Primers for LF82 unique genes

Table A1.1. **A list of primers designed using Primal.pl with stringent settings**. These will be used for the amplification of the majority of LF82 unique gene set in preparation for PCR R2.

Library primers designed using Primal stringent (Tm 50-55) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_004	ACGTATGGAGAAGCG	TGCCTGGTCTTGTG
LF82_030	ATAAACGATATCGTTCAAATAACG	GTCTGTTATTAACTCTAGTGTT
LF82_032	CGTGACTGCCATTC	GCCTATATTTACCGTACC
LF82_064	CTGAAAGCACAGCTC	TCTTTCGCTCCTTTGT
LF82_071	TTTATATTGCCTCCATGC	GGGTGGTAAGGCAC
LF82_089	AATGTCCTAATTATCTCACGC	CATGTTGCGAAACACA
LF82_091	CACTTGACGGTTTCTC	TTCTAAATCGTCATATTCATTAAAGG
LF82_092	GCAAAGGCTCCTGT	TTCCGATTTCTCACCA
LF82_093	GATCACTACCTTGAGATCC	AAACCAGGGAATAGTGG
LF82_101	AAAAAGACTCCATGGGA	ATTTATTCGTATGCCTGC
LF82_102	GATTCAGATTTAGAGCCAG	TTTGATAAGAAGTGAAGGTTTC
LF82_104	GAAGAACAAGCAAACAAGA	CTTAACCAATGATGCTGC
LF82_105	GAGTTTCATGAAAGTGCG	GGCAGCCTTGTTGT
LF82_106	ACAGTAGTCATTACATATCTGG	AATAAGTGGTTTGCTGC
LF82_114	CTTTTAATTCAACCTGGATTTG	AGCATGGCTTTCACC
LF82_129	GAAAACGAACTGATCATCG	TTCGATTTTTGCGCC
LF82_133	GCGGATTCTAATCTCAATG	CTGCTCCGCGATTA
LF82_138	ATGGCAAAGATAAGGAGA	TCTGATGTTACTGCTACT
LF82_152	AAATCGAACATCATCCAC	TTTTAATCTATCAAGGAGTGAC
LF82_153	AAAGATAACAGACAAAAAGTCG	TTTTGCAACCAATTCTCC
LF82_159	AAAATTAACAAGATATTATCATCTGC	CAGGTCAGCGTTTG
LF82_217	CCCTCATTTCTTCAGG	AGACAAACGTGGGT
LF82_220	GCCCCCATCGTAAC	GGATGATTTATTCATAGCAGG
LF82_223	AATGTCAGAATCGTGTTAC	CACCGTAGCAAAACTAA
LF82_230	GCTGAAAATCATATTGGTTG	TAAATAATCCTCTATGTTTCCTAAAA

Library primers designed using Primal stringent (Tm 50-55) settings		
Gene name	Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology	Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology
LF82_234	CCTTTAGCTGCCAGA	GCCTATATTTACCGTGC
LF82_240	AGCAGTAAAGTCAGTGG	TTTTGGCAATAATACTCCG
LF82_245	TCATATGACAGTCAGTCTT	CTCAGCTATGAGGACC
LF82_246	AGAATACTCGTTGCTGG	GATTTTTGAATTTTTCGAATACAGA
LF82_247	CTTAAAGAAAACTTCAACGAAC	GGCGATACCAAATCC
LF82_248	CTGGCCATCTGCAT	TTGCAGATTTTTCTCTTTCA
LF82_250	ATCAGACAGTTACATTATCTTG	ATGGGCGTGGTATT
LF82_281	CAGAATTTATCTCAACAGTTGT	TGATTGCTGTGATTTAATTATTTG
LF82_311	GGGATACTGGAGCA	ATCTCTTCTTCCGTAAACC
LF82_314	GCTACGATCCCCAC	TGCTTTGGCCTCCA
LF82_319	AGTGTTCAGGAAAATACTATTAAT	TTCCAGATCAGCATGT
LF82_321	GCAAAAATATTACCTTCTATATTTGG	CTGAACAGCGATTTTTAATG
LF82_323	ATATCGATCAGTGTTGCC	TTGGATTATTTTGGACACAA
LF82_325	GAACTGATCACGCAAT	AAGCGCTTTTTCTAAATCC
LF82_326	AAATTTGAAATTCTGGTGAATG	GGATGCTGCGAGAG
LF82_327	GAAAAGAGAATGACGAAACA	ATCCAGAAGTTTAATAGCCG
LF82_350	GGCGAAAACTGCTC	CCTCTTAGAAGAATTGCG
LF82_353	AATATTAAAGTATCAATAGTTATCCCT	ATCTTTGAATATTGCTAAGCAT
LF82_355	AGAAATAGAATACTTACCTTGATG	ATGACACCAAGCGT
LF82_361	TGGAGTTCGGCAAT	GAACGTCCAGACTACA
LF82_363	CACCAATCTGGTTCTG	TATATGCTGTTCTCCGC
LF82_364	TTTGAGCGTAGCGT	AAAGCATTGGGGACA
LF82_415	CATATTCAACAAGAACTCGA	TGGTTTAATATGGGCATATAC
LF82_434	TTGAAAGAATGGATGATATTTACG	TTGAATTTTACTCTCTGAAGG
LF82_435	AGTACCAATAAATCAGAGCC	CACATTACGTAACCGG
LF82_464	AGCAAAGCAGTTATTGTC	TGCGTTTTCAGCCA
LF82_535	CAAGGCATACAGTTTCA	TATTTCACCTCGTTCTTTC
LF82_536	AACGATATCGCGCA	GGCATAAACCGTGC
LF82_537	CCTGATAACAGCGC	GTCAACATAGTGCTCC
LF82_572	TCAGACAAAGACTTAACCC	CTTTTTCGACTCTTTATTCCA
LF82_580	GTACAGAATATCGTGATCACTA	TTTATTCTCACGTTGCAC
LF82_583	AGTGATGAAATTTACCTGAC	TCGACCAAATCTATTGATCC

Library primers designed using Primal stringent (Tm 50-55) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_603	ACGCTCAAGCAGAT	TAACGAGAAGATCCGAA
LF82_606	ATTACGCTAATTGGTCATC	TACATGGTGTTCTCTCA
LF82_619	TCCACGTTTAATACGCG	GTAGTAGTAAATTTGTGGATAACA
LF82_660	AATAACATCGAACAGATATCG	GAAGTCATAGCGACCT
LF82_662	AAAGCTGACTGGGT	GCTTTCATCGTATCCTC
LF82_668	AAGCTCTTTCTCTGCG	GACACTTTTCTTTGCCG
LF82_669	CAATCGTTAATGGAGATGG	AGTTTTATTTTTAACAGTTCGAAG
LF82_679	TCAATCTACGCAGTCT	CTCAATGTGCTGTACTAAA
LF82_682	TTTTGTTCTGCATTATCTTCT	GAGCAGAACCGTTG
LF82_683	GCGACTATCGAATTTATGC	CTCATGGATTTCCTCCT
LF82_685	AGTTTACGACGATTTTTACTG	ATCTGCATCGACCTT
LF82_686	ATTGGCGTGAAGTCC	ATCCAGCTCGTGTT
LF82_689	ATGCGATTCAGGAACA	AATACCTAATAAAGTCAATGAGC
LF82_692	AAACAGCTTCGACGG	TTGAAAATCGACGCTG
LF82_696	GAAGGCGTAATAAAATGGT	TTTCCTTTTGTTAGCTATGATG
LF82_697	TCGAAATTAAGTAAGCAGC	TTCCTCGATATCAGATTTATTATAG
LF82_703	AATGTTTGTACATTTGGGAC	CATCAACTCCGTTTCC
LF82_715	GCTACTTCACTGAAAAACC	GTCGCCGAGAGTTG
LF82_719	GATATGCAAAAAGAAACGATTT	GTTCCTTACCACCAAATC
LF82_722	GTTGCCATTGTTATTGTTTC	GTGGGAGGTTCTGA
LF82_734	GAATTTTATCTGGAACCCG	AAAGACTTTTACGCCATTT
LF82_747	AAATCATTTTATCAATTTCTCGC	TCTCCTGTAGTGAATCTC
LF82_750	CGTCAGTCAGCTCA	CGCATACCAGAGCC
LF82_753	ACTGCATTTGTGATCC	CAACTGCTGTGTATTAACC
LF82_759	ATGATTCTGACTCTGAATGA	TGCTGCTTCTCCAA
LF82_765	GATATAACAGCTTTTCTTCACG	CGATGCAAGAATCCATAC
LF82_769	ATTAAGAATATACAATCAAGGCTT	AAGATTGCCATCTTCATTG
LF82_770	AAAGATTTAACACTCAAATTTGCC	CATCCAGCCACGAA
LF82_774	CAGTCTGACTGGCA	GCGATAGCGTGAAG
LF82_780	CCTTTTTCTGTTTCTATAGAATCA	TGCAGTTTGCCATCC
LF82_782	AGTAAGGAGACTTATCGG	TCGTTCATGCTTTTCTC
LF82_783	GAAGAACTTCACTTTGTTTAC	TTCATCTGGGATCTCAC

Library primers designed using Primal stringent (Tm 50-55) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_787	ACACAAATACAATTCCAATCA	CGCTCTTAATAAACTTAAATAACTTT
LF82_788	GTTATGTTAGTGGATGACTG	TTTGTTTCTTCTTTGTTCACG
LF82_789	GATTTAGTCTTTAAAATTCTGGC	CGTAACCACAGTTAATGT
LF82_792	AGTACTGATATTTCAATTCGTG	CTTACCAACAATGACTCG
LF82_796	AGTCAAACCATTTTATTTGTCT	CTTTTTAAGGATCGAAACAATTT
LF82_801	TCAGATAAGCACGCC	GCGCAATACGTCAAA
LF82_802	CCACTAACAATTTTTCCTGA	CATAAGGCTCATGTGCC
LF82_803	AACGATGCCCAGAC	GCTGGCGAAGAGAT
LF82_804	AACAATGATAATAAACCATCTAGC	AGCATCATTGTCAAAATGC
LF82_808	ACATCGAAATACACTTATCTTC	TCCATGACATACCTCC
LF82_809	GAAATCACGGCTGATC	GTACTTATAGAGACGAACCC
LF82_811	GCTACTACACCATTCTATTATG	TTTATTGTTCGATTTGGTGG
LF82_098	ACCACTATGACCAAAGA	CTTCACCTCCTGCT
LF82_100	AAACAAATGTCACTAATTGAGA	CTGGTTGCCTCCTT
LF82_137	GCTAAAGCATGGAAAGA	TTCTGACAACGCATCA
LF82_139	AAGAGGATTATCGGCG	GTCTTGCTTTGTTGATGG
LF82_280	AGCATGAATATCGCACA	TTTTTGTTTTATGGTAAATGAAATAACA
LF82_312	GACAGAGGGGAAATCT	GCTCAGAATGGCCT
LF82_785	ATGCCAGTTAGCCG	TACACCGATTGCTACC
LF82_808	ACATCGAAATACACTTATCTTC	TCCATGACATACCTCC
LF82_809	GAAATCACGGCTGATC	GTACTTATAGAGACGAACCC

Table A1.2. **A list of primers designed using Primal.pl with less-stringent settings**. These will be used for the amplification of a portion of the LF82 unique gene set in preparation for PCR R2.

Library primers designed using Primal less-stringent (Tm 50-60) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_031	CGTAAAAAATACATTTCTGTTAGAG	ATTTTTTAATATTTCCATTCTGGATG
LF82_034	ATAGAAATGAAAAATAACACGCC	CCCATAATATTTTTTCCCTATACA
LF82_044	TTGATTTTTATTTTAAATTGCGTGG	ATTACCCCACCATCG
LF82_086	CAATTATCTAACCTGTTAAAAAATAAAAG	TTCCTGTGGATAACCAT
LF82_087	TTTTGTAGCCCGCC	CATATATGAAAAAATCGTGCC
LF82_090	GAAAAAACGGCGCT	TTTCAGCGCCTCCT
LF82_124	ATAGCGATAATTTTACTCATCATC	ACGCAGCCCCTTAC
LF82_130	CCTAACAATCTCGACAG	CGGATTACCGAAGAAC
LF82_140	ACAGACATTACAGCCA	TTTTAACACCCATCCTGT
LF82_147	GCGATGACGCATCC	ACTCAACGGGAGATG
LF82_177	GTGAGAACGCCGAA	CACGAAGGTCGATG
LF82_179	GGTAAAGGCAGCAG	ACGAATACTGATGCCG
LF82_189	ATTGATACAATGAAAACCTTAAGT	TTTTTTTGAAACGAGAAGTTG
LF82_219	ATGATAATTTCTAAAAAACAGCTTATT	CGATGGGGGCCATC
LF82_229	AATATTGAAAATGAAACTGGTTTTC	GCCATGAGGTCCCC
LF82_233	AGTGAAAAGCGTCATG	CGTCTTAACCCTGCC
LF82_249	AACAATACCACGCCC	TGTGTTCTCAAGCGC
LF82_251	CTACCGGGAATGGA	CTCAGGCCGGACTT
LF82_256	GGATGTATAAATATCCAGGAC	TCCGTTAGTTGTCATGT
LF82_259	GATATAACACCTTTCCTTCAT	CCCACGATTAACCAAC
LF82_260	GCTACATCGACAGTAATTC	TTTAGACGCAAGCGC
LF82_261	GCAGTACAGATTTCAGG	GATTGACTCCCCCA
LF82_282	TTTAGCTTAATCATGGCC	TGATACCGATAAAAACTTTTTTG
LF82_316	ATGAACTATAACGACGAAGT	CTCGGTCAAAGAGGC
LF82_317	CACATTCATCAGGTCG	GACATCAGCGCTGG
LF82_318	TCTGAGATCCATTCGC	CGGTGCTGGTATTATAG
LF82_320	AGTAATTTTGCAGGTAAAGTG	AGCAATAGTCCGGC

Library primers designed using Primal less-stringent (Tm 50-60) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_322	GAAATGTTGAAAAATTTAATGGAAAC	ATGTAATACGGTGCTCA
LF82_324	TCGGCAGGTACGGT	GCAGGTTATTACTGACG
LF82_352	GGTGATAATTATAATATTATACCGCA	ATCGTGGCTACCATATAA
LF82_354	GTGCTGACGCTTGC	ATATTCATTTACTCGACTCTTATTT
LF82_373	CTGAGCTTCTTTTTTGCT	AATTAACTTCCATAACATTAAAATATGTAT
LF82_378	AATTTAGTTAGCAGGAAAATGG	CGGCATAATCCGGT
LF82_398	ACGCTTGAGCGAAT	ACGATTGATGGATTCATAGT
LF82_436	GAAAGTGATATGGGAAAAAATG	GCCTATCTCCACATCC
LF82_437	GGGTGGTCAAAACC	TTCCTTCTGGGCGC
LF82_467	TTGACAAAGCTTACCGT	AAGGTGCTCGCTAA
LF82_469	CATGTCATGACTTCAGATAA	TGAAAAATAAGTATCATACCTTTTTTTAT
LF82_471	TCAATAAAGAAAACACTCATAAAAAAA	CTTTGTAATTTGTTCAAAAATCATTC
LF82_499	TCAAATATCATATACTTATCAATTAAAGGG	AAAAATCCTTTCTTCCCAAC
LF82_534	AAATCATTGAGTAAGTCGTC	ACTCACTTCCCGCC
LF82_538	CCATTAGTGAATGGACG	CTTACCCAACCGATG
LF82_0557	AACAGGAAAGAATATGCTATG	ATGCGGTGCACTTC
LF82_582	GCATTTCCCGATAACTAC	TTTTTTTGCATCCGGAT
LF82_602	ATCTGGCGGTTAATCG	TGAACTGAACCCCTG
LF82_604	GCAAAGCCGATGAAT	TCGTTTATGCGTCTCC
LF82_605	TATAACGTAACAGACGAAGT	TTCATCGGCTTTGCC
LF82_658	CATGTAATTTCAGTAATTTCTACTAAA	TTTATCCCCCTTCGG
LF82_663	AACGGACTGTTAACGC	GGCGGCACTCCATC
LF82_664	GCTGTTGTTCCCGT	TGAGCCAAAATAATTCCC
LF82_665	AATGTCTATAAACTCTCAGTTC	AGCGATCAAATCCCC
LF82_666	TCAGACAAGACAGAAAAGA	TTGCGGATCATCCC
LF82_670	TCTGACACCCCAGC	GTATGATGAGTCGGATG
LF82_672	AACGAGTCCGCTAG	TCGCACTGTCAATGT
LF82_673	AAAAAGCTCAAAGTTGTGT	GTTTACGCGCTGCC
LF82_674	TGGCTGGCAATGTT	TTGTGTTCCTCCCTG
LF82_675	ATGATGCGTTTTCCG	TTGCTCCAGCCCCT
LF82_676	AGCGACCGCTATGT	AAAAGTGGCAGCATC
LF82_678	TCGCAGATGAACACG	CAAATACTTCTTAAAACTCCCC

Library primers designed using Primal less-stringent (Tm 50-60) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_684	AGTAAGTTTAAAAATGCCCTG	GTTAAACACTCCCTTTTTCT
LF82_687	AAAGGATACGGTTTATTCATG	GAGTGCCTCCAGCC
LF82_693	ACCATTAAATACGTGTTTTGT	ATCCCCCGCCTTTT
LF82_695	ACGGCTGACAGTTAT	ACCTTTCCCTTTAGATGC
LF82_700	AAAGAGAAAAAAACATCCCT	GTGATTTAATGCTTTGGTATAG
LF82_701	GATGTATTCAAAATTTCCGTTG	GGCCAAGACGCTTC
LF82_702	TTGAATCTGAAATCTCTTCTTC	TATAACAATCAATAACGGGTTATC
LF82_705	GCATTAATCGGAGTGTATG	AATTGCCCCCCTGA
LF82_720	AATGTCTTTGTATCCGC	TCCAACTTGTTTAATCATTTCA
LF82_723	AAAATTCAAAACAAACATGTTATTGC	TCCCGCTAATCCTTC
LF82_724	AAGAAATTAATTAACCAGATTGATTC	ACATCCCCAACGAA
LF82_725	GATAAGATTATTATAAGTCCGGC	AAGTCCCAAATGATGTG
LF82_727	GTGGACATGATTAATGAAAGT	TGCCAGACGTACGC
LF82_728	CAATATTTAAAAAAACGCTTTTCTTT	TTCATCTCTCGCGG
LF82_730	AATATTGGATTTATCGGCAC	TTTACGGATAGTGAAATTGG
LF82_731	GCAAAGTTTGCAGCA	AAATTGAGTCTCGTTTTTTAGC
LF82_732	GCGAGGATGCGTGC	TAAATCGAGAATTTCTTCAAATTC
LF82_773	TGCACAGAAAACCCC	CAGTTCCATCGGGT
LF82_777	AAAAAAATCTTAGTAGTTTGCG	TTTCAACATCTCTTCTTTAACT
LF82_779	TTAAAACATCGTTATGCAAATAC	AGGTACTGGAGTCGT
LF82_781	GCAACTTATTACAGCCA	AGCTTTCGGTGTACC
LF82_790	GGTAAAAAAGACAGTAATCACC	ATCAAACAGGGAATAATTCG
LF82_797	TTGACTAAAACCGTGGA	TTCAGCTTCACCGA
LF82_799	TTGCATTTTGCGTGG	TCGTCCCACTGTTT
LF82_805	TCTGAAAAAACTAAACTTGAGA	CAGAACATCCTTTACCC
LF82_806	ACTAATCGCTTATCTCTGG	TTCTCCCAGGGTTAC
LF82_807	AAAATTCAGCTTGAACTGA	AGCGAGACAGCCGT
LF82_810	AACCTAAGCGGGCA	TGACGGGTATCCCA
LF82_136	GCTACTTGGCAAGG	ATCACCCCATAGTGAAG
LF82_313	GTCGCCGTTTTACC	CAGGATCTCCTTGCC
LF82_440	ACGATGGGAAGAGG	TTTCAATTCCCATTTACTCAC
LF82_698	ATTGACATGCAGCCC	GTCTATCGCGGCCA

Library primers designed using Primal less-stringent (Tm 50-60) settings		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_762	AGCCGCGTAATCAA	GGTTGAGGACTCTTTC
LF82_786	CCTGCTGACATTGTTG	CACCTCCGGTAACT

Table A1.3. **A list of primers designed using the NEB Tm calculator**. These will be used for the amplification of a portion of the LF82 unique gene set in preparation for PCR R2.

Library primers designed manually using Tm Calculator (NEB)		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_052	CAAAAACGTTATCTGGGT	GCGGCCTACGCGTG
LF82_146	CACAAAGCATCTCCC	TACCTGGTCTCTCTCAT
LF82_188	CACCTCCCCCCA	TCGTTCTTTTGCGTC
LF82_252	GATAAACCTGCGAATCG	CAGCTGCCAGCGCTT
LF82_315	AGTGCCCCTTTAATACG	TGCCCTTTCATGAATATTC
LF82_349	TTATCAATTAATTTAACAACTACAGC	ATTATCAAAATAACGATTTACTTT
LF82_351	GTAGCCACGATTAAAAAAATAAC	ATCACAATTCACCCTCATC
LF82_356	GCACTAATAATCTTTAGTAGGT	TGATTTTATTATTTTAGTAGATAAT
LF82_394	GCTGAACTGCAGAAG	ATTTTTCATCAATAAATACTTTAAAA
LF82_441	ACTATAAATATATTAAAAGAAA	ATTATTGCCATTTTTTTCGC
LF82_468	AAAAATTTTTTAATGTGTTTGA	CTCATCAATAAGTTTATCTAAAGGA
LF82_470	AAAGTAGGATTTATAATCTGGAATC	TACTTTTAATTTTCTAGAAAATAAA
LF82_498	AAAAGCGATCTCACTAT	CTCATTAATCAATTTTTTGATATGTTC
LF82_607	GAAAAGAAATTTTTTATAGAT	TTCAGTAAACATCGGATTC
LF82_649	GATATCAGAACGCTG	CTCTGGCGTTAACCA
LF82_659	AAACTCTATCGTGCAAC	TGCTGGCCTCCGGC

Library primers designed manually using Tm Calculator (NEB)		
Gene name	**Forward. All with "ACGGAGATCTGATATG" added to 5' end for SLiCE vector homology**	**Reverse. All with "GAGGGTACCGATTCA" added to 5' end for SLiCE vector homology**
LF82_661	ACTTCTAAAAAATCAAGCCA	TGCTGCACCCACCCT
LF82_667	ATCCGCAATAAAACAGC	TCGAATCGTCCCTTCC
LF82_671	AACCTTCATATCCCAA	TTTCAGCATCCTCTGC
LF82_677	ATGAAACAACGTTATCGG	ATGCACCGCCAGCGG
LF82_680	GCGCATTCTGCGC	GACTCGGCGGCCTG
LF82_681	AAGCAGTTGGAGCA	CGGGCGACGCTCCA
LF82_688	TTCAGTTTATTCACCAAAAAAC	TGCCGCCTCCTTCGT
LF82_690	GAGAGAATTGTGAGTGATG	ATTTATTAAATAATCATATTCATCT
LF82_691	AAATTTTCAAAGTTCCTGATCT	TCGCGCCCCCTGAAA
LF82_694	ATTCATAAACCGCGC	CATCCCCCAGCGCG
LF82_699	ACGAAGAAAACTACCAAC	TGCTGCCAGCGCTTG
LF82_729	AAGATAGTTATCTCCC	ACTCAGGATCGCAC
LF82_733	GATATTATTATAATGAATAAAG	ATCTGCATGCTCAACA
LF82_757	GCCGGGCAGTGG	TAAACATCCCGATGCG
LF82_776	GAGATCGTGTACGATCC	ACCCTCAGATTTTTCATCAT
LF82_794	ACGGGCGGCGA	ATGCAGGGAGTCTATTAT

Table A1.4. **'Common' primers used for PCR R2**. These were used for the addition of homology sequences to the destination vector, pCMV-mEm-4GS during SLiCE.

Primer name	Primer sequence
PCR_R2_Common_F	CGGTGGCGGAGGTTCCCGAATTCGTCGACGGAGATCTGATATG
PCR_R2_Common_R	GGATCCCCGCGGCCGCAAGCTTCTCGAGGGTACCGATTCA

Appendix II: LF82 unique genes

Table A2.1. **LF82 unique genes identified using InParanoid.** Asterisks indicate genes added from the list compiled by Miquel et al. (2010). The presence of these genes in other comparison sets is shown. Of the total unique LF82 gene set reported by Zhang et al. (2015), 73% were represented in the unique gene set used in this research.

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_004		
LF82_030	Y	
LF82_031	Y	
LF82_032		
LF82_034		
LF82_044		
LF82_052		
LF82_0557		
LF82_064		
LF82_071		
LF82_086		
LF82_087		
LF82_089		Y
LF82_090		Y
LF82_091		Y
LF82_092		Y
LF82_093		Y
LF82_098	Y*	
LF82_100	Y*	
LF82_101		
LF82_102		
LF82_104	Y	
LF82_105		
LF82_106		Y
LF82_114		
LF82_124		
LF82_129		
LF82_130		

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_133		
LF82_136	Y*	Y
LF82_137	Y*	Y
LF82_138	Y	
LF82_139	Y*	Y
LF82_140		Y
LF82_146		
LF82_147		
LF82_152	Y	
LF82_153	Y	
LF82_159	Y	Y
LF82_177		
LF82_179		
LF82_188		
LF82_189		
LF82_217		
LF82_219		
LF82_220		
LF82_223		
LF82_229		Y
LF82_230		Y
LF82_233		
LF82_234		
LF82_240		
LF82_245		
LF82_246		
LF82_247		
LF82_248	Y	Y
LF82_249	Y	Y
LF82_250	Y	
LF82_251	Y	Y
LF82_252		
LF82_256		
LF82_259		
LF82_260		
LF82_261		Y

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_280	Y*	
LF82_281	Y	
LF82_282	Y	
LF82_311	Y	Y
LF82_312	Y*	Y
LF82_313	Y*	
LF82_314	Y	Y
LF82_315	Y	Y
LF82_316	Y	Y
LF82_317	Y	Y
LF82_318	Y	Y
LF82_319	Y	Y
LF82_320		Y
LF82_321		Y
LF82_322		Y
LF82_323		Y
LF82_324		Y
LF82_325		Y
LF82_326		Y
LF82_327		Y
LF82_349	Y	Y
LF82_350	Y	Y
LF82_351	Y	Y
LF82_352	Y	Y
LF82_353	Y	Y
LF82_354	Y	Y
LF82_355	Y	
LF82_356	Y	Y
LF82_361		
LF82_363		
LF82_364		
LF82_373		
LF82_378		
LF82_394		
LF82_398		
LF82_415	Y	

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_434		
LF82_435		Y
LF82_436		
LF82_437		
LF82_440	Y*	Y
LF82_441		Y
LF82_464		
LF82_467		
LF82_468	Y	Y
LF82_469	Y	Y
LF82_470	Y	Y
LF82_471	Y	Y
LF82_498		
LF82_499		
LF82_534		
LF82_535		
LF82_536		
LF82_537		
LF82_538		
LF82_572		
LF82_580		
LF82_582		
LF82_583		
LF82_602	Y	Y
LF82_603	Y	Y
LF82_604	Y	Y
LF82_605	Y	Y
LF82_606	Y	Y
LF82_607	Y	Y
LF82_619		
LF82_649		
LF82_658	Y	Y
LF82_659	Y	Y
LF82_660	Y	Y
LF82_661	Y	Y
LF82_662	Y	Y

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_663	Y	Y
LF82_664	Y	Y
LF82_665	Y	Y
LF82_666	Y	Y
LF82_667	Y	Y
LF82_668	Y	Y
LF82_669	Y	Y
LF82_670	Y	
LF82_671	Y	Y
LF82_672	Y	Y
LF82_673	Y	Y
LF82_674	Y	Y
LF82_675	Y	Y
LF82_676	Y	Y
LF82_677	Y	Y
LF82_678	Y	Y
LF82_679	Y	Y
LF82_680	Y	
LF82_681	Y	Y
LF82_682	Y	Y
LF82_683	Y	Y
LF82_684	Y	Y
LF82_685	Y	Y
LF82_686	Y	Y
LF82_687	Y	Y
LF82_688	Y	Y
LF82_689	Y	Y
LF82_690	Y	Y
LF82_691	Y	Y
LF82_692	Y	Y
LF82_693	Y	Y
LF82_694	Y	Y
LF82_695	Y	Y
LF82_696	Y	Y
LF82_697	Y	Y
LF82_698	Y*	Y

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_699	Y	Y
LF82_700	Y	Y
LF82_701	Y	Y
LF82_702	Y	Y
LF82_703	Y	Y
LF82_705		
LF82_715		
LF82_719	Y	Y
LF82_720		
LF82_722		
LF82_723		Y
LF82_724		Y
LF82_725		Y
LF82_727		Y
LF82_728		Y
LF82_729		Y
LF82_730		Y
LF82_731		Y
LF82_732		Y
LF82_733		Y
LF82_734		
LF82_747		
LF82_750		
LF82_753		
LF82_757		
LF82_759		
LF82_762		
LF82_765		Y
LF82_769		Y
LF82_770		
LF82_773		
LF82_774		Y
LF82_776		Y
LF82_777		Y
LF82_779		
LF82_780		

Gene	Identified by Miquel et al. (2010)	Identified by Zhang et al. (2015)
LF82_781		
LF82_782		
LF82_783		
LF82_785		
LF82_786		
LF82_787		
LF82_788		
LF82_789		
LF82_790		
LF82_792		Y
LF82_794		
LF82_796		
LF82_797		
LF82_799		
LF82_801		
LF82_802		Y
LF82_803		Y
LF82_804		Y
LF82_805		Y
LF82_806		Y
LF82_807		Y
LF82_808		
LF82_808		
LF82_809		
LF82_809		
LF82_810		
LF82_811		

Appendix III: Layout of screen transfection plates

a)

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		LF82_004	LF82_030	LF82_032	LF82_064	LF82_071	LF82_089	LF82_091	LF82_092	LF82_093	LF82_101	
C		LF82_105	LF82_106	LF82_114	LF82_129	LF82_133	LF82_138	LF82_152	LF82_153	LF82_159	LF82_217	
D		LF82_230	LF82_234	LF82_240	LF82_245	LF82_246	LF82_247	LF82_248	LF82_250	LF82_281	LF82_311	
E		LF82_321	LF82_323	LF82_325	LF82_326	LF82_327	LF82_350	LF82_353	LF82_355	LF82_361	LF82_363	
F		LF82_434	LF82_435	LF82_464	LF82_535	LF82_536	LF82_537	LF82_572	LF82_580	LF82_583	LF82_603	
G		mEm	mEm	NH_4Cl	NH_4Cl	Rap	Rap	NH_4Cl + Rap	p62	SseD	Untr.	
H												

b)

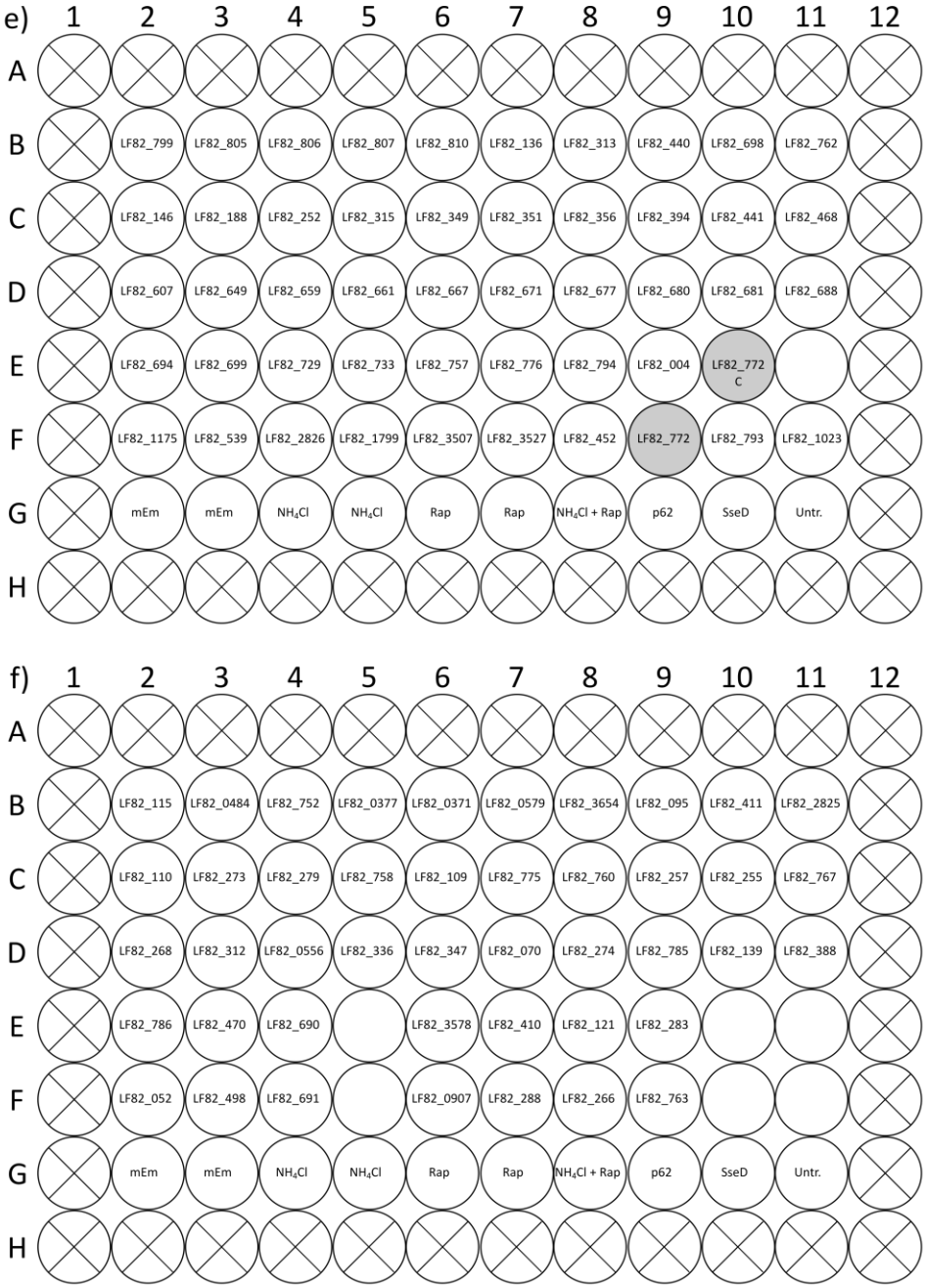
	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		LF82_660	LF82_662	LF82_668	LF82_669	LF82_679	LF82_682	LF82_683	LF82_685	LF82_686	LF82_689	
C		LF82_697	LF82_703	LF82_715	LF82_719	LF82_722	LF82_734	LF82_747	LF82_750	LF82_753	LF82_759	
D		LF82_770	LF82_774	LF82_780	LF82_782	LF82_783	LF82_787	LF82_788	LF82_789	LF82_792	LF82_796	
E		LF82_102	LF82_220	LF82_314	LF82_364	LF82_606	LF82_692	LF82_765	LF82_801			
F		LF82_104	LF82_223	LF82_319	LF82_415	LF82_619	LF82_696	LF82_769	LF82_802			
G		mEm	mEm	NH_4Cl	NH_4Cl	Rap	Rap	NH_4Cl + Rap	p62	SseD	Untr.	
H												

c)

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		LF82_803	LF82_804	LF82_808	LF82_809	LF82_811	LF82_098	LF82_100	LF82_137	LF82_139	LF82_280	
C		LF82_808	LF82_809	LF82_031	LF82_034	LF82_044	LF82_086	LF82_087	LF82_090	LF82_124	LF82_130	
D		LF82_177	LF82_179	LF82_189	LF82_219	LF82_229	LF82_233	LF82_249	LF82_251	LF82_256	LF82_259	
E		LF82_282	LF82_316	LF82_317	LF82_318	LF82_320	LF82_322	LF82_324	LF82_352	LF82_354	LF82_373	
F		LF82_436	LF82_437	LF82_467	LF82_469	LF82_471	LF82_499	LF82_534	LF82_538	LF82_0557	LF82_582	
G		mEm	mEm	NH_4Cl	NH_4Cl	Rap	Rap	NH_4Cl + Rap	p62	SseD	Untr.	
H												

d)

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		LF82_605	LF82_658	LF82_663	LF82_664	LF82_665	LF82_666	LF82_670	LF82_672	LF82_673	LF82_674	
C		LF82_678	LF82_684	LF82_687	LF82_693	LF82_695	LF82_700	LF82_701	LF82_702	LF82_705	LF82_720	
D		LF82_725	LF82_727	LF82_728	LF82_730	LF82_731	LF82_732	LF82_773	LF82_777	LF82_779	LF82_781	
E		LF82_312	LF82_140	LF82_260	LF82_378	LF82_602	LF82_675	LF82_723	LF82_790			
F		LF82_785	LF82_147	LF82_261	LF82_398	LF82_604	LF82_676	LF82_724	LF82_797			
G		mEm	mEm	NH_4Cl	NH_4Cl	Rap	Rap	NH_4Cl + Rap	p62	SseD	Untr.	
H												


e)
1 2 3 4 5 6 7 8 9 10 11 12
A
B LF82_799 LF82_805 LF82_806 LF82_807 LF82_810 LF82_136 LF82_313 LF82_440 LF82_698 LF82_762
C LF82_146 LF82_188 LF82_252 LF82_315 LF82_349 LF82_351 LF82_356 LF82_394 LF82_441 LF82_468
D LF82_607 LF82_649 LF82_659 LF82_661 LF82_667 LF82_671 LF82_677 LF82_680 LF82_681 LF82_688
E LF82_694 LF82_699 LF82_729 LF82_733 LF82_757 LF82_776 LF82_794 LF82_004 LF82_772 C
F LF82_1175 LF82_539 LF82_2826 LF82_1799 LF82_3507 LF82_3527 LF82_452 LF82_772 LF82_793 LF82_1023
G mEm mEm NH4Cl NH4Cl Rap Rap NH4Cl + Rap p62 SseD Untr.
H
f)
1 2 3 4 5 6 7 8 9 10 11 12
A
B LF82_115 LF82_0484 LF82_752 LF82_0377 LF82_0371 LF82_0579 LF82_3654 LF82_095 LF82_411 LF82_2825
C LF82_110 LF82_273 LF82_279 LF82_758 LF82_109 LF82_775 LF82_760 LF82_257 LF82_255 LF82_767
D LF82_268 LF82_312 LF82_0556 LF82_336 LF82_347 LF82_070 LF82_274 LF82_785 LF82_139 LF82_388
E LF82_786 LF82_470 LF82_690 LF82_3578 LF82_410 LF82_121 LF82_283
F LF82_052 LF82_498 LF82_691 LF82_0907 LF82_288 LF82_266 LF82_763
G mEm mEm NH4Cl NH4Cl Rap Rap NH4Cl + Rap p62 SseD Untr.
H

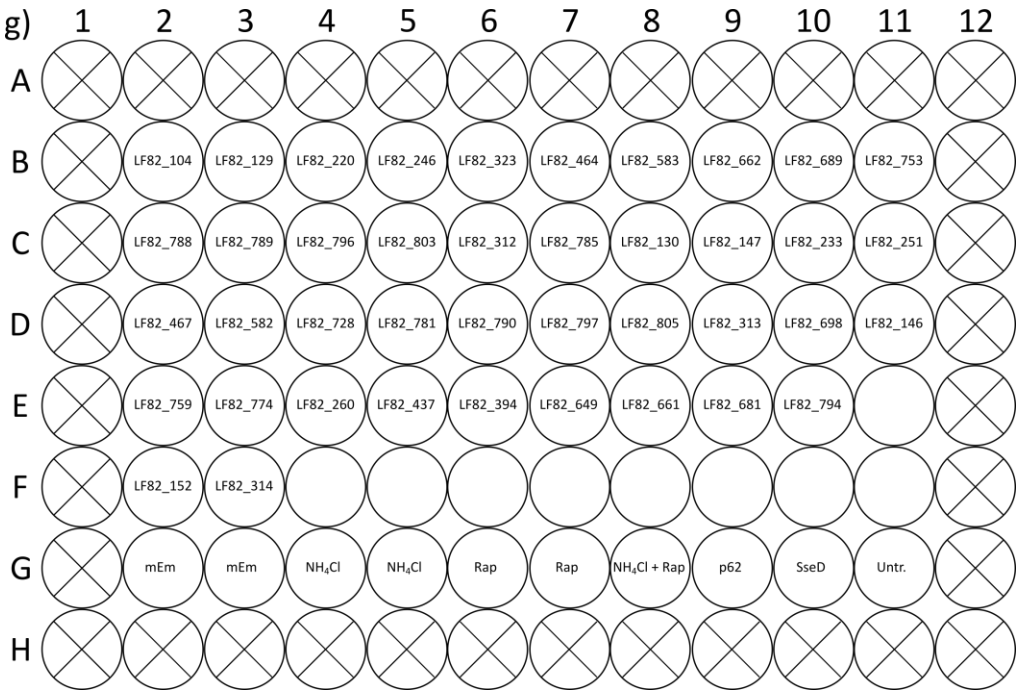
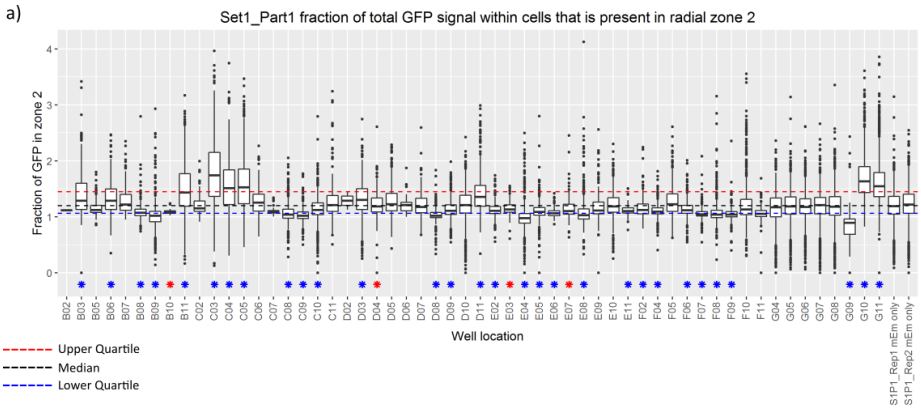


Figure A3.1. **Layout of Set 1, 2, 3, and 4 for screen transfections**. a) and b) show Set 1 Part 1 and Set 1 Part 2, c) and d) show Set 2 Part 1 and Set 2 Part 2, e) and f) show Set 3 Part 1 and Set 3 Part 2, and g) shows Set 4. Empty wells are blank, and wells not used due to limitations of automated microscopy are marked with an X. Control wells can be seen at the bottom of each respective figure (for more information, see: Transfection plate controls). Grey wells are those thought to be involved in cross contamination, 'C' designates the contaminated well (Set 3 Part 1, well E10) – this well should not have resulted in any transfection, and was the only obvious case of cross contamination seen.

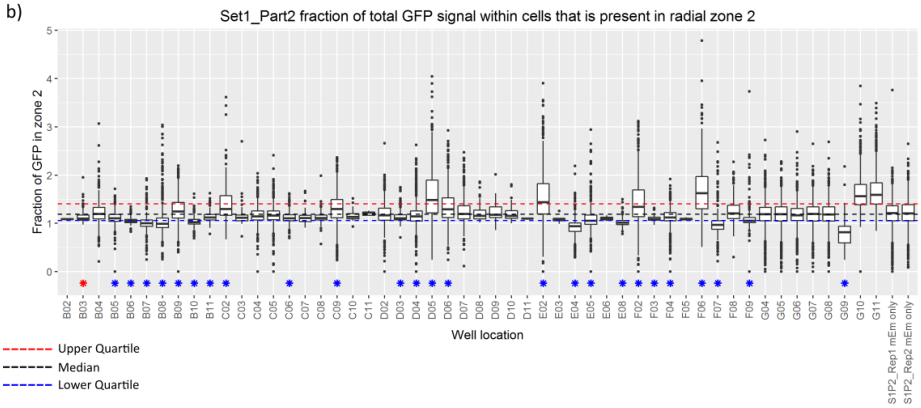
Appendix IV: Screen analysis graphs

6.2 Radial zone 2

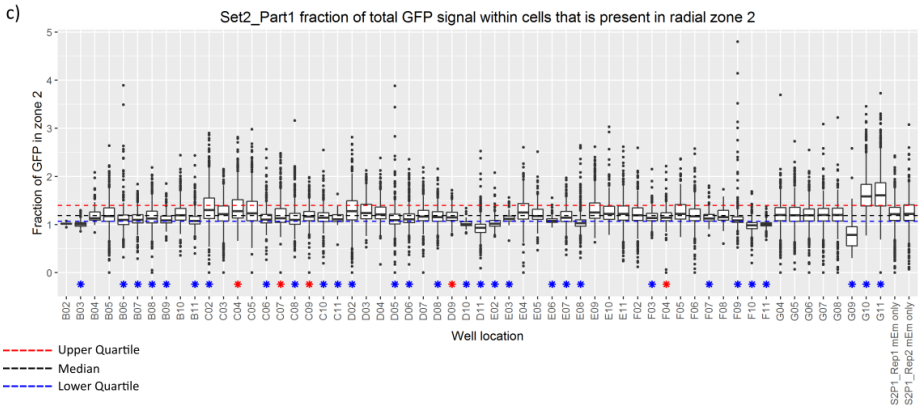
a)



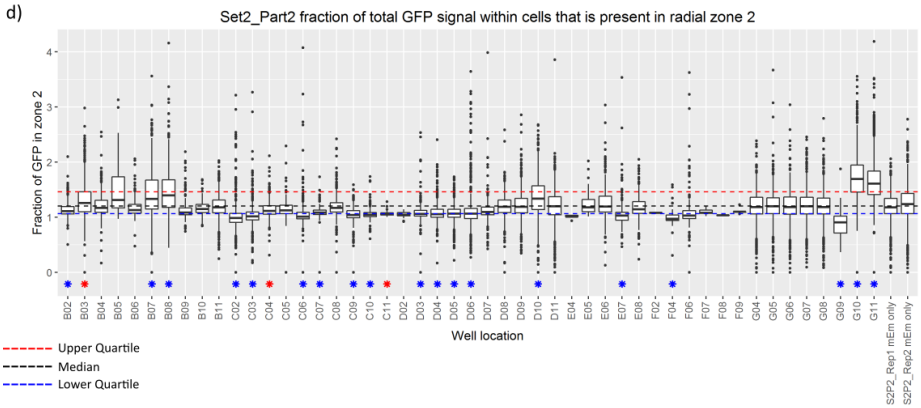
b)



c)



d)



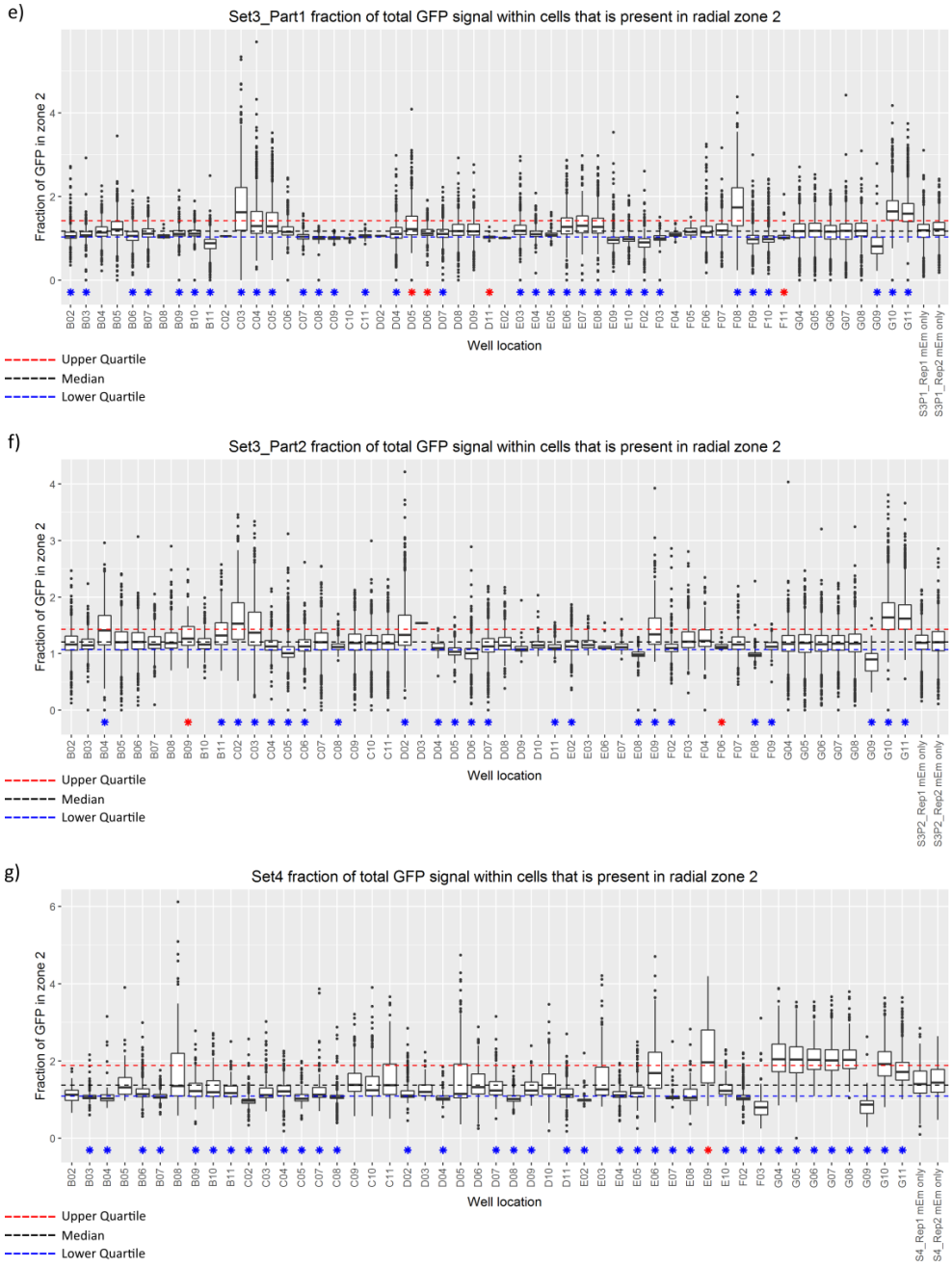


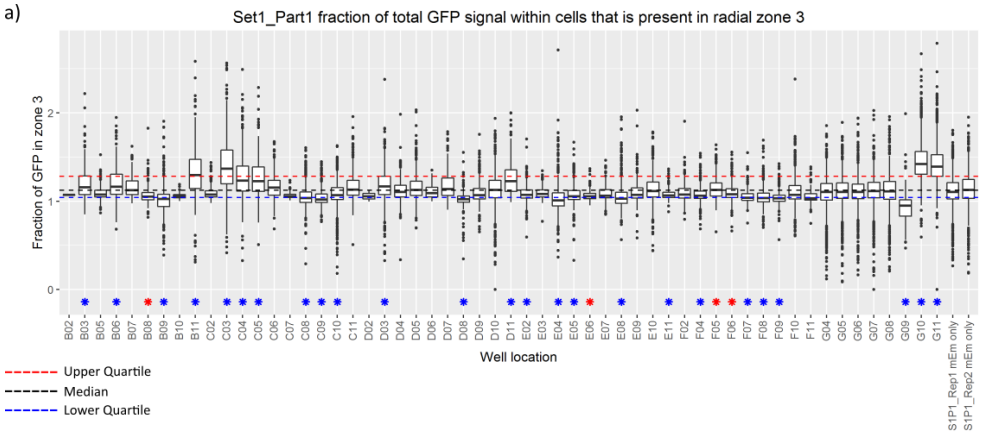
Figure A4.1. **Radial zone 2 measurements**. Boxplot graphs showing the measurement of GFP present in radial zone 2 of transfected cells. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table A4.1. **Top 20 hits from radial zone 2 data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest.

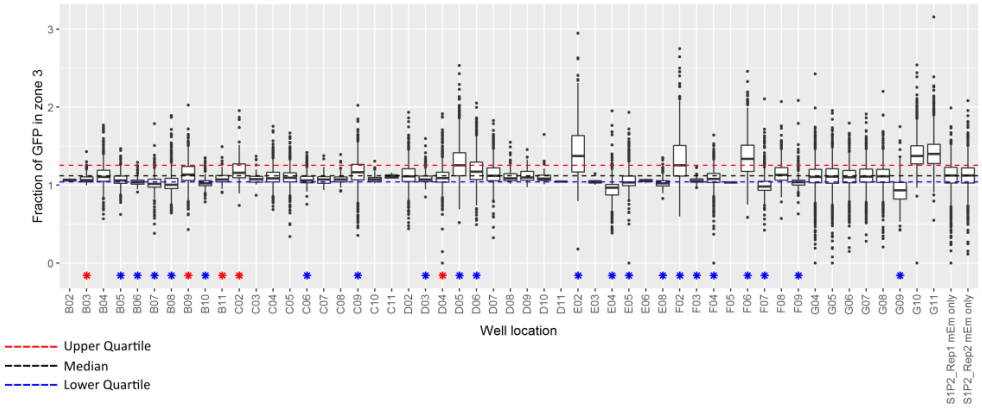
Top 20 hits for the fraction of total GFP signal within cells that is present in radial zone 2				
Plate	Well name	Gene	Median	Standard deviation
Set4	E09	LF82_681	1.11	0.50
Set1_Part1	C03	LF82_106	1.10	0.36
Set3_Part1	F08	LF82_452	1.08	0.40
Set3_Part1	C03	LF82_188	1.01	0.47
Set3_Part2	C02	LF82_110	0.99	0.31
Set1_Part2	F06	LF82_619	0.98	0.30
Set1_Part1	C04	LF82_114	0.97	0.32
Set1_Part1	C05	LF82_129	0.96	0.31
Set4	E06	LF82_394	0.95	0.38
Set3_Part2	C03	LF82_273	0.90	0.28
Set1_Part2	D05	LF82_782	0.90	0.33
Set1_Part2	E02	LF82_102	0.90	0.31
Set1_Part1	B11	LF82_101	0.89	0.27
Set3_Part2	B04	LF82_752	0.88	0.26
Set3_Part2	D03	LF82_312	0.88	NA
Set3_Part2	B11	LF82_2825	0.86	0.21
Set3_Part2	E09	LF82_283	0.85	0.24
Set2_Part2	B08	LF82_100	0.85	0.28
Set1_Part1	D11	LF82_311	0.85	0.25
Set3_Part1	C04	LF82_252	0.84	0.38

6.3 Radial zone 3

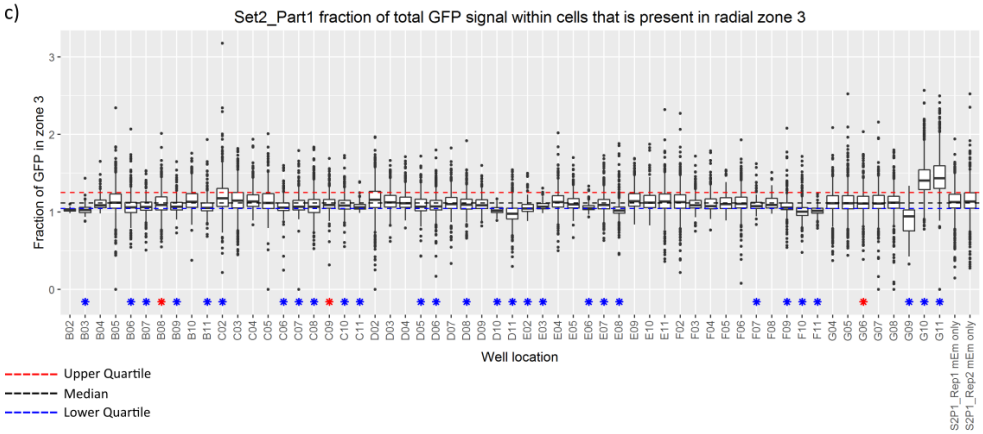
a)



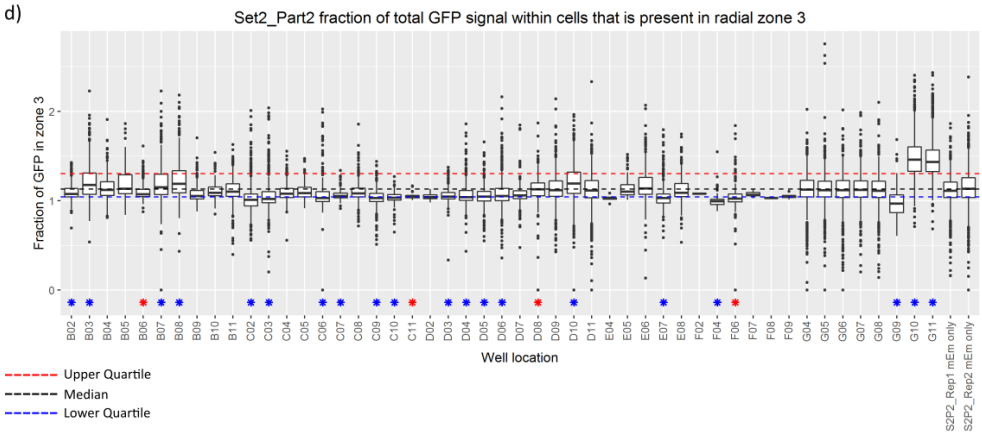
b)



c)



d)



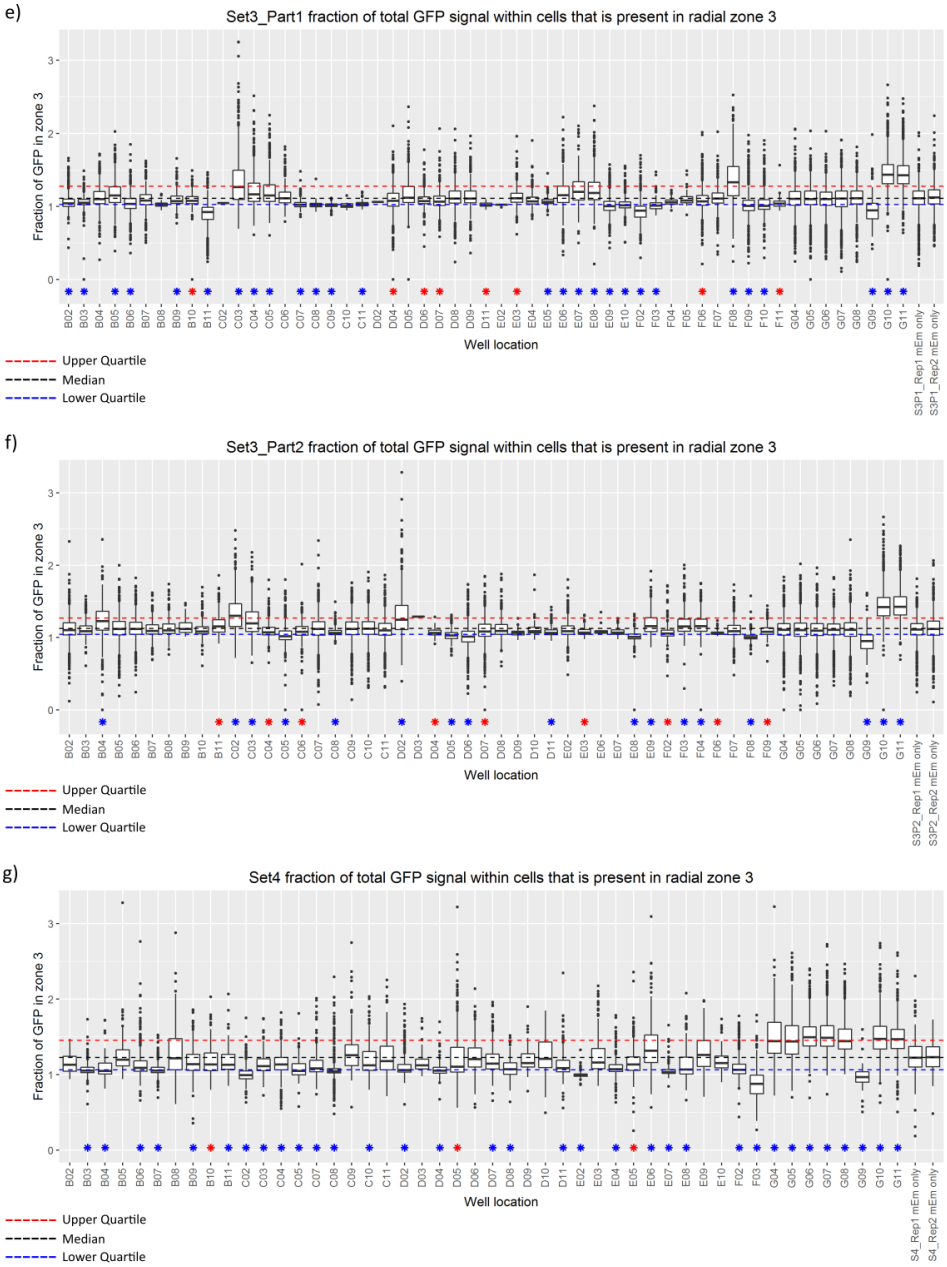


Figure A4.2. **Radial zone 3 measurements**. Boxplot graphs showing the measurement of GFP present in radial zone 3 of transfected cells. Red asterisks indicate a P value of <0.05, and blue asterisks indicate a P value of <0.01.

Table A4.2. **Top 20 hits from radial zone 3 data**. Data was normalised between plates using untreated cell data. The top row in italics displays an example of a control well, which is treated with/is expressing a chemical/protein known to affect the phenotype of interest.

Top 20 hits for the fraction of total GFP signal within cells that is present in radial zone 3				
Plate	Well name	Gene	Median	Standard deviation
Set1_Part1	C03	LF82_106	0.98	0.21
Set1_Part2	E02	LF82_102	0.96	0.23
Set3_Part1	F08	LF82_452	0.94	0.21
Set3_Part2	C02	LF82_110	0.94	0.18
Set1_Part2	F06	LF82_619	0.93	0.17
Set1_Part1	B11	LF82_101	0.92	0.19
Set3_Part2	D02	LF82_268	0.89	0.21
Set4	E06	LF82_394	0.89	0.21
Set1_Part1	C04	LF82_114	0.89	0.17
Set3_Part1	C03	LF82_188	0.88	0.24
Set1_Part2	D05	LF82_782	0.88	0.18
Set1_Part1	C05	LF82_129	0.88	0.15
Set3_Part2	B04	LF82_752	0.87	0.16
Set1_Part1	D11	LF82_311	0.87	0.13
Set3_Part2	C03	LF82_273	0.87	0.16
Set1_Part2	F02	LF82_104	0.87	0.22
Set3_Part2	D03	LF82_312	0.87	NA
Set4	E09	LF82_681	0.85	0.18
Set4	C09	LF82_147	0.85	0.17
Set3_Part1	E07	LF82_776	0.85	0.15