HIGH THROUGHPUT RNA INTERFERENCE BASED SCREEN FOR THE IDENTIFICATION OF GENES IMPLICATED IN LISTERIA MONOCYTOGENES INFECTION

By

Student: Asongwe Lionel Ateh Tantoh

Student number: TNTASO001

Thesis Presented for the Degree of DOCTOR OF PHILOSOPHY
In the Department of Medicine UNIVERSITY OF CAPE TOWN
FEBRUARY 2014

Supervisors

Prof. FRANK BROMBACHER

Division of Immunology, Institute of Infectious Diseases and Molecular Medicine
& International Centre for Genetic Engineering and Biotechnology (ICGEB) University
of Cape Town Medical School. 7925

Dr. NEIL EMANS

Persomics (PTY) LTD, Barinor’s Vineyards North. The vineyards office Etate, 99 Jip de
Jager. 7530

DEDICATION

This thesis is primarily dedicated to the Almighty God for the strength, comfort, protection and the uncounted blessings He showered upon my feeble shoulders during this period. This thesis is also dedicated to my parents; Monica Mangwi Tantoh and Sylvester Tantoh who through their relentless moral support, incredible motivation and inspiration and financial support, I was able to accomplish this difficult journey. Also, this thesis is dedicated to my late grandmother mama Beatrice Nanga Asongwe, who once said to me, “Life is a battlefield and as long as we are alive, we have to keep fighting”. Finally, I dedicate this thesis to my siblings, Doris, Clarisse, Valerie, Ferdinand, Jervis, Kelly, Ateh and to my nephew and nieces Lionel, Pashukeni, Titi and Ngwe for all the joy they brought into my life during this period.

DECLARATION

I, Asongwe Lionel Ateh Tantoh hereby declare that the work on which this thesis is based on my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: …………………………………

Date:………………………………………

ACKNOWLEDGEMENTS

It would not have been possible to write this dissertation without the help and support of some special people, to only some of whom it is possible to give particular attention here.

This thesis would not have been possible without the help, support, encouragement and patience of my principal supervisor, Prof. Frank Brombacher, not to mention his good academic advice and unsurpassed knowledge on *L. monocytogenes* infection. Thanks is also due for the good mentorship, technical support and friendship of my co-supervisor, Dr. Neil Emans, not to mention his good technical understanding in high content screening. I would also like to thank Dr. Musa Mhlanga for the patient guidance, encouragement, advice and his ceaseless moral supports he provided at the initial stages of this thesis.

I will like to thank Dr. Anita Schwegmann and Dr. Suraj Parihar for their guidance and technical contributions during the initial stages of this work. I owe a great deal of gratitude to Hannah Wambui Njau for her enormous contributions in optimizing and setting up the siRNA array platform without which the work presented in this dissertation would have not been possible. My sincere gratitude also goes to Dr. Grace Mothepane London who through her scientific discussion, support and motivation kept me in the race. I would like to thank Gillian De Villiers for taking her valuable time and helping to review this dissertation. It is well said, “All work and no play makes Jack a dull boy”. In this regard, I will be forever grateful to Hannah W. Njau, Charlotte Mashaba, Zodwa Mbambo and Cornelius Kano, who made the social part of this journey unforgettable.

I am most grateful to the Council for Scientific and Industrial Research (CSIR) and the International Centre for Genetic Engineering and Biotechnology (ICGEB) for creating the perfect platform for this work and also for all their financial supports.

ABSTRACT

The facultative intracellular pathogen *L. monocytogenes* has acquired a versatile arsenal of adaptor proteins. It uses these proteins to invade and undermine the host immune surveillance system, as well as to survive and propagate within the host cell. Our understanding of how *L. monocytogenes* infects cells points towards an interplay between bacterial proteins and human host factors. It has become apparent that this interplay is critical for the establishment and spread of infection. Here we describe a high throughput screening approach to assay for potential human host factors involved in *L. monocytogenes* infection. In this methodology, we applied a high-density siRNA array comprising of the kinome, phosphatome, ubiquitome and protease encoding human genes that enabled rapid parallel screening for human host factors through reverse transfection into overlaid HeLa cells. The siRNA transfected HeLa cells were exposed to *L. monocytogenes* expressing the green fluorescent protein. The high content screen comprised a total of 15 replicate screens or 47, 250 individual siRNA experiments. Using a combination of image-based analysis and the data-mining tool of Principal Component Analysis, genes whose silencing visually impaired or enhanced bacteria invasion/proliferation were selected as candidate genes. The 65 strongest genes were subjected to a secondary screening for validation. Among the 15 confirmed hits, we recognized some cellular membrane proteins that might function as cell receptors for bacteria entry and others that may be related to calcium release triggered by bacteria during cell invasion. In addition, we identified a transcription factor bromo adjacent homology domain-containing 1 (BAHD1) protein, whose interplay with the bacterial virulence factor LntA has recently been shown to modulate IFN-λ-mediated immune response to control bacterial colonization of the host cell. Also, we identified the inositol polyphosphate-5-phosphatase (INPP5B) that may regulate bacterial internalization by modulating actin dynamics at bacterial internalization sites as shown with oculocerebrorenal syndrome of Lowe (OCRL), whose OCRL gene shares an identical sequence and overlapping enzyme activities with INPP5B. We also identified the large tumor suppressor kinase 2 (LATS2) that may play a potential role in bacterial colonization by modulating the expression of the host E-cad receptor gene. This study represents a significant step toward unveiling the key molecular requirements for host cell invasion and undoubtedly provides an avenue of new therapeutic drugs against *L. monocytogenes* infection.

ABBREVIATIONS

ActA	Actin assembly-inducing protein
Arp2/3	Actin related protein 2/3 complex
BAHD1	Bromo adjacent homology domain-containing 1 protein
BMDMØs	Bone marrow derived macrophages
CDCs	Cholesterol-dependent pore-forming cytolysins
CFTR	Cystic fibrosis transmembrane conductance regulator
DNA	Deoxy-ribonucleic acid
E-cad	E-cadherin
GFP	Green fluorescent protein
GILT	ϒ-interferon-inducible lysosomal thiol reductase
h	Hour
Hly	Hemolysin
IFN-λ	Interferon gamma
InIA	Internalin A
InIB	Internalin B
INPP5B	Inositol polyphosphate-5-phosphatase
IR	Inter repeat
KDa	Kilo-Dalton
LATS2	Large tumor suppressor kinase 2
L. monocytogenes	*Listeria monocytogenes*
LLO	Listeriolysin O
LRR	Leucine-rich repeats
miRNA	Micro Ribonucleic acid
MOI	Multiplicity of infection
NF-κB	Nuclear factor-kappaB
OCRL	Oculocerebrorenal syndrome of Lowe
PCA	Principal Component Analysis
PFO	Perfringolysin O
PLCs	Phospholipases

PLL	Poly-L-Lysine
PMID	PubMed ID
RELA/NF-κB/p65	Nuclear factor NF-Kappa-B P65 subunit
RGB	Red, green and blue
ROI	Regions of interest
RNA	Ribonucleic acid
RNAi	Ribonucleic interference
siRNA	Small interfering Ribonucleic acid
shRNA	Short hairpin Ribonucleic acid
SLO	Streptolysin O
STDEV	Standard deviation
TNF-α	Tumour Necrosis Factor Alpha
VASP	Vasodilator-stimulated phosphoprotein
XPO1	Exportin 1
μl	Micro-liter
nM	Nano-molar
ng	Nano-gram
ml	Milli-liter
μm	Micro-meter

LIST OF FIGURES AND TABLES

CHAPTER 1 FIGURES AND TABLE

CHAPTER 4 FIGURES

CHAPTER 5 FIGURES

CHAPTER 6 FIGURES AND TABLES

LIST OF PUBLICATIONS AND CONFERENCE

AN ON-THE-FLY HIGH-CONTENT RNAI SCREEN FOR REGULATORS OF THE TNF INDUCED RELA NUCLEAR TRANSPORT PATHWAY

Manuscript (chapter five of the dissertation) is in preparation for submittion

HIGH THROUGHPUT RNA INTERFERENCE BASED SCREEN FOR HUMAN MODULATORS IMPLICATED IN LISTERIA MONOCYTOGENES INFECTION

Manuscript (chapter six of the dissertation) is in preparation for submittion

SOCIETY FOR LABORATORY AUTOMATION AND SCREENING (SLAS) 2012 1ST ANNUAL CONFERENCE AND EXHIBITION (SAN DIEGO, CA, USA).

TABLE OF CONTENTS

THESIS SUMMARY

Infectious diseases continue to be a leading cause of childhood and adult morbidity and mortality, with greater than 14 million deaths per year (Haghighat and Seveau 2010). Over 90% of these deaths occur in developing countries, where disadvantaged communities are predominately affected. The current strategy for managing infectious diseases has mainly focused on targeting processes or enzymes unique to pathogens with specific drugs to eradicate the causative pathogen. The advantage of this approach is the minimal toxic side effects to the host due to evolutionary divergence between host and pathogen. However, despite success stories, a serious disadvantage of this pathogen-directed strategy has been the development of drug resistance, resulting in the re-emergence of infectious diseases notably tuberculosis. This has become an increasingly serious problem, necessitating the development of more effective vaccines and identification of new drug targets that can be used against resistant or mutated microbial pathogens. The goal of this thesis was to screen for and identify human host factors involved in *L. monocytogenes* infection. This was approached using a combination of visual cell based assays with genome wide RNAi screening.

Understanding disease and the processes of disease has been advanced through genome scaled methods. Genome wide RNA-interference is one such technology that has emerged as a potent means to identify genes involved in disease through their function. As an example, this has been applied to defined human host factors required for HIV infection in human cells (Brass, Dykxhoorn et al. 2008; Goff 2008; König, Zhou et al. 2008; Zhou, Xu et al. 2008). Current RNAi technology for capturing such interactions between host and pathogens is limited to large-scale laboratories and expensive, cumbersome technology. A consequence of this is that RNAi screening is costly and limited to researchers in the developed world. The central goal of this project was to apply a novel miniaturized technology to genome scale screening that will exceed these limitations.

The miniaturization processes implemented in this research involved the development of a high-density siRNA arrays. A printer was used to produce a

high-density array of siRNAs contained in an encapsulation mixture on to the surface of a cover glass or optical wafer. These arrays comprised of optically addressable spots that contain a pool of siRNA targeting one gene. Cells were then seeded on to the surface. In the case of HeLa cells, the cells adhered over the entire surface of the slide. However, only cells in contact with the spotted siRNA were transfected. In these spots, a given gene is robustly silenced 48 hours post transfection. Since the arrays were printed using fractional amounts of siRNA, this enabled both mass production and very efficient screening. A single array covered the entire RNAi complement of the kinome, phosphatome, ubiquitome and protease encoding genome and was screened by imaging individual siRNA experiments. Because of the inherent miniaturization on the arrays, 3000 experiments were imaged and screened faster than conventional technologies allow: in less than 1 hour using an automated confocal microscope. This screen occurs in just one plate compared to the 30 or more required with conventional methods, reducing the time for a single screen from two weeks to an hour. Since the position of each siRNA was known, as if they were wells, observed phenotypic changes could be attributed to specific siRNAs.

The research work presented set out to exploit the miniaturized high throughput screening technology to find human host factors involved in *Listeria* infection. To achieve this, arrayed siRNA was combined with image analysis of a cell-based assay of *Listeria monocytogenes* infection.

CHAPTER I

LITERATURE REVIEW

1.1 Introduction

Intracellular pathogens remain a serious threat to quality of life. They are responsible for a vast number of infectious diseases, morbidity and mortality worldwide (Vannberg, Chapman et al. 2011; Thi, Lambertz et al. 2012). These pathogens encompass bacterial, viral and eukaryotic parasite infections that enable reproduction by co-opting the host cell's endogenous pathways. The impact of these pathogens is widespread, for despite their relatively small genome component compared to humans, they have evolved sophisticated mechanisms to counter and evade immune surveillance. Several bacterial pathogens share this capacity to evade immune responses, such as microorganisms from the genus *Listeria*.

1.2 Genus *Listeria*

1.2.1 Taxonomy of the genus *Listeria*

For several years, the genus *Listeria* generic taxonomy remained unclear. The organism was originally classified based on the sixth and seven editions of Bergey's manual of Determinative Biotechnology in the family Corynebacteriaceae (Rocourt and Mollaret 1988). However, based on results obtained from many biochemical studies, DNA homology, numerical taxonomic and 16S rRNA sequencing homology, organisms under this genus are now classified into the domain: **bacteria**, kingdom: **eubacteria**, division: **firmicutes**, class: **Bacilli**, order: **bacillales**, family: **Listeriaceae** and genus: **Listeria**.

The genus *Listeria* currently includes six species: *L. monocytogenes, L. ivanovii, L. innocua, L. seeligeri, L. welshimeri,* and *L. grayi* (Vazquez-Boland, Kuhn et al. 2001). However, recent studies have revealed a new species (*L. marthii*) as a member belonging to the genus *Listeria* (Graves, Helsel et al. 2010). Despite their relative taxonomical similarities, only two species of the genus *Listeria* have been significantly associated with severe illnesses in both humans and animals: *L. monocytogenes and L. ivanovii* (Hain, Chatterjee et al. 2007; AbdEl-Malek, Ali et al. 2010).

Reports suggest organisms under this genus descended from a common ancestor. Phylogenetic studies, based on the 16S and 23S rRNA genes and the *prs, iap, vclA*, *vclB*, and *ldh* genes reveals genes required for intracellular survival, transcription factors, and LPXTG- and LRR-containing proteins and 55 genes involved in carbohydrate transport and metabolism that were deleted from the nonpathogenic species over time (Hain, Steinweg et al. 2006). The latter statement has been supported through characterization of the genetic unit containing the virulence genes cluster of *Listeria* species. The virulence genes (*prfA, plcA, hly, mpl, actA and plcB*) required for its intracellular life cycle are physically linked in a 9-kb chromosomal island called the PrfA-dependent virulence island or the *Listeria* pathogenicity island 1 (LIPI-1).

The LIPI-1 chromosomal fragment is found in two of the six *Listeria* species: *L. monocytogenes and L. ivanovii* and are flanked by two housekeeping genes *prs* and *Idh*, both encoding for the enzymes phosphoribosyl-pyrophosphate synthase and lactate dehydrogenase respectively (Schmid, Ng et al. 2005). In the *plcB-Idh* intergenic region, one virulence cluster locus (*vcl*) or open reading frame (*orf*), *orfB* is found in all six *Listeria* species, indicating that the insertion point of the virulence island is between the *prs* and *orfB* loci. Another orf, present in all species, with the exception of *L. grayi*, is the *orfA*. Other genes potentially controlled by the master virulence regulator PrfA (positive regulatory factor A) include: *dplcB, orfD, orfE, orfK, orfL, orfX,* and *orfZ* while others such as the *orfC, orfF1, orfF2, orfG1*, and *orfG2* of the *L. seeligeri and L. grayi* are not controlled by the PrfA (Vazquez-Boland, Kuhn et al. 2001). The fact that the LIPI-1 fragment is located at the same position within members of the same phylogenetic groups (*L. monocytogenes* and *L. ivanovii*) but missing in other members of the same groups, clearly indicates that LIPI-1 has been acquired by or evolved in a common ancestor of the present *Listeria* species and has been deleted from the chromosome of *L. innocua* and *L. welshimeri* (Vazquez-Boland, Kuhn et al. 2001), (Figure 1.1).

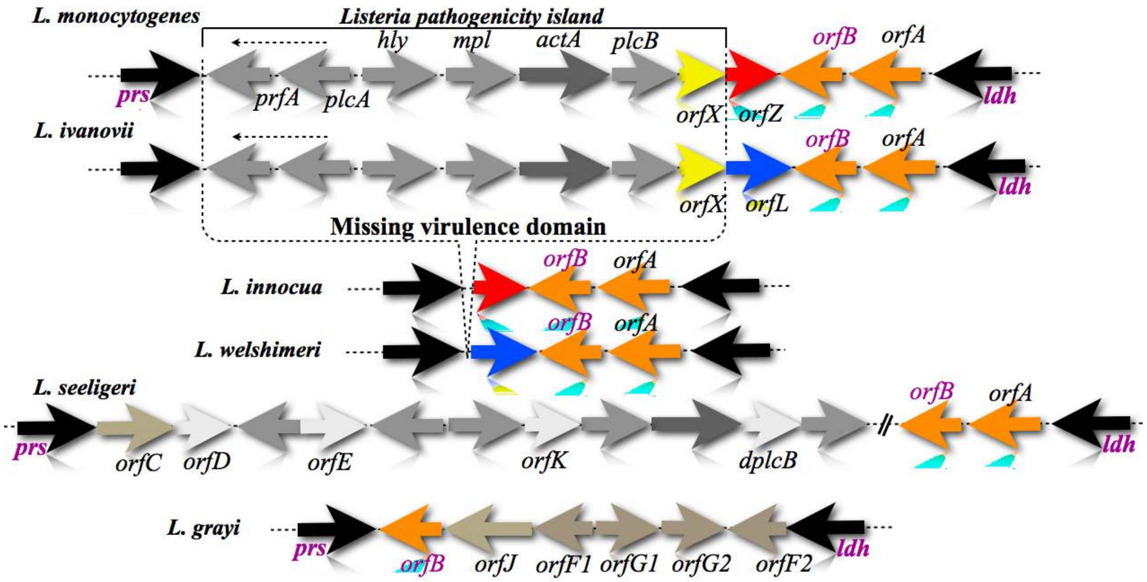
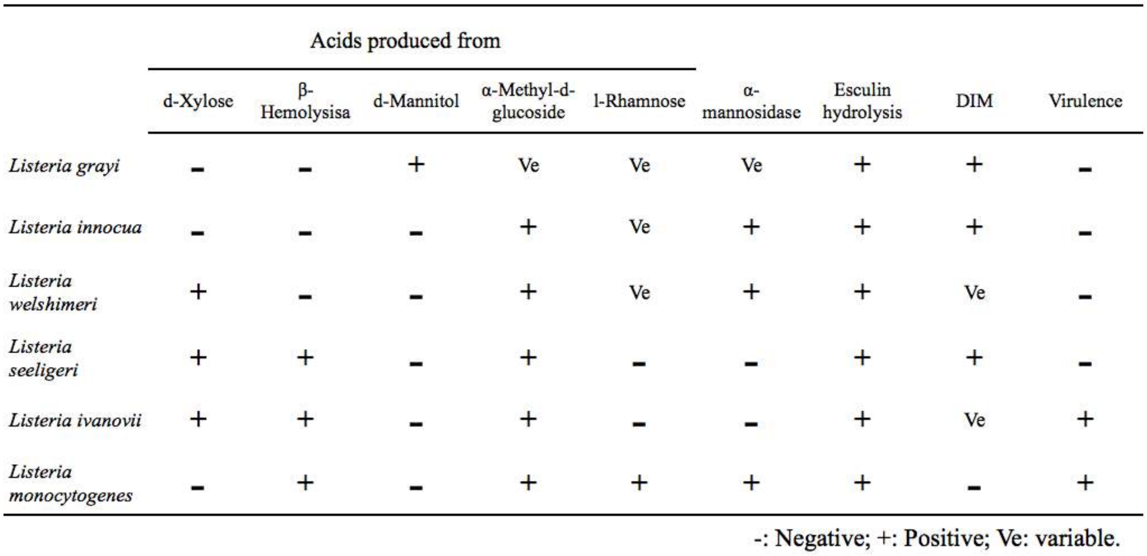


Figure 1.1: Transcriptional organization of the virulence gene cluster locus in the six *Listeria* species.

The key genes (*prfA, plcA, hly, mpl, actA*, and *plcB*: colour coded grey) are responsible for intracellular parasitism and are physically linked in a 9-kb chromosomal island: the *Listeria* pathogenicity island 1(LIPI-1). This figure was adapted from Schmid, M. W. et al., 2005; Vazquez-Boland, J. A. et al., 2001 with modification.

Generally, organisms in this genus express close morphological and biological similarities but varied pathogenicity (Vazquez-Boland, Dominguez-Bernal et al. 2001). Therefore, it is important that pathogenic and nonpathogenic *Listeria* species are adequately differentiated. Several molecular and phenotypic (such as biochemical and serological tests) procedures have been implemented for the identification and differentiation of *Listeria* species. Some of which include: multiplexed PCR (Bubert, Hein et al. 1999 ; Liu 2006), hemolysis, acid production from D-xylose, L-rhamnose, alpha methyl-D-mannoside, and mannitol tests, as depicted in Table 1 (Bille, Catimel et al. 1992; Allerberger 2003; Rocourt and Buchrieser 2007).

Table 1.1: Identification of Listeria species by biochemical tests and their virulence.



	Acids produced from								
	d-Xylose	β-Hemolysisa	d-Mannitol	α-Methyl-d-glucoside	l-Rhamnose	α-mannosidase	Esculin hydrolysis	DIM	Virulence
Listeria grayi	-	-	+	Ve	Ve	Ve	+	+	-
Listeria innocua	-	-	-	+	Ve	+	+	+	-
Listeria welshimeri	+	-	-	+	Ve	+	+	Ve	-
Listeria seeligeri	+	+	-	+	-	-	+	+	-
Listeria ivanovii	+	+	-	+	-	-	+	Ve	+
Listeria monocytogenes	-	+	-	+	+	+	+	-	+

-: Negative; +: Positive; Ve: variable.

1.2.2 History of the genus *Listeria*

The origin of the discovery of these intracellular microorganisms dates back a century. In 1911 the Swedish microbiologist G. Hülphers isolated a bacterium causing liver necrosis in the rabbit and because of its characteristic affinity to the liver, the organism was named *Bacillus hepatitis* (Hulphers 1911). The first officially documented record of the pathogen was when Murray et al. isolated a 1–2 μm long, 0.5 μm wide, round-ended Gram-positive rod as the etiological agent of a septicemic disease affecting rabbits and guinea pigs in their laboratory in Cambridge, United Kingdom (Murray, Webb et al. 1926). This organism was given the name *Bacterium monocytogenes* based on the potent monocytosis-producing activity this microorganism had in rabbits and guinea pigs. In 1927, Pirie isolated the same organism from a gerbil suffering from "Tiger River" disease. The causative intracellular pathogen isolated was named *Listerella hepatolytica*, a generic name in honour of the English pioneer of sterile surgery, Joseph Lister. Because the name Listerella was generally given to a group of slime molds in the 1906, Pirie changed the name in 1940 to *Listeria monocytogenes* (Breed, Murray et al. 1948).

1.2.3 Characterization of the genus *Listeria*

The genus *Listeria* are a group of closely related micro-organisms with a low GC DNA content (36-42%) and measuring about 0.5 μm in width and 1-1.15 μm in length (Liu 2006). They are Gram-positive, non-spore-forming, motile (at 20-25^{0}C), catalase positive, rod shaped and facultative-anaerobic bacteria. These intracellular pathogens are ubiquitous and are generally considered to be hardy bacteria. They are exquisitely adaptable environmental bacteria capable of surviving and multiplying in a wide range of environmental temperatures (0 to 45^{0}C), high salt concentrations (10% NaCl) and at a broad pH range (4.5 to 9) (Hain, Chatterjee et al. 2007). Because of their ability to invade and proliferate under adverse environmental conditions, these bacterial pathogens have been isolated from a wide variety of origins including healthy individuals and animals, soil, water, decaying plants, meat, faeces, dairy products and food processing plants (Ivanek, Grohn et al. 2006). Since these micropathogens are widely distributed in the environment, humans and animals are likely to come in contact with the pathogens. Upon infection, the intracellular pathogens manipulate their way through the host defense mechanisms and propagate their survival within host cells. If the presence of the bacteria within the host is not controlled, it may lead to the severe disease state of listeriosis, especially in hosts with impaired immune systems.

1.2.4 Human Listeriosis

Listeriosis is a severe food-borne disease generally caused after consumption of food contaminated with bacteria from the genus *Listeria* (Allerberger and Wagner 2010). Upon ingestion, the microorganism crosses the intestinal barrier and later disseminates from the mesenteric lymph nodes to the spleen and the liver (Figure 1.2). If not properly controlled by the immune surveillance system or antibiotic administration, the bacteria will migrate to other organs such as the brain or the placenta, resulting in fatal disease (Lecuit 2007).

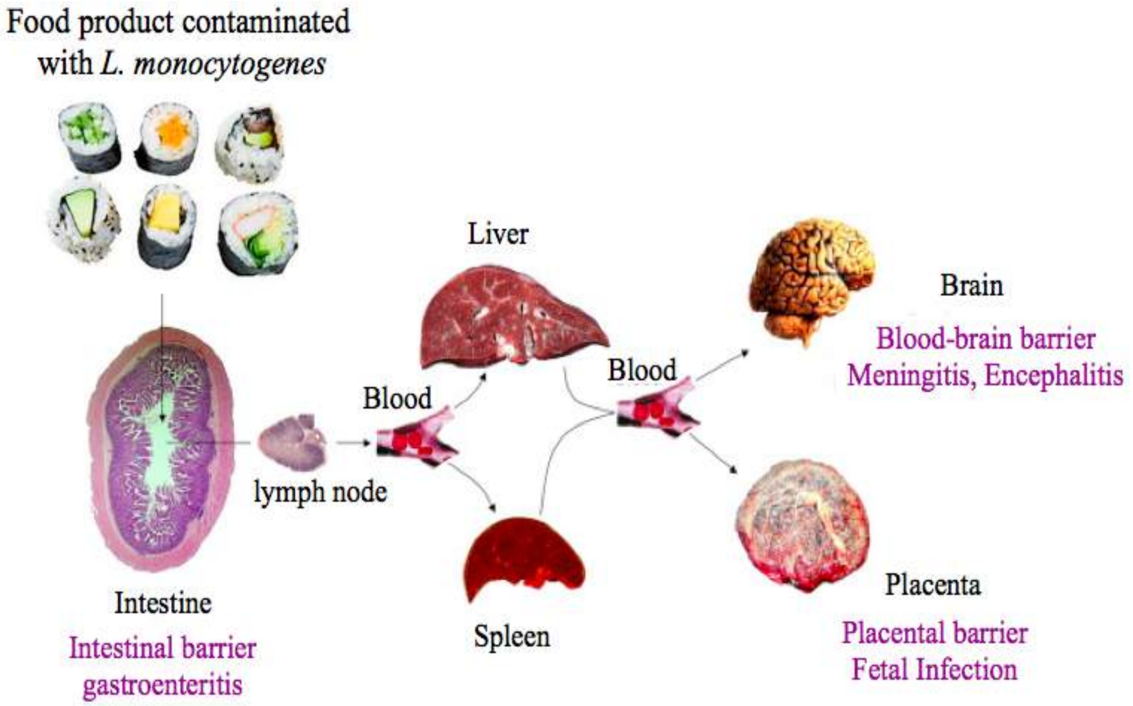


Figure 1.2: Stages in Human listeriosis.

Successive stages of human listeriosis affecting the different organs in humans. Adapted from Lecuit 2007.

The disease is generally considered to be contracted from food contaminated by the pathogen, although there have been reports attributing infection to other modes of transmission. For instance, in 1989 an outbreak of nosocomial listeriosis was reported in Costa Rica involving the use of contaminated mineral oil for cleaning newborn babies after delivery (Schuchat, Lizano et al. 1991).

Listeriosis is generally classified into two forms: non-invasive listeriosis and invasive listeriosis. In non-invasive listeriosis, which generally occurs after infection of a healthy person, the disease is mild and causes symptoms including gastrointestinal illness, fever, vomiting and diarrhea. Meanwhile, invasive listeriosis can be more severe leading to serious complications, such as septicemia, meningitis, meningoencephalitis or febrile gastroenteritis and can cause stillbirths and abortions (Ramaswamy, Cresence et al. 2007; Cossart and Archambaud 2009). Invasive listeriosis is fairly common among newborn children (100 cases per million), the aged (14 cases per million), and pregnant women (Lecuit 2007; Poulsen and Czuprynski 2013). Also, individuals with an immunocompromised system, such as bone marrow transplant patients, blood-borne cancer patients and those with full-blown AIDS are 20 to 40 times more vulnerable to the infection than a healthy individual (Patil, Nadiger et al. 2007; Ramaswamy, Cresence et al. 2007).

Despite efforts by research organizations and food industries to reduce the incidence of the disease, infection as result of this pathogen remains a threat to human health. This disease has been of serious concern in most European countries as well as in the United States (Goulet, Jacquet et al. 2006; Goulet, Hedberg et al. 2008). For example, in the United States, approximately 2,500 individuals are diagnosed with listeriosis with an estimate of 1,520 hospitalizations and 200 to 500 related deaths each year (Mead, Slutsker et al. 1999; Nadon, Woodward et al. 2001; Cartwright, Jackson et al. 2013). Isolation and characterization of the microorganism from infected humans and animals indicated *Listeria monocytogenes* (*L. monocytogenes*) is the principle *Listeria* strain responsible for listeriosis.

1.3 *L. monocytogenes*

1.3.1 General life cycle of *L. monocytogenes*

L. monocytogenes is an opportunistic intracellular bacterial pathogen that has evolved specialized adaptor molecules that interact with specific host cell receptors to facilitate its internalization into host cells (Braun and Cossart 2000). After internalization, the bacterium is transiently trapped in a tight phagosomal compartment before accessing its replicative niche. In this compartment, the bacteria begins recruiting bnactin filaments for generating an actin-“rocket” tail that it uses to propagate itself within the host cytoplasm then to access neighboring cells (Cossart and Lecuit 1998; Schuppler and Loessner 2010). Extensive studies have been conducted on several of the bacteria proteins (virulence factors) facilitating the intracellular life style of the pathogen.

Virulence factors playing important roles during invasion and proliferation of *L. monocytogenes* have been identified (Weiglein, Goebel et al. 1997; Cossart and Lecuit 1998; Schuppler and Loessner 2010). These factors include: Internalins (Inl), hemolysin (Hly), the actin assembly-inducing protein (ActA) and the Phospholipase

(Plc), which are regulated by the bacteria transcriptional regulatory factor PrfA. During invasion, the bacteria uses bacterial surface proteins, called Inl or invasins, which exist in two forms, InlA and InlB, to induce its uptake into the host cell. lnlA and lnlB facilitate the entry of the pathogen by activating signaling pathways via interaction with host cell membrane-bound receptors. InIA binds the cell adhesion protein E-cadherin (E-cad), whereas InIB binds to and activates the MetA tyrosine kinase receptor (Marino, Banerjee et al. 2002). These interactions lead to the internalization of the pathogen into the host cell phagosomal compartment (Marino, Banerjee et al. 2002). After internalization, the pathogen produces the endotoxin *hly*, which generates Listeriolysin O (LLO). LLO hydrolyses the phagosomal membrane to allow the bacteria to into the cytosol. LLO not only mediates the escape of the pathogen from the phagosome into the cytosol but is also required for the pathogen escaping, with the help of *plc* and ActA, from a double-membraned vacuole established during cell-to-cell spread (Pentecost, Otto et al. 2006).

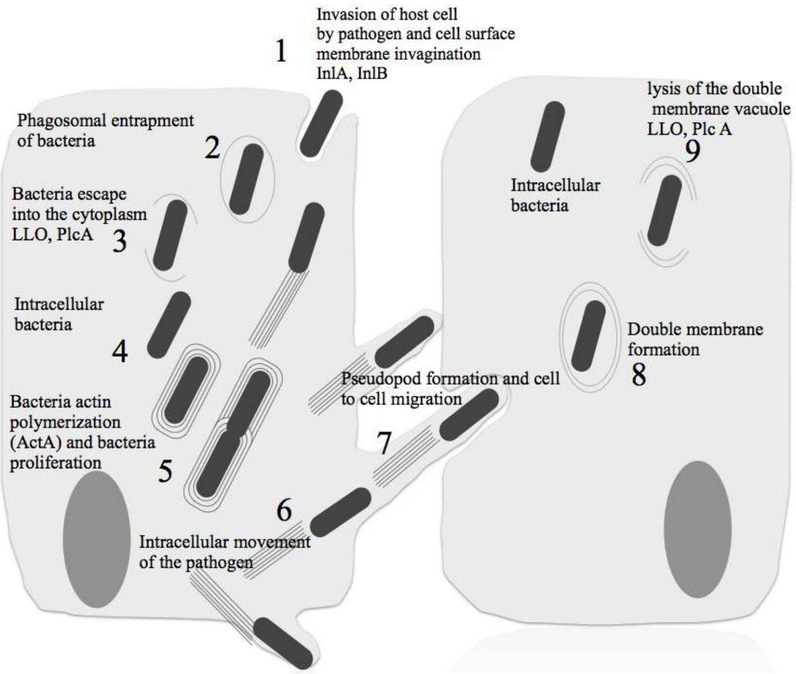


Figure 1.3: Invasion and intracellular life cycle of *L. monocytogenes*.

The invasion and intracellular life cycle of *L. monocytogenes* can be grouped into 9 stages: [1] Cell membrane invagination, [2] Bacterial entrapment by cell phagosomal membrane, [3] Phagosomal membrane disintegration, [4] Freed cytoplamic bacteria, [5] Actin recruitment and replication, 6. Bacteria intracellular movement mediated by the actin rocket tails, [7] Cell-to-cell spread by formation of listeriapods, [8] Double membrane formation, and [9] Lysis of the double membrane vacuole by LLO and PlcB releasing a free bacteria in the adjacent cell where the cycle is repeated. Adapted from Tilney & Portnoy 1989 with modifications.

It is therefore critical to understand how this pathogen manipulates the host cell to induce its uptake, survival, replication and cell-to-cell migration. Beginning with host cell receptor binding and phagocytosis, the pathogen processes through multiple cellular compartments and pathways. Once inside the cell, the bacteria are protected from immune surveillance and upon escaping the phagosome, they hijack the cellular actin machinery to drive themselves into the neighboring cell (Vazquez-Boland, Kuhn et al. 2001) as shown in Figure 1.3.

1.3.2 *L. monocytogenes* host cell invasion

The invasive intracellular bacteria *L. monocytogenes* has evolved a versatile arsenal of virulence proteins for its adhesion and internalization in a variety of host cell types. These virulence factors form part of a large family of proteins, the InI family. Also known as adhesins, they are generally characterized by proteins sharing a similar structural organization that includes a variable leucine-rich repeats (LRR) region at their amino (N)-terminal, a highly conserved inter-repeat (IR) region and a carboxylic (C)-terminal domain (Bonazzi, Lecuit et al. 2009). Based on their C-terminal association with the bacteria surface, the InI are usually classified into three different groups; InI covalently linked to the bacteria surface peptidoglycan by the Leu–Pro–X–Thr–Gly (LPXTG) motif, those that are loosely anchored to the bacteria surface via a C-terminal that contains repeats of 80 amino-acids carrying the Glycine-Tryptophan-rich (GW) dipeptide motif at their starting sequence (Jonquières, Pizarro-Cerda et al. 2001) and finally, the InI molecules lacking the anchoring motifs. Though several of these bacterial association proteins have been identified (Seveau, Pizarro-Cerda et al. 2007), InIA and InIB are the only two InI proteins that have been extensively studied and associated with *L. monocytogenes* adhesion and internalization within the host cells (Bierne and Cossart 2007). Despite the structural similarities existing between the two InI molecules, they bind different human host cell receptors to initiate their respective invasion processes (Bierne and Cossart 2007).

1.3.3 Characterization of *L. monocytogenes* internalin molecules

The first characterized InI protein whose function was associated with mediating *L. monocytogenes* anchoring and invasion of host cells was InIA (Gaillard, Berche et al. 1991; Mengaud, Ohayon et al. 1996). InIA consists of approximately 800 amino acids characterized by two functional units; the N-terminal comprising in addition to the signaling peptide, 15 LRR units (A repeats) and the C-terminal comprising of three longer repeat sequences (B repeats). Also incorporated within the InI structure is a cell wall anchor sorting motif, the LPXTG motif and a 20 amino acid hydrophobic membrane (Bonazzi, Lecuit et al. 2009), Figure 1.4A. The adhesin protein is covalently linked to the bacterial cell envelope by the LPXTG motif through a process catalyzed by the enzyme sortase (Bierne, Garandeau et al. 2004). Because most human cells lack the E-cad receptors, InIA/E-cad mediated entry has been associated with a limited number of human cells, such as epithelial Caco-2 cells (Pentecost, Kumaran et al. 2010).

InIA mediates entry by binding to its host receptor E-cad, a protein first identified as a receptor for InIA by running an extract of Caco-2 cells through a covalently linked InIA affinity column (Mengaud, Ohayon et al. 1996). E-cad is a calcium-dependent transmembrane glycoprotein composed of a short cytoplasmic component of 152 amino acids that can be subdivided into a membrane proximal cytoplasmic domain (MPCD) and a catenin binding domain (CBD), Figure 1.4B. The MPCD is important for p120 catenin binding while the CBD binds the αβ-catenin molecules (Bossche, Malissen et al. 2012). Also incorporated in the domain structure of the E-cad is a transmembrane region and an extracellular component of 555 amino acids that consist of five tandemly repeated cadherin-motif subdomains with calcium binding sites (Serra and Chetty 2008; Bonazzi, Lecuit et al. 2009). They are localized to adherens junctions at the basolateral surface in epithelial cells and are generally involved in cell-to-cell interaction and recognition (Serra and Chetty 2008). The InIA binding site exploits the E-cad signaling pathway to propagate bacterial entry by binding onto a specific epitope located on the E-cad receptor. As revealed by recent studies and X-ray crystal structure, the InIA molecule through its LRR domain interacts with the proline residue in amino acid position 16 (Pro 16) located at the extracellular compartment of the E-cad (Pizarro-Cerdá, Sousa et al. 2004 ; Hamon, Bierne et al. 2006), as shown in Figure 1.4C.

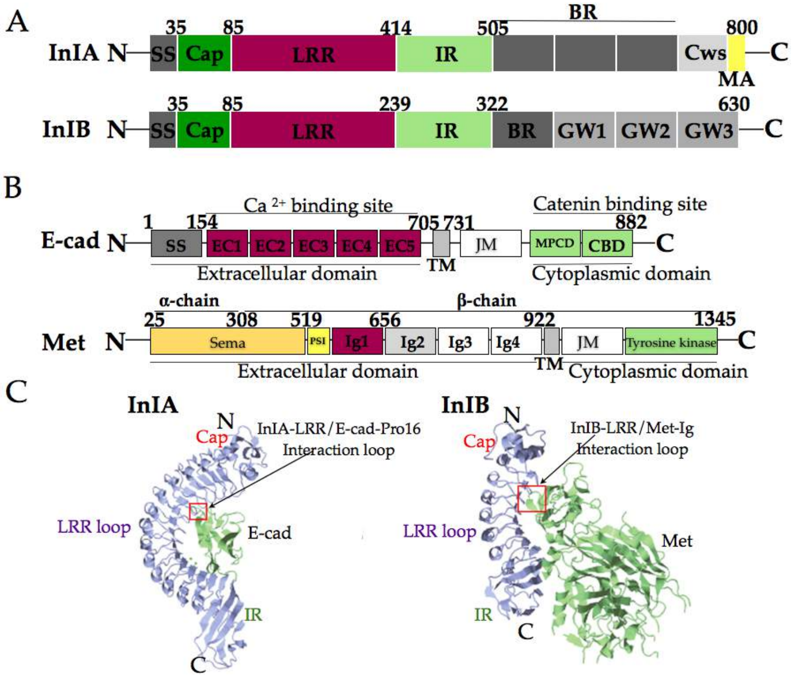


Figure 1.4: Domain representations and crystal structures of InlA/B and their respective cell receptors E-cad and Met.

[A] Domain organization of InlA and InlB. Homologous regions are color coded: the cap domain; green, LRR domain; maroon, inter-repeat region (IR); lime, B-repeats (BR) and signal sequence (SS); dark gray, cell wall spanning region (Cws); light grey, membrane anchor (MA); yellow, Gly-Trp-repeats (GW); grey. [B] Domain organization of the cell E-cad Met receptors. The extracellular domain of the E-cad comprises of five extracellular cadherin repeats (EC) harboring the p16 amino acid that mediates its interaction with the LRR region of InlA while its cytoplasmic domain interacts with cytoplasmic adaptor catenin molecules p120 and α/β catenin to mediate its association to the cell actin cytoskeleton. The extracellular region of the Met receptor comprises of two main domains; the sema and Ig domains. The latter is responsible for mediating binding with the LRR of the InlB. The cytoplasmic domain of the Met harbors the tyrosine kinase that recruits downstream adaptor proteins to initiate signal transduction. Abbreviation TM: transmembrane, JM: juxta membrane, MPCD: membrane proximal cytoplasmic/conserved domain, CBD: catenin binding domain, sema: semaphorin, PSI: plexin-semaphorin-integrin, Ig: immunoglobulin. [C] Crystal structure illustrating the interplay between the InlA/B and their cell surface receptors E-cad and Met. Adapted from Niemann, Jäger et al. 2007; Schubert, Göbel et al. 2001; Cossart 2011) with modifications.

The second well-characterized *L. monocytogenes* adhesin molecule is the InIB bacteria surface protein (Braun, Ohayon et al. 1998; Shen, Naujokas et al. 2000). These adhesin molecules have been reported to mediate bacterial invasion in a wide variety of mammalian cells including, hepatocytes, fibroblasts and epitheloid cells such as HeLa, CHO, L2, Vero, HEp-2, and S180 cells (Pentecost, Kumaran et al. 2010). InIB is a 67 kDa molecule whose primary structure is mainly characterized by two regions of repeats: the C-terminal, comprising of a GW dipeptide motif also known as the cell surface anchor (Csa) domain and the N-terminal which carries the bacterial signaling proteins LRR, IR and the B repeats (Figure 1.4A). The Csa domain is responsible for mediating a loose association of the InI molecule with the bacteria cell surface, mainly through non-covalent interactions with the lipoteichoic acid polymer present on the bacterial cell surface (Jonquieres, Bierne et al. 1999).

InIB mediates host cell invasion by interacting with the hepatocyte growth factor receptor; generally known as the Met receptor (Niemann, Jäger et al. 2007; Seveau, Pizarro-Cerda et al. 2007; Bonazzi, Lecuit et al. 2009). Met is a disulfide-linked heterodimer transmembrane protein composed of an extracellular domain of a 42 kDa α-subunit harboring the InIB binding receptor and a 145 KDa β-subunit spanning the cell membrane harboring a tyrosine kinase catalytic domain on the cytoplasmic region (Furge, Zhang et al. 2000 ; Schubert, Göbel et al. 2001), Figure 1.4B. Phosphorylation of the tyrosine receptor initiates the downstream processes, which eventually lead to the internalization of the bacteria. Unlike InIA where bacteria internalization is solely mediated by the interaction of InIA with the E-cad receptor, InIB mediated internalization is not only achieved through its N-terminal interaction with the Met receptor, but also through its C-terminal (GW domains) interaction with extracellular glycosaminoglycans (GAGs) and with the gC1q-R protein (Marino, Banerjee et al. 2002; Pizarro-Cerdá and Cossart 2009).

1.3.4 Mechanisms of internalization of *L. monocytogenes*

L. monocytogenes invades non-phagocytic host cells through the zipper mechanism (Lecuit, Ohayon et al. 1997; Bierne and Cossart 2002). The mechanism involves the binding of the bacteria surface proteins to the cell-membrane receptors. This interaction triggers a signaling cascade that reorganizes the cell actin cytoskeleton to facilitate internalization of the pathogen. As previously discussed, *L. monocytogenes* can use two different cell surface proteins, InIA and InIB for its invasion (Figure 1.5 and 1.6).

1.3.5 Model for InI A-mediated internalization of *L. monocytogenes*

InlA-mediated internalization occurs when the protein binds to its cell surface receptor, the E-cad. As revealed by genetic and biochemical studies, the major interaction between the bacterial surface protein and the cell surface receptor is site and domain specific. The interaction occurs between the pro 16 region located on the outermost extracellular domain of the E-cad and the LRR/IR domain of the InIA (Figure 1.4). The LRR is believed to interact directly with E-cad, while the IR region is required for a proper folding of the LRR region (Lecuit, Ohayon et al. 1997). Interaction between InIA and its main signaling receptor E-cad leads to the phosphorylation of the E-cad: a process involving both the lipid raft associated caveolin-1 protein and Src cytoplasmic protein (Pizarro-Cerdá and Cossart 2009). Phosphorylation of the E-cad triggers its ubiquitination by the ubiquitin ligase Hakai which enables the receptor to recruit several molecules particularly members of the catenin family (p120, α- and β-catenin), septin and clathrin leading to the formation of adheren junctions (Mostowy, Nam Tham et al. 2009; Pizarro-Cerdá, Bonazzi et al. 2010; Bonazzi, Vasudevan et al. 2011; Mostowy and Cossart 2011). Recruitment of the α-catenin has been shown to be dependent on ARHGAP10, a Rho-GAp domain protein that associates with the small GTP-binding protein Arf6 (Seveau, Pizarro-Cerda et al. 2007). The formation of the signaling cascade subsequently triggers the rearrangement of the cell actin cytoskeleton and membrane extension, which leads to the invagination of the cell membrane entrapping the bacteria in a tight phagosomal compartment within the cell cytosol (Figure 1.5). It is believed other cytoplasmic proteins including Rac-1, Arf6, Myosin VIIA and vezatin may be involved in the formation of adheren junction/bacteria internalization, albeit their participation in this process needs to be fully established (Pizarro-Cerdá, Sousa et al. 2004).

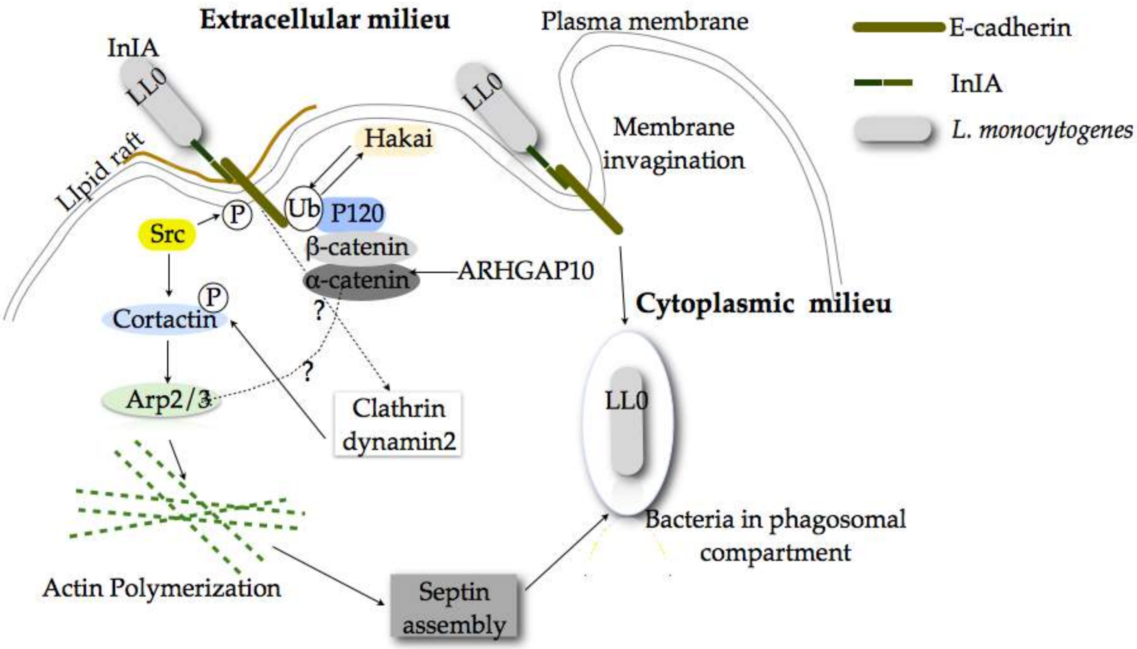


Figure 1.5: InIA-Mediated internalization pathway of *L. monocytogenes*.

Interaction of InIA with the E-cad located within the lipid rafts initiates the activation of some cytoplasmic adaptor proteins, which in turn triggers the rearrangement of the cell actin cytoskeleton and the internalization of the bacteria. Dashed arrows indicate hypothetical steps.

1.3.6 Model for InI B-mediated internalization of *L. monocytogenes*

InIB binds to the extracellular domain of its Met receptor tyrosine kinase and induces multifunctional cellular responses leading to the internalization of the bacteria. Studies have shown InIB can mediate entry into the host cells in two ways; by dissociating from the bacterial surface and its soluble form interacting with its host proteins GAGs, gC1q-R and Met or by interacting with the host cell receptors while conjugated to bacteria. Upon interaction, the bound InIB induces membrane aberration and Met activation. This process triggers a series of downstream biological reactions particularly the phosphorylation of several adaptor proteins, including Gabl, Cbl, Shc and CrkII (Dokainish, Gavicherla et al. 2007). These adaptors then recruit the phosphatidylinositol 3-kinase (PI3K) which in turn activates Rac-1 (Schmeck, Beermann et al. 2006). The activation of Rac-1 orchestrates a signaling cascade via the

Arp2/3 and wave/N-wasp proteins that leads to cell actin rearrangements and the subsequent internalization of the bacteria. Once in the cell cytoplasm, the internalized bacteria are entrapped in the phagosomal membrane (Figure 1.6). Reports suggest the involvement of other cellular effectors such as the Sept2 and PI4K in the internalization process, although their functions need to be characterized (Pizarro-Cerdá and Cossart 2009).

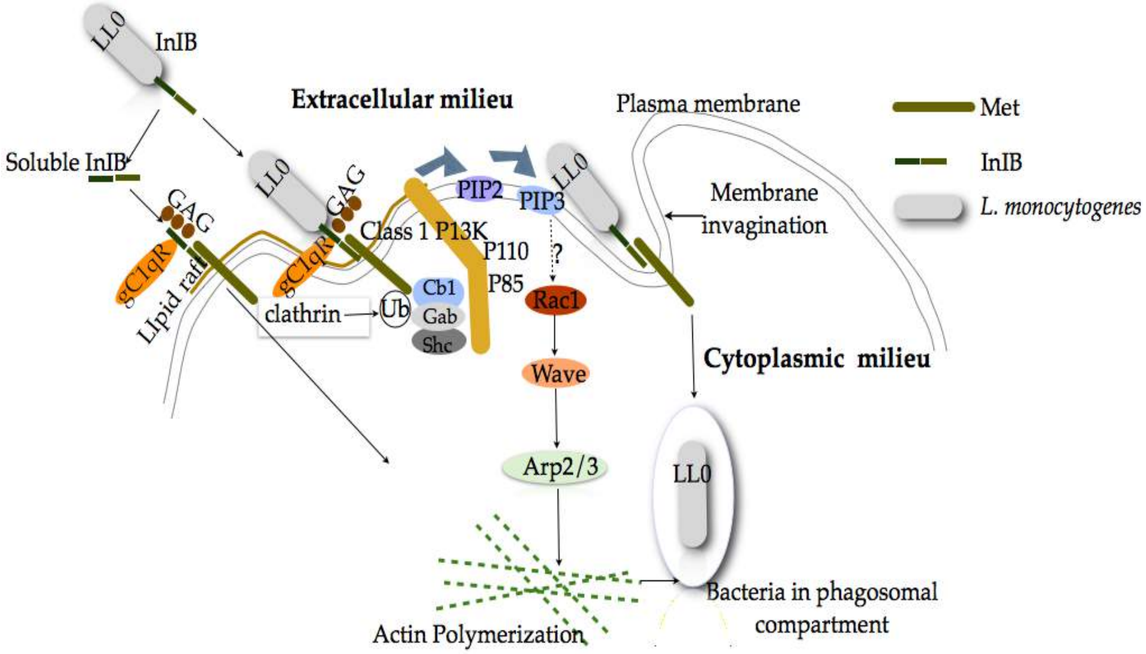


Figure 1.6: InIB-Mediated internalization pathway of *L. monocytogenes*.

Interaction of InIB with the Met located within the lipid rafts initiates the activation of some cytoplasmic adaptor proteins, which in turn triggers the rearrangement of the cell actin cytoskeleton and the internalization of the bacteria. InB can either bind to its host receptor as a soluble molecule or be conjugated to the bacteria cell wall. Dashed arrow indicates poorly understood steps.

1.3.7 Listeriolysin O (LLO) and phospholipases C (PLCs): *L. monocytogenes* escape route from the cell phagosomal compartment

Following internalization, the bacteria are localized in an acidic phagosomal compartment. How the intracellular organism progresses from this point will depend on its ability to subvert or avoid the host autophagy surveillance innate immune system (Py, Lipinski et al. 2007).

Autophagy is a catabolic process put in place by the host cell to target cellular or foreign materials and process them for degradation through the action of the cytolytic lysozyme (Hamon, Ribet et al. 2012). Once in the phagosome, some intracellular bacteria have found a way around the host defense machinery, by activating their virulence protein(s), which firstly delays maturation of the membrane vacuole and later lyses the phagosomal membrane to release bacteria from the phagocytic vacuole into the cell cytosol. *L. monocytogenes* has adopted the latter process to evade autophagic surveillance and promote its intracellular growth within the host cell (Tweten 2005). *L. monocytogenes* achieves this process by secreting its pore-forming cytolysin listeriolysin O (LLO) and two phospholipases C (PLCs) (Gründling, Gonzalez et al. 2003), that are both largely responsible for mediating perforation of the phagosomal membrane, thereby allowing the bacteria to access its replicative niche in the cytoplasm (Kayal and Charbit 2006; Schnupf and Portnoy 2007).

1.3.8 Characterization of listeriolysin O

L. monocytogenes LLO belongs to a superfamily of thiol-activated cholesterol-dependent pore-forming cytolysin (CDCs) which includes perfringolysin O (PFO) and streptolysin O (SLO). The three dimensional (3D) monomeric structure of LLO has not yet been determined. However, based on the existing sequence similarities between members of the CDC family (40-70 %) and the known 3D structures of PFO and SLO (Rossjohn, Fell et al. 1997), it is assumed with a reasonable degree of certainty that the LLO monomer (50 KDa-80 KDa) shares the same structural organization as PFO and SLO. Figure 1.7 depicts a proposed 3D model of the LLO.

Despite their sequence similarities, LLO is unique among members of the CDCs in that it has a low pH activity range (5.5 to 6) which permits LLO to induce pores in phagosomal membranes, causing a passive flux of ions and macromolecules without destroying the host cell (Dramsi and Cossart 2002). Four domain regions, based on the PFO structure, were used to characterize the monomeric structure of LLO. The first domain is located on the N-terminal, the C-terminal is comprises of the conserved fourth domain while domains two and three are located between them. Found on the fourth domain are two conserved loops: the tryptophan rich undecapeptide motif made of 11 amino acids (ECTGLAWEWWR) and the cholesterol binding motifs that are believed to be crucial in recognizing the cholesterol raft within the vacuolar membrane. Located on domain three are two sets of α-helices transmembrane β-hairpins (TMH 1 and 2) that make up the β-barrel of the pore complex as well as the pH sensor acidic triad D208, E247 and D320.

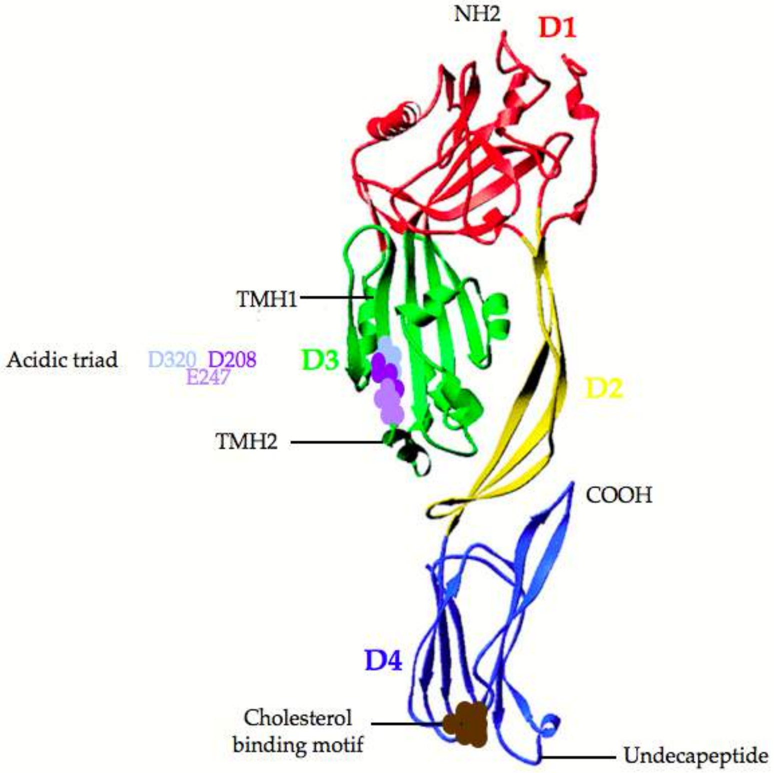


Figure 1.7: Ribbon structure of the water-soluble LLO monomer.

Four domains (D1-D4) characterize the monomer as depicted. D3 comprises of two transmembrane β-hairpins (TMH 1and 2), which are responsible for forming the β-barrel of the pore complex while D4 comprises of two conserved loops, the cholesterol binding and undecapeptide motifs that are responsible for binding the cholesterol on the cell membrane. Relevant amino acids (D208, E247 and D320) within D3 are highlighted by space filled models of their side chains. The structure is based on the PFO crystal structure (Rossjohn, Fell et al. 1997).

1.3.9 LLO in host cell signaling

In addition to its ability to induce pore formation, LLO has been reported to mediate several other cellular processes that enable prolonged survival of the pathogen in the host cell. These include: activation of the NF-κB, MAPK, calcium, protein kinase-C and phosphatidyl-inositol signaling pathways (Hamon, Bierne et al. 2006;

Kayal and Charbit 2006). In this regard, LLO alters bacterial uptake, induces mucus exocytosis in intestinal cells, induces apoptosis in dendritic cells, causes degranulation and leukotriene formation in neutrophils, induces pro-inflammatory cytokines and chemokines in macrophages, phosphoinositide metabolism and lipid mediator generation (Vazquez-Boland, Kuhn et al. 2001; Kayal and Charbit 2006; Schnupf, Zhou et al. 2007). Recent experimental evidence has also shown that the bacterial toxin regulates host gene expression through histone modification. This adaptive process is orchestrated by the intracellular pathogen's ability to silence its host immune defenses thus propagating pathogen survival and intracellular growth (Hamon, Batsche et al. 2007; Gekara, Zietara et al. 2010).

1.3.10 *L. monocytogenes*: suggested mechanism of vacuolar escape

Although there is evidence that supports the role of LLO in mediating pore formation within the phagosomal membrane, it is important to note that the actual mechanisms by which LLO-pore formation results in bacterial escape into the cytosol still remain unclear (Tweten 2005). Recent studies have demonstrated that successful bacterial intracellular growth depends on restricting LLO activity to the phagosomal compartment (Schuerch, Wilson-Kubalek et al. 2005). As documented, unregulated expression of the LLO within the bacteria intracellular niche could lead to the perforation of the host organelles and plasma membrane causing cell death and destruction of the intracellular bacterial compartment (Glomski, Decatur et al. 2003). This results in exposing the bacteria to the host surveillance immune system and the subsequent degradation of the bacteria. Though several mechanisms have been suggested to regulate the LLO activity within the phagosome, the best understood rely on its sensitivity to pH and temperature (Schuerch, Wilson-Kubalek et al. 2005). As shown in previous studies (Beauregard, Lee et al. 1997), vacuolar membrane integrity is lost once the phagosome reaches a pH between 4.9 and 6.7. The pH sensitivity of the LLO is regulated by a triad of acidic residues located within domain 3 of the LLO monomer. At neutral pH at temperatures above 30^0C, a rapid denaturation of the LLO occurs, resulting from premature unfolding of the domain 3 TMHs, the structures responsible for the formation of the transmembrane β-barrel pore (Shatursky, Heuck et al. 1999).

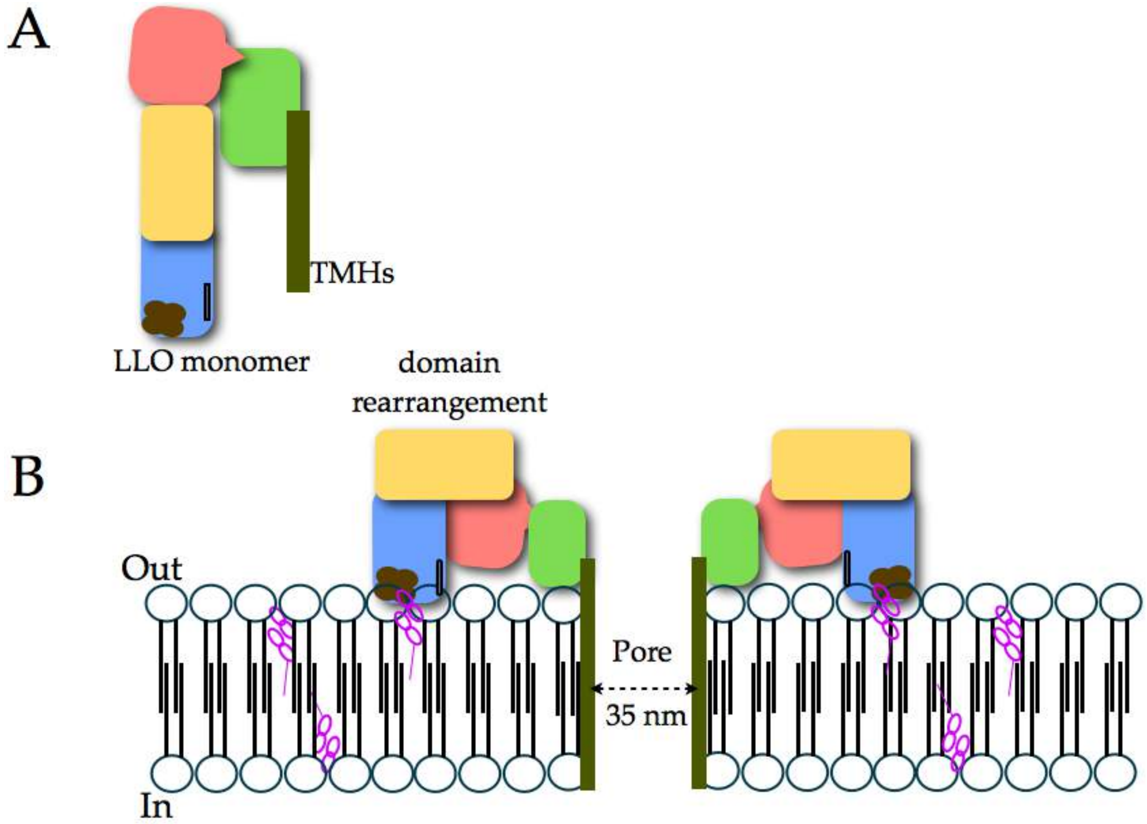


Figure 1.8: Schematic representation of the LLO mechanism of pore formation.

[A] Shows a structural model of the *L. monocytogenes* LLO monomer with each domain color-coded similarly to the structural model depicted in Figure 1.7. [B] Illustrates the mechanism of pore formation. Following secretion, the bacterial LLO monomers bind to the cholesterol (purple ring clusters) within the host membrane's lipid layer and diffuse laterally to oligomerize in a pre-pore complex. A conformational change followed by a vertical collapse in the monomer height allows the TMHs in D3 to insert into the membrane and to form a large transmembrane β-barrel pore which the bacteria then uses to exit into the cell cytosol. Adapted from Hamon, M. A. *et al* 2012 with modifications.

Shortly after accessing the phagosomal acidic milieu, *L. monocytogenes* LLO induces small membrane perforations, which allow the efflux of vacuolar Ca^{2+} leading to an increase in vacuolar pH and a subsequent delay in vacuolar membrane maturation (Henry, Shaughnessy et al. 2006). During this period, it is assumed that LLO binds to the cholesterol rich regions of the host biological membranes using their cholesterol-binding motif located on domain 4 of the LLO. This interaction triggers the toxin monomers to diffuse laterally and homo-oligomerize into a ring-like pre-pore complex.

A structural rearrangement of domain 2 causes two set of α-helices present on domain 3 to unfurl forming extended amphipathic β-hairpins (TMH1 and TMH2), thus allowing the switch from a pre-pore complex into an effective pore. Subsequently, a 40-Å vertical collapse of domain 2 during the conformational transition facilitates the insertion of the β-hairpins into the membrane bilayer to form a large transmembrane β-barrel pore of approximately 35 nm (Figure 1.8).

However, studies conducted using an LLO_{L461T} mutant demonstrated the involvement of other human host factors acting with LLO to facilitate phagosomal escape of *L. monocytogenes* (Hamon, Ribet et al. 2012). These host factors include the enzyme GILT (ϒ-interferon-inducible lysosomal thiol reductase and the CFTR (cystic fibrosis transmembrane conductance regulator) which are both highly expressed in macrophages. GILT knockout studies performed in macrophages demonstrated a delay in bacteria phagosomal escape and *in vivo* GILT knockout mice were more resistant to bacterial infection than wild-type mice (Singh, Jamieson et al. 2008). CFTR on the other hand is a chloride channel located on pathogen-containing phagosomes that acts in concert with LLO to potentiate pore formation by increasing phagosomal chloride concentration (Cossart 2011; Radtke, Anderson et al. 2011).

Notably, the requirement of LLO in vacuolar lysis has been demonstrated to be cell specific (Gründling, Gonzalez et al. 2003). Several human-derived cell lines including: WS1, Henle 407, HeLa, HEp-2 and human dendritic cells have recently been identified as cells in which LLO is not required by the *L. monocytogenes* to facilitate their escape into the cell cytosol (Portnoy, S. et al. 1988; Paschen, Dittmar et al. 2000; Hense, Domann et al. 2001; O'Riordan, Yi et al. 2002). However, in the above listed cell lines, bacterial escape into the cytosol is mediated by the bacterial PLCs (Gründling, Gonzalez et al. 2003).

1.3.11 Vacuolar escape is enhanced by two bacterial phospholipases C

Apart from the LLO mediated phagosomal escape, two bacteria phospholipases C; phosphatidylinositol-specific and broad-range phospholipases C, PI-PLC and PC-PLC, have been reported to facilitate rupturing of the phagosomal membrane (Schnupf and Portnoy 2007). PI-PLC is secreted by the bacteria in an active state and catalyses the production of inositol phosphate (IP) and diacylglycerol (DAG) through cleavage of the membrane phospholipids. PC-PLC is generally secreted as an inactive 33-KDa proenzyme and cleaved to an enzymatically active 28-KDa form by the bacterial metalloprotease (mpl) enzyme (Smith, Marquis et al. 1995; Gründling, Gonzalez et al. 2003). Studies conducted in bone marrow derived macrophages (BMDMØs) following the deletion of PI-PLC resulted in a 15 % decrease in the escape efficiency of the bacteria. A similar study carried out using PC-PLC mutant in BMDMØs, had no effect on bacteria phagosomal escape but the virulent activity of the bacteria was decreased by 20 fold (Camilli, Tilney et al. 1993). Notably, when both PI-PLC and PC-PLC were deleted, the bacterial ability to escape from the phagosomal compartment was more impaired than with either mutation alone, suggesting the overlapping involvement of both PI-PLC and PC-PLC in phagosomal disruption (Smith, Marquis et al. 1995). Though studies have demonstrated the involvement of the PLCs in bacteria phagosomal escape, the actual bacteria/host protein interactions leading to this process remain rudimentary.

1.3.12 Actin-based intracellular motility of *L. monocytogenes*

Dissemination of *L. monocytogenes* within the host cell cytosol is a key step in its intracellular life cycle (Vazquez-Boland, Kuhn et al. 2001; Ireton 2013). As previously discussed, following internalization and entrapment of the bacteria pathogen within the phagosomal compartment, by means of its pore forming virulence factors (LLO and PLCs), the bacterium accesses the cell cytoplasm where it initiates recruitment of F-actin. Recruitment of the F-actin cloud is triggered by the bacterial surface protein ActA, which stimulates polymerization of actin monomers (G-actin) following its interaction with several host cellular proteins (Ireton 2013). Actin filament polymerization generates an actin “rocket” tail that allows for the directional force, actin-based motility that propels the bacteria through the cells’ cytoplasm. Intracellular movement eventually propels the bacteria to the peripheral membrane of the host cell where it pushes the membrane outward, forming distinct filopods. Adjacent cells often digest the filopods, and the cycle is repeated. It is important to note that the actual molecular events leading to bacterial dissemination as well as filopod formation are still under investigation.

1.3.13 Characterization of the bacterial ActA protein, the mammalian Arp2/3 complex and VASP proteins

Intracellular movement of *L. monocytogenes* within the host cytoplasm is achieved following polymerization of the host actin on the surface of the bacteria. The bacteria surface protein ActA orchestrates the hijacking of the host actin network. ActA is a 610-amino acid protein and comprised of three main domains; the N-terminal domain that binds and activates the host Arp2/3 complex, the proline rich repeats that recruit proteins from the vasodilator-stimulated phosphoprotein (VASP) family and a C-terminal domain that anchor the protein to the bacterial surface (Travier, Guadagnini et al. 2013), (Figure 1.9).

The human receptor for bacterial ActA is the Arp2/3 protein complex (Welch, Rosenblatt et al. 1998). The Arp2/3 complex is generally located in cellular regions characterized by dynamic actin filament activity, it plays a major role in host actin regulation (Humphries, Balcer et al. 2002). The Arp2/3 complex is composed of seven conservative subunits (ARPC1-5, Arp2 and Arp3) with Arp2 and Arp3 harboring the surface features required for dimerization and to initiate an actin filament growing in the barbed end direction (Beltzner and Pollard 2004). The complex is intrinsically inactive in actin nucleation and branching and becomes activated by members of the Wiskott-Aldrich-syndrome family protein (WASP, N-WASP, WAVE, and WASH) (Beltzner and Pollard 2004). The C-terminal region of the WASP share the same sequence similarity to the 144 to 162 amino acid residues within the N-terminal of the *L. monocytogenes* ActA (Pistor, Gröbe et al. 2000). Following stimulation, the Arp2/3 complex through side branching facilitates nucleation of monomeric actin from existing filaments (Schoumacher, Goldman et al. 2010; Firat-Karalar and Welch 2011).

In addition to the Arp2/3 complex, other cellular proteins especially proteins belonging to the Enabled/vasodilator-stimulated phosphoprotein (Ena/VASP) family have been identified to be involved in *L. monocytogenes* actin polymerization. This was confirmed by studies where deletion of VASP resulted in a slower rate of bacterial movement (Smith, Theriot et al. 1996; Haas, Kusinski et al. 2007). Members of the Ena/VASP family (Mena, n-Mena, EVL and VASP) contain a number of conserved proline-rich domains spread throughout their central regions. They are responsible for binding cytoplasmic profilin, which forms a 1:1 complex with the monomeric actin as well as concentrating the profilin-actin complex near the sites of actin polymerization (Machesky 1997).

1.3.14 Model for *L. monocytogenes* actin assembly

L. monocytogenes initiates the process of actin polymerization by releasing its ActA protein. The ActA directly interact with, and activate, the Arp2/3 complex. The Arp2/3 complex is responsible for recruiting monomeric actin to the barb end of the bacteria. The cycle of assembly of actin filaments is predicted to include seven steps: [1] free profilin in the cytoplasm binds to actin monomer; [2] profilin–actin complex binds to VASP protein; [3] VASP-profilin-actin complex binds to the proline rich domain of the bacterial ActA, [4 and 5] VASP hands over the profilin-actin complex to the Arp2/3 complex where nucleation and elongation of the actin tail on the bacteria surface takes place, alternatively [6], profilin–actin may add directly to the barbed end of the actin filament, possibly by being handed over to the Arp2/3 complex, which binds both profilin and actin; or [7] profilin dissociates from the barbed end complex and may be released to restart the cycle as illustrated (Figure 1.9).

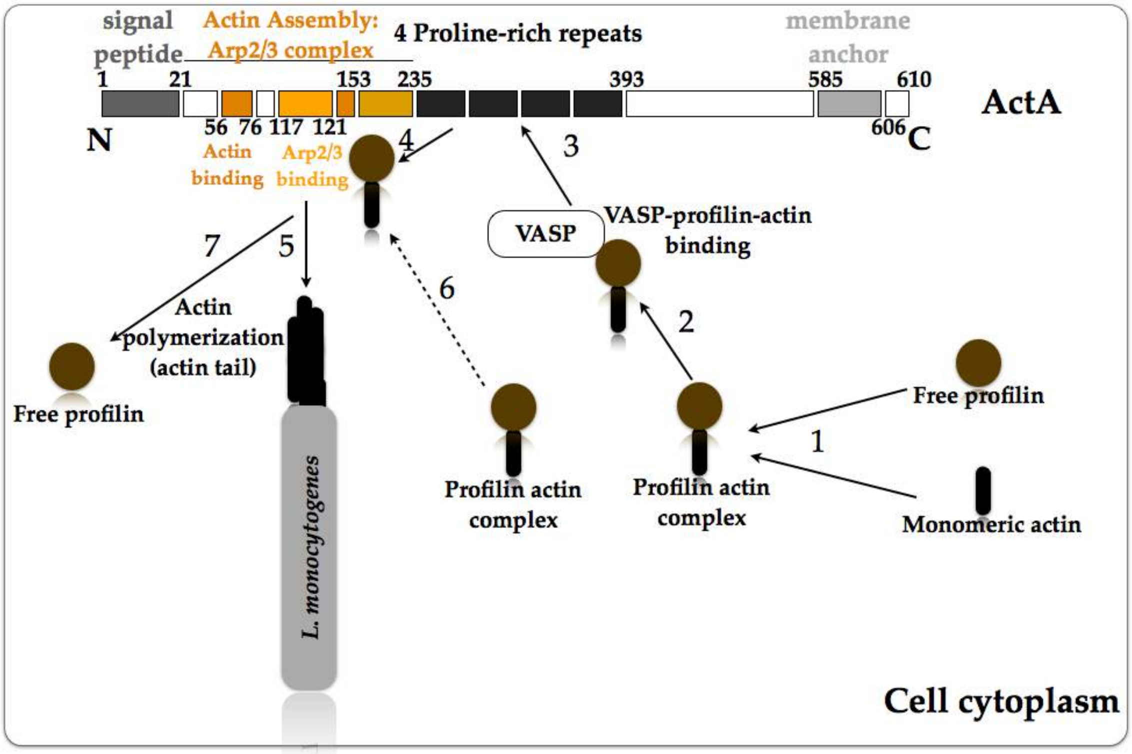


Figure 1.9: Model for actin assembly by ActA of *L. monocytogenes.*

In the cytoplasm of an infected cell, *L. monocytogenes* ActA protein serves as a scaffold, recruiting host proteins leading to the formation of an actin-rich "comet tail" on the bacterial surface. Briefly, free circulating profilin binds to actin monomer and the VASP protein then recruits the complex. The VASP-profilin-actin complex is then bound to the proline rich domain on the ActA. Binding of VASP to the bacterial ActA protein moves the profilin-actin complex closer to the Arp2/3 complex where actin polymerization and nucleation take place.

1.3.15 Conclusion

For the past twenty years, *L. monocytogenes* has been the stereotypical bacterium used to understand how intracellular pathogens invade and cause pathological conditions in humans (Cossart 2011). Although there is significant knowledge on the virulence proteins used by the intracellular pathogen to invade and subvert the host defense machinery (Vazquez-Boland, Kuhn et al. 2001; Poulsen and Czuprynski 2013), it is unclear how the bacterial proteins interact with their host receptors to trigger the expected responses. Several models have been established to explain bacteria/host proteins interactions leading to the bacteria entry, phagosomal escape and actin polymerization (Lee and Liu 2008; Cossart 2011). However, based on the literature discussed above it is without doubt that to fully understand how this intracellular pathogen invades and circumvents the host immune system, more of the human adaptor proteins involved in the process need to be established.

A major challenge faced by researchers in the past was the unavailability of appropriate tools to carry out their research work. With the introduction of new biological techniques such as reverse transfection and the RNAi technology (Ziauddin and Sabatini 2001), researchers are now better equipped to study and understand the bacteria/host protein interaction at cellular and sub-cellular levels. A major challenge using the latter method is the time it takes to screen a pool of siRNAs that usually comprises of hundreds of thousands of individual siRNA molecules. There is therefore a rising need for new and robust experimental approaches that will facilitate the easy but rapid identification of human host factors manipulated as a result of the bacteria infection.

CHAPTER 2

THESIS OBJECTIVES

2.1 Specific objectives

Specific objective [1] is to validate the screening technology using known model/control siRNAs, to determine if it is suitable for screening using gene silencing.

Specific objective [2] will extend this: to show if a cohort of arrays retains similar characteristics, therefore demonstrating that the production process itself is robust and suitable for screening.

Specific objective [3] will then test the arrays in screening. This will be by applying arrays to screen a well-known cell based model: TNF-α induced nuclear translocation of RELA/p65. This is a well-understood pathway and accessible to image analysis. If successful, this will validate the arrays as a screening platform.

Specific objective [4] is to develop a cell-based assay of *Listeria* infection then screen this assay to identify host factor genes required for infection.

2.2 Brief description of each objective

2.2.1 Objective 1: Validation of gene silencing and reverse transfection

This objective was geared at assessing the potency of gene silencing achieved using reverse transfection of cells from printed spots. A major challenge generally encountered in this technology is developing a suitable transfection encapsulation mixture for the siRNA for adequate spotting and delivery of the siRNA by reverse transfection into the cells.

2.2.2 Objective 2: siRNA array production validation

The second objective demonstrated whether a cohort of glass chips retain similar gene silencing characteristics. Therefore demonstrating that the production process itself is robust. For this validation, > 2000 siRNA spots comprising of known controls were printed onto a glass chip and that was repeated for more than 1500 contiguous copies of the arrayed siRNA in a single print cycle. Gene silencing, spot size and spot intensity were sampled across the entire array production. The latter quantification parameters were introduced to check for printing accuracy.

2.2.3 Objective 3: siRNA array validation in screening: Identification of genes implicated in the regulation of the TNF-α induced p65 (NF-κB) signaling pathway using high throughput cell based RNA interference

The third objective was to demonstrate that the arrayed siRNA on glass chips were effective at defining genes involved in a cellular pathway through screening. TNF alpha induced nuclear translocation of RELA/p65 was selected as the model, since it is well-understood and accessible to image analysis. Furthermore, the pathway is significantly well understood compared to *Listeria* and therefore a suitable validation model.

Arrays of siRNA comprising of all kinases, phosphatases, ubiquitinylases, and proteases in the human genome were printed on glass chips. These were used to identify proteins implicated in the TNFα induced p65 (NF-κB) signaling pathway. The screen comprised a total of 32 replicates, and > 100,000 individual extracted and analyzed phenotypic experiments. Each experiment consisted of image analysis of cells exposed to a single siRNA spot for 48 h, then stimulated, and analyzed individually for the phenotypic effect of the reverse transfected siRNA on RELA/p65 nuclear translocation.

The screen yielded 51 hits affecting the pathway, of these, 26 (51%) were identified as direct interactors with the TNF to NF-κB pathway, including two of its known regulators CHUK and IKK. This validated the glass chip siRNA technology, as well as the entire image-handling pipeline from cell handling, staining through to acquisition and analysis as a powerful platform for functional genomics and high content screening.

2.2.4 Objective 4: Screening for genes involved in *L. monocytogenes* infection of cells

The final objective is to apply arrayed siRNA screening to identify genes implicated in *Listeria* infection.

The first sub-objective was to develop a cell-based assay of *Listeria* infection. For assay development, two well-characterized human cell lines were used: HeLa and U2OS. A human cell model is required because the screening library silences the human genetic repertoire but not other species. The assay used a *L. monocytogenes* strain that expresses green fluorescent protein (GFP) to monitor cell infection by confocal imaging. Algorithms were developed to measure intracellular bacteria and changes in cell surface morphology from images of living cells exposed to a serial titration of *Listeria* and varying infection times. Based on the statistical measurements obtained, the optimal *Listeria* multiplicity of infection was determined [MOI: 25] and time of infection was set at 3 h.

The assay was then combined with the glass chip arrays to screen for human host factors required in *Listeria* infection. The screen was performed on 15 replicate glass chip arrays, comprising a total of $> 40,000$ individual experiments. Data were extracted and analyzed from which 67 human host factors out of 3150 genes were identified to potentially play a role in the bacterial infection.

CHAPTER 3

EXPERIMENTAL PROCEDURES/MATERIALS AND METHODS

3.1 Introduction

There have been major advances in molecular biology techniques in the past few years that have led to the discovery of new analytical tools to study and manipulate human host factors at the subcellular level. Two such techniques include siRNA-mediated RNA interference (RNAi) screens in which the effects of gene silencing on biological phenotypes can be systematically explored and reverse transfection in which nucleic acid (siRNA, miRNA, shRNA, DNA) coupled to a biological substrate is taken up by the cells. When coupled, these techniques generate a powerful experimental approach that seeks to identify possible regulators implicated in general biological processes: cell size, shape and cell proliferation, mitosis or the efficiency of viral/bacterial infection. In this chapter, all the chemicals and general molecular biology techniques that were used throughout the thesis are described.

3.2 Chemicals

All fine chemicals were purchased from Sigma-Aldrich (SA). DRAQ5 was from BioStatus (Shepshed, UK). All siRNA duplexes were purchased from Dharmacon (USA). The SMARTPool human whole-genome siRNA library was purchased from Thermo Scientific Dharmacon (West Lafayette, CO. USA). Primary antibodies were purchased from Santa Cruz Biotechnology (CA) and fluorescent secondary antibody was purchased from Molecular Probes, Invitrogen (Carlsbad, CA). Transfection reagents were purchased from Qiagen (Valencia, CA). All culture media and their supplements were purchased from Gibco (Invitrogen, Carlsbad, CA).

3.3 siRNAs array printing

The siRNA library comprises a collection of pools of 4 individual siRNAs. Each pool is designed to target a single gene as a mixture of four siRNAs per target gene. The library comprises a total of 18,174 pools of siRNAs. The siRNA library is supplied as lyophilized siRNA in 96-well plates with each well containing 0.5 nMol of the dried nucleotide. Lyophilized siRNAs were resuspended in a 1X siRNA buffer to a final concentration of 12.5 μM, sealed and stored at -20^{0}C until required for printing. The siRNA microarray printing solutions were prepared following protocols adopted from methods previously described (Erfle and Pepperkok 2005; Erfle, Neumann et al. 2007; Erfle and Pepperkok 2007; Simpson, Cetin et al. 2007). Prior to printing, frozen siRNAs were thawed at room temperature for 60 minutes and samples concentrated at the bottom of the wells by centrifugation at 1000 rpm for 1 minute (22^{0}C). The siRNA samples used for array printing were prepared by dispensing 5 *μ*l of 12.5 μM (final 2.72 μM) of the individual siRNAs into 96-well matrix microplates (Thermo Scientific, Denmark). The siRNA were encapsulated for reverse transfection by dispensing 18 *μ*l of the encapsulation mixture (1.5 M sucrose: Sigma-Aldrich, USA, 0.4% w/v gelatin: Sigma-Aldrich, USA, EC buffer: QIAGEN GmbH, Hilden, Germany and lipofectamine: Invitrogen, USA) by means of a WellMate liquid dispensing robot (Matrix, USA). The encapsulation mixture was supplemented with the addition of red fluorescent non-targeting siRNA (3 μl of 20 μM, final 2.6 μM) as an optically addressable marker (red si-GLO, Thermo-Fisher, USA) (Genovesio, Giardini et al. 2011; Genovesio, Kwon et al. 2011).

Microarray printing was performed with a custom-built microarray printer from Dr Emans' laboratory. Prior to printing, siRNA encapsulation mixtures were loaded by capillary action using micro-fluidic filling plates into 330 μm diameter printing elements pre-mounted onto printing plates. Each print element was individually filled with a different encapsulated siRNA. The print plates were mounted onto a custom built Z axis movable printing head. Using a 2.5 seconds contact time, an array of the individual siRNA spots was printed onto a poly-L-Lysine (PLL) modified glass Substrate. Arrays were kept under a stream of humidified air and 55-65% RH during the print run. Post printing, trays of arrays were stored in light- and air-tight custom-made desiccators (Pieter Fourie Design and Engineering cc, Pretoria). All array manipulations were carried out in an ISO4 custom-built 2.5 m^3 clean bench with two down facing HEPA filters (VividAir PTA, Pretoria).

3.4 Array coverslip Poly-L-Lysine coating

For optimal adhesion of the siRNA, array coverslips 24X60 mm (Marienfeld, Germany) were chemically treated with 0.01% Poly-L-Lysine (Sigma-Aldrich, USA), prepared by dissolving PLL powder (100 mg) to 10 ml in autoclaved MilliQ water. PLL Solution was filtered with a 0.45 μm acrodisc syringe filter (Nalgene, USA) and a 0.01% PLL coating buffer was made by making up the solution to 1L using MilliQ water. Prior to coating, coverslips were carefully loaded into custom-built racks with each rack containing a maximum of 50 coverslips. Racks were placed into chambers and autoclaved for 30 minutes at 121^0C. 500 ml of the PLL solution was dispensed into each chamber containing the array coverslips and incubated at room temperature for 1 hour. To prevent uneven coating, coverslips were rinsed in MilliQ water and excess water on the coverslips was removed by centrifugation. Array coverslips were air dried overnight and later stored in sterile airtight boxes for subsequent use.

3.5 Routine experimental cell culture

Three commonly exploited cell lines were examined in this thesis. HeLa (H-3) is a human negroid cervix epitheloid carcinoma cell line. U2OS is a human Osteosarcoma cell line and TZM-BL a human epithelial cell line derived from a HeLa cell clone that constitutively expresses CD4, Co-receptor CCR5, and CXCR4 (Platt et al., 1998) and contains two reporter genes, firefly luciferase and *E. coli* Beta galactosidase under the control of the HIV-1 LTR promoter. The characteristics of the HeLa cells have allowed the cell line to be widely used in many *in vitro* models of infection. HeLa, U2OS cells were utilized between passages 5 to 30 and maintained in Dulbecco's modified Eagle Medium 1X (DMEM, GIBCO, Grand Island, NY) supplemented with penicillin (100 U/ml)/streptomycin (100 μg/ml) (Pen-strep, GIBCO, Invitrogen) and 10% fetal bovine serum (FBS, GIBCO, Invitrogen). The TZM-BL cells were maintained in DMEM supplemented with 50 μg/ml gentamicin (gentamicin, GIBCO, Invitrogen), 26 μM HEPES solution (HEPES, GIBCO, Invitrogen) and 10% FBS. Cells were routinely grown in T75 tissue culture flasks at 37^0C in an atmosphere of 5% CO_2 in air.

3.6 Reverse transfection of printed siRNA into cultured cells

Array coverslips were transferred from the airtight custom made desiccators into Ø100 X 20 mm tissue culture dishes (TPP, Switzerland) prior to cell culturing. Culture cells were trypsinized and resuspended in Opti-MEM tissue culture medium (Opti-MEM, GIBCO, Grand Island, NY) supplemented with 5% FBS and penicillin (100 U/ml)/streptomycin (100 μg/ml). For reverse transfection, typically $1X10^6$ cells in 20 ml Opti-MEM (final $5X10^4$ cells/ml) were plated per siRNA array coverslip. Prior to cell seeding, the siRNA array coverslips were washed for 10 minutes by incubating the coverslips with 10 ml of the culturing medium at 37^0C. Medium was aspirated and the medium containing the cells was dispensed into tissue culture dishes containing the siRNA array coverslips and grown for 48 hours at 37^0C in a humidified atmosphere containing 5% $CO_{2.}$

3.7 Immuno-labeling

Coverslips with adherent cultured cells were mounted in a custom watertight chamber (Chamlide, Korea). Cells were gently washed in Dulbecco's phosphate buffered saline (PBS++, GIBCO, Invitrogen), fixed for 10 minutes in 4% (w/v) paraformaldehyde (Powder 95%, Sigma-Aldrich, USA): PBS and permeabilized for 10 minutes with 0.1% TX-100 (v/v) in PBS at room temperature. For immunostaining, NFκB p65 (A) rabbit polyclonal IgG (Santa Cruz Biotechnology) (0.2 μg/ml) or XPO1 (H-300) rabbit polyclonal IgG antibody (Santa Cruz Biotechnology) (1 μg/ml) was used with (1 μg/ml) Alexa 488 goat anti-rabbit IgG (H+L) (molecular probes, Invitrogen, USA) as secondary antibody. 10% goat serum (Sigma-Aldrich) was used as a blocking buffer to eliminated all non-specific binding. Cells were stained in primary antibody diluted in blocking buffer (1 hour, room temperature) and secondary antibody diluted in blocking buffer (30 minutes, room temperature). Array coverslips were washed four times in PBS between each staining step for 10 minutes. Nuclei were counterstained with 1 µM DRAQ5 in PBS (Biostatus, UK). All incubation steps were carried out on an ultraRocker (Bio-rad Laboratories, Hercules, CA).

3.8 4% (w/v) paraformaldehyde buffer

For fixation of culture cells, a 4% (w/v) paraformaldehyde (PFA) in PBS ++ solution was used. Briefly, for a 500 ml solution 20 g of PFA powder was added into a beaker containing a magnetic stirrer and 350 ml of boiled milliQ water. The beaker was placed on a magnetic stirring plate with temperature set at 60^0C. Adding 10N Sodium hydroxide solution in 100 μl steps cleared milky PFA solution. Using a pH strip, the pH was checked and adjusted to pH 8 using fuming hydrochloric acid. 50 ml of 10X PBS++ was added to the solution and using MilliQ water, the final volume of the solution was adjusted to 500 ml using a volumetric flask. The solution was aliquoted into 50 ml falcon tubes and stored at -20^0C for subsequent use.

3.9 Microarray acquisition and spot analysis

Immuno-labelled arrays in chambers were mounted in a microtiter plate dummy for imaging. Arrays were acquired with a point scanning confocal reader (Image express Ultra, Molecular Devices, USA). DAPI (excitation 405 nm, emission 400-450 nm), FITC (488 nm, emission 500-550 nm), TEXAS RED (excitation 561 nm, emission 600-650 nm) and CY5 (excitation 635 nm, emission 650-670 nm) fluorescence were captured using sequential acquisition to generate separate image files for each. A 10X extra long working distance magnification objective, pixel binning of 2 and five averages was used during all acquisitions. Images were acquired as 16 bit TIFF files written directly to an external database on an external 8TB server. Image processing was performed using ImageJ v1.44 and Array-pro analyzer v6.3 (Media Cybernetics, USA) softwares. Adaptive grid fitting was applied to identify, annotate and export siRNA spots within an entire array. The siRNA spots were fitted and annotated before exporting as cropped, annotated 48 bit RGB TIFF spot images. Image analysis of spots and cell morphology used both commercial and in-house algorithms.

3.10 Curve fitting and pathway analysis

Measurements were expressed as the mean and standard deviation for all cells in four images per well, per treatment and exported to a spreadsheet. Assay development data was analyzed using Graphpad Prism (Graphpad, USA). Curve fitting to determine was performed using sigmoidal dose response fitting of six titrations. Hits from the screen were analyzed using Pathway Studio (Elsevier, USA).

3.11 TNF-α induced p65 (NF-κB) nuclear translocation cell based assay validation

HeLa (H-3) cells were trypsinized from cell culture flasks and plated into 96-well tissue culture Greiner plates at plating density of 7.5 X 10^3 cells per well and cultured overnight at 37°C in a humidified, 5% CO_2 incubator. Twenty-four hours after plating, cells were washed twice in DMEM supplemented with 1% FBS. Cells were later starved for an hour with DMEM (1% FBS). For TNF-α induced p65 nuclear translocation, cells culture plates were stimulated for 1 hour with different TNF-α concentrations (0 to 200 ng/ml) (GIBCO, Invitrogen) in 1% FBS DMEM medium. Cells were subsequently treated for immunofluorescence as described in section 3.7.

3.12 Forward transfection of siRNAs and TNF-α stimulation of HeLa cells

For validation of the individual pooled siRNA, 3.75 X10^3 HeLa cells per well were cultured in 96-well tissue culture Greiner plates. Cells were cultured in 100 μl antibiotic free Opti-MEM medium supplemented with 5% FBS for 5 to 7 hours. Cells were later supplemented with 20 μl (antibiotic and serum free medium) of the transfection mix containing the individual siRNAs (20 nM) and lipofectamine. Cells were further incubated for 48 hours. Prior to TNF-α stimulation, cells were washed twice in DMEM supplemented with 1% FBS. Cells were later starved for an hour with DMEM (1% FBS) before stimulated with TNF-α (100 ng/ml) for 1 hour. Transfected cells were subsequently treated for immunofluorescence as described in section 3.7.

3.13 Bacteria strains, culture conditions, media and maintenance

The bacteria strain used in this study was GFP-*L. monocytogenes* obtained from the Institut of Pasteur France (Paris). Bacteria were cultured in enriched tryptic soy broth (Becto, Dickinson and company, USA) supplemented with 7 μg/mg of chloramphenicol. Briefly, culture flasks containing bacteria were grown at 37^0C by agitation (110 rpm) for 12 hours. At OD_{600} ~1 the bacterial medium was diluted 1 to 3 and further grown until

OD_{600} was between 0.3 to 0.6. Culture medium was supplemented with glycerol (20% final concentration), aliquoted and stored at -80^0C until required.

3.14 GFP-*L. monocytogenes* invasion assay

H-3 HeLa cells were routinely cultured in T75 TC flasks as described in section 3.5. Cultured cells were washed twice with 10 ml PBS++, which was then aspirated. 3 ml of Trypsin-EDTA was added to the flasks and incubated for 1 minute at room temperature. Trypsin-EDTA solution was removed and TC flask further incubated for 3 to 5 minutes at 37°C with 5% CO_2. Trypsinized cells were then resuspended in full serum medium. Prior to infection, cells were seeded in a humidified atmosphere of 5% CO_2 at 37°C in 96 well imaging (glass based) culture plates (Greiner Bio-one CELLSTAR®, Neuburg Germany) at a number of 7.5 X 10^3 cells per well. Cells were grown for Twenty four hours, washed three times with PBS and infected with GFP-*L. monocytogenes* at different multiplicity of infection concentrations (MOI 0 to 100) in an antibiotic free medium. Infection was carried out at different time points and at the end of each time point, cells were washed and incubated further (1 hour) in medium containing 50 μg/ml gentamicin to eliminate extracellular bacteria and to prevent reinfection.

3.15 Immuno-labeling of GFP-*L. monocytogene*s infected cells

Arrays with adherent cultured cells were mounted in a custom watertight chamber and gently washed once in PBS. GFP-*L. monocytogenes* infected cells were fixed in 4% w/v Paraformaldehyde: PBS for 10 minutes followed by treatment of cells for 10 minutes with 0.1% v/v Triton X-100 in PBS. Cells were washed four times with PBS++. Infected cells were labelled for 1 hour with 0.1 μM ATTO 647 Phalloidin (ATTO-TEC, Germany) in PBS to stain for the actin cytoskeleton, and the nuclei were counterstained with 12 μg/ml Hoechst 33324 (Invitrogen, Eugene, Oregon, USA) in PBS. Array coverslips were imaged as described in section 3.9 using the ImageXpress confocal imaging system. All fixing, staining and washing steps were carried out at room temperature with gentle mixing on an ultra-rocker (Bio-rad, Hercules CA) at 20 rpm.

3.16 Forward transfection of pooled siRNAs and GFP-*L. monocytogenes* infection of HeLa cells

For validation of the individual pooled siRNA, 3.75 X10^3 HeLa cells per well were cultured in 96-well tissue culture Greiner plates. Cells were cultured in 100 μl antibiotic free Opti-MEM medium supplemented with 5% FBS for 5 to 7 hours. Cells were later supplemented with 20 μl (antibiotic and serum free medium) of the transfection mix containing the individual siRNAs (20 nM) and lipofectamine. Cells were further incubated for 48 hours. Prior to GFP-*L. monocytogenes* infection, cells were washed twice in DMEM supplemented with 10% FBS. Cells were subsequently infected with 100 μl of DMEM (10 % FBS) containing the GFP-*L. monocytogenes* at MOI of 25. Bacteria internalization and growth was allowed to proceed for 3 hours. Extracellur bacteria were eliminated from the infected cells by treating the infected cells with 50 μg/ml of gentamicin. Infected cells were later treated for immunofluorescence as described in section 3.7 with a slight modification in the nuclei staining and the addition of cell actin cytoskeleton stain. The nuclei were counter-stained with 12 μg/ml Hoechst 33324 while the cell actin cytoskeletons were counter-stained with 0.1 µM Phalloidin 647.

CHAPTER 4

ARRAY PRODUCTION AND CHARACTERIZATION

ABSTRACT

Systematic screening and characterization of gene function is a key challenge, especially after the sequencing of the human genome and those of many other organisms.

High throughput screening technology using RNA-interference (RNAi) has been widely used by researchers to study gene function. RNA interference results in the silencing of expression of a desired target gene and a loss-of-function phenotype. This enables identification of genes involved in a cellular process.

Here we demonstrate and validate a new high-throughput screening technology in which a cohort of siRNAs are mechanically spotted onto a chemically modified glass cover slip. This approach differs from others in that thousands rather than a few dozen well defined nuclei acid spots are simultaneously deposited on a support.

To validate the process, we applied criteria to determine if the printed spots were uniform and also to verify if the nuclei acid activity within the printed spot was optimal. Using a computational-based approach, we were able to characterize: the spot size, relative distance between the spots, mean intensity of the printed spots, functionality of the nucleic acids through reverse transfection of two human cell lines (HeLa and U2OS) as well as spot to spot crossover.

4.1 Introduction

Given the size of the human genome, strategies for screening the genome have relied on automation of experiments. These include systematic well-plate based screening using robotic or manual handling of samples, as well as arrayed screening of printed siRNA, and pooled screening of shRNA libraries (Pust, Morrison et al. 2005; Mullenders, Fabius et al. 2009; Sims, Mendes-Pereira et al. 2011; Dow, Premsrirut et al. 2012; Strezoska, Licon et al. 2012) Pooled screening is achieved by introducing a library of different shRNA simultaneously into a population of cells and the cells are then subsequently selected for a phenotype of interest.

Classical siRNA screening exploits fluid phase delivery of an siRNA complexed with a transfection reagent into cells grown in well plates (Cheng, Magdaleno et al. 2011). Introduced siRNAs then enter cells and trigger silencing of their cognate gene product. This is a labor-intensive process and for scale up to the level of the human genome, the strategy has been to automate the process, rather than miniaturizing it.

Reverse transfection technology inverts the classical siRNA process (Ziauddin and Sabatini 2001) in which the siRNA transfectant complexes are deposited on the base of a well and then overlaid cells grow over the complexes and are transfected. Reverse transfection has been efficiently miniaturized to where up to 3888 siRNA: transfectant complexes can be mechanically deposited onto a glass slide (Genovesio, Giardini et al. 2011; Genovesio, Kwon et al. 2011). These arrays have been used to screen for cellular proteins involved in HIV infection as well as in the identification of human host factors required for *Trypanosoma cruzi* infection.

Reverse transfection arrays emerged through the application of arraying technology to cDNA or siRNA deposition, using robotic arrayers depositing small volumes through contact printing. Typically, spots are deposited using a print element that iteratively prints a sample, and a library one sample at a time (Genovesio, Giardini et al. 2011).

Here, the objective was to demonstrate that gene silencing using reverse transfection was effective before moving onto large-scale screening. Three measurements based on the known outputs expected from RNAi array experiments were applied: siRNA spots should be uniformly printed, gene silencing is effective and measurable and spot-to-spot cross contamination should be minimal.

4.2 Results and discussion

4.2.1 siRNA spot characterization

When printing microarrays, consistency in spot morphology is a very important aspect to take into consideration. Printed spots should have similar shape, size and uniformity as discrepancies in the latter parameters could lead to inconsistent results. To demonstrate the production of siRNA spots for reverse transfection, the deposition of siRNA onto the glass substrate was tested. For this method to work the following conditions need to be met, the siRNA must be retained in the printed spot by encapsulating the siRNA using the appropriate encapsulation chemistry, printed spots should be fluorescently labelled for rapid identification and the encapsulated siRNA must be printed on a chemically modified glass substrate to minimize dissipation of the siRNA during use.

To achieve this, a print element was filled with an encapsulation solution containing a red fluorescent siRNA, mimicking transfection conditions. The print element was coupled to a Z-axis drive and brought into contact with a glass substrate, then imaged using a fluorescent confocal microscope. The size of printed spots varied as a function of print element size, contact time and relative humidity (Figure 4.1).

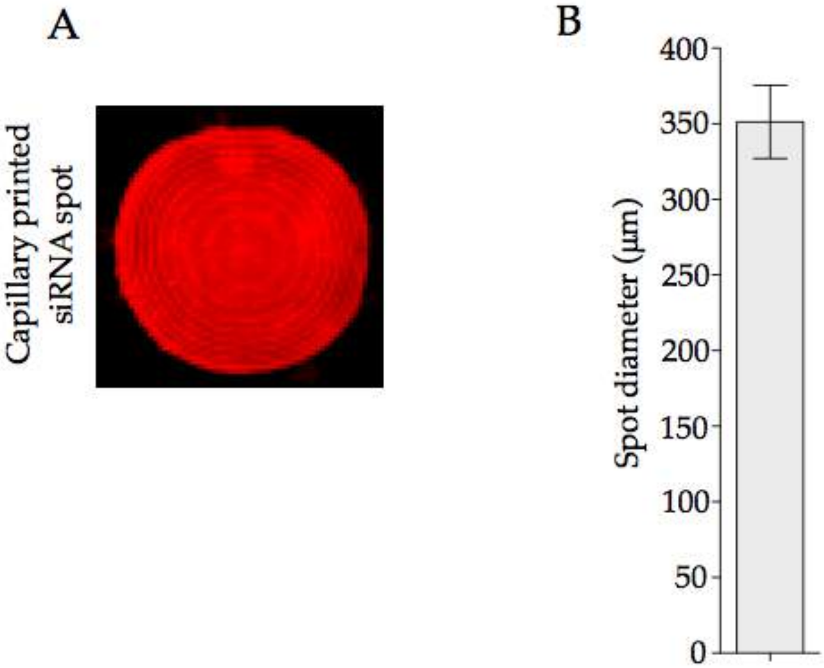


Figure 4.1: siRNA array spot quantification.

Printed encapsulated siRNA spots demontrate reproducibility in spot intensity and shape after measuring a few hundreds of the siRNA printed spots (n=750 spots, >10 individual capillaries). Contact printed optically addressable siRNA spot [A], and spot average diameter [B].

After optimization, printed spots were imaged as relatively regular and circular with an average diameter of 351 µm±0.5 µm (SEM; n=2377 imaged spots) and a tight diameter distribution (Figure 4.1 A and B). These results indicated reproducibility of the siRNA printed spots was possible.

4.2.2 Printed siRNA(s) do not cross-contaminate neighbouring spots

In the course of printing siRNA microarrays there is a possibility of adjacent spots cross contaminating each other. However, for data reliability, cross contamination between the individual printed siRNAs spots should be minimal. This could result from loading the individual printing element or from the spreading of the encapsulation solution during the print cycle.

To assess for cross contamination from delivering encapsulated siRNAs into the printing element or during the print cycle, printing elements were alternately loaded with solution containing either green fluorescent labelled or red fluorescent-labelled encapsulated siRNA. A line of spots with intercalated red and green spots was then printed to a substrate. Printed spots were air dried and subsequently imaged. Image analysis of the printed siRNA spots show no detectable overlap indicating that there were no crossovers between neighbouring siRNAs during filling or in the course of printing the individual nuclei acids spots. It should however be noted that the image presented in Figure 4.2 is a section of a much larger array.

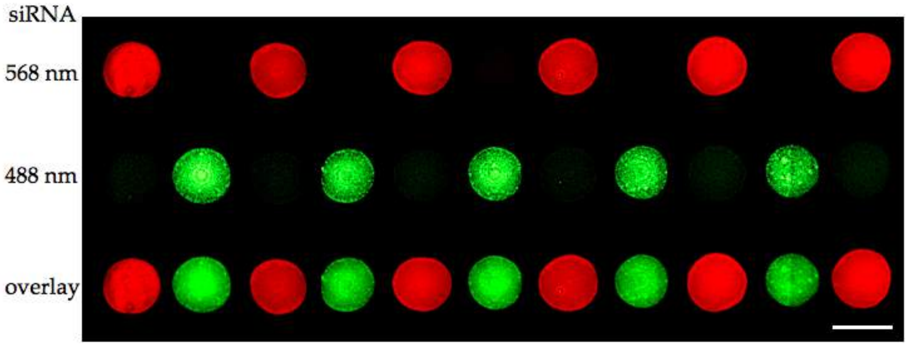


Figure 4.2: Array spot-to-spot crossover fluorescent image verification.

Spot to spot crossover detection using 488 nm and 568 nm excitable fluorescently conjugated siRNA loaded into capillaries and printed onto a glass wafer. Scale bar 200μm.

4.2.3 Quantification of gene silencing by reverse transfection

In order to achieve this objective, it was necessary to focus on optimizing the chemistry of the siRNA spots for optimal cell transfection. Conventional forward transfection siRNA yields 60-70% gene knock down within 24 to 48 hours (Devling, Lindsay et al. 2005). Therefore if reverse transfection is used then the silencing efficiency obtained should be comparable (Huang, Hancock et al. 2009).

To accurately evaluate the efficiency of gene silencing via reverse transfection, p65 (NF-κB subunit), XPO1 (nuclear export of cellular proteins bearing a leucine-rich nuclear export signals), INCENP (gene required for the normal segregation of the chromosome during cell division) and control (non-targeting) siRNA were encapsulated with gelatin, printed, dried and overlaid with HeLa cells for 48 hours. These samples were then fixed and stained by indirect immunofluorescence for RELA/p65 & XPO1 expression, or imaged for the formation of polynucleate cells after INCENP silencing and nuclear staining. Silencing was quantified after image analysis of > 20 experiments.

Images of the cells overlaying siRNA spots were acquired in three channels red, blue and green depicting the printed siRNA spots, the nuclei and immuno-labeled cells respectively (Figure 4.3 to 4.5). Here we demonstrate efficient gene knockdown when HeLa cells are reverse transfected with the siRNA. Reverse transfection with the p65 siRNA gave $> 75\%$, ($p<0.0001$, $n = 2918$) silencing efficiency while XPO1 gave $> 70\%$, ($P<0.0001$, $n= 297$ spots) when both siRNAs were compared to cells transfected with the control non-targeting siRNA (Figure 4.3 and 4.4).

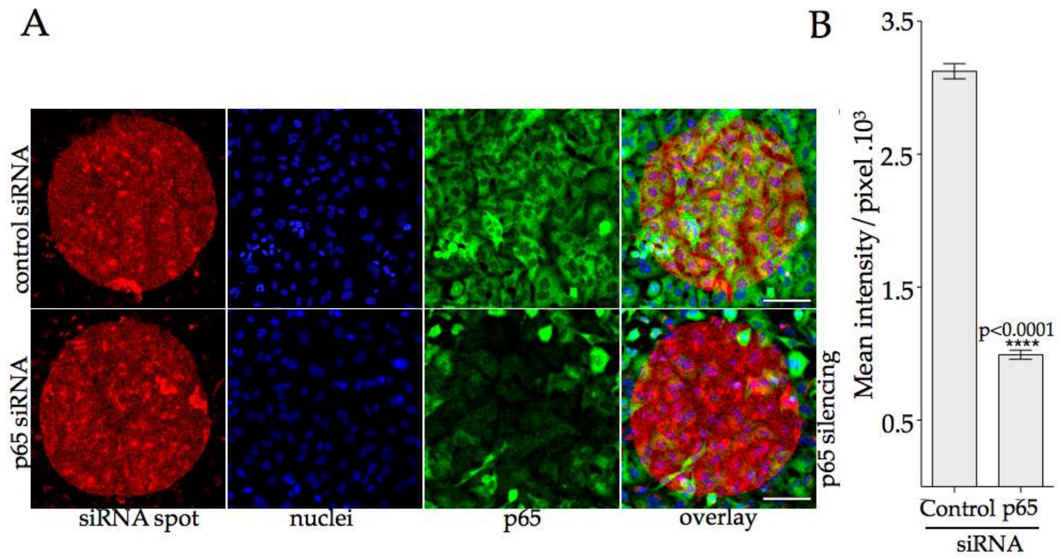


Figure 4.3: Silencing of p65 directed siRNA printed spot.

[A] On the upper row, the red spot shows the non-targeting control siRNA. Next to it is an image of the nuclei stained with 1 μM DRAQ5. The third image shows HeLa cells stained for the NF-κB p65 subunit by immunofluorescence while the last image is a 3-color overlay. On the bottom row, the red spot shows the p65 siRNA, followed by the nuclei, p65 staining and the overlay in the right most panels. Scale bar 100 µm. [B] Quantification of p65 silencing after automatic cell detection and image analysis. (n=2918 spot pairs and P value <0.0001).

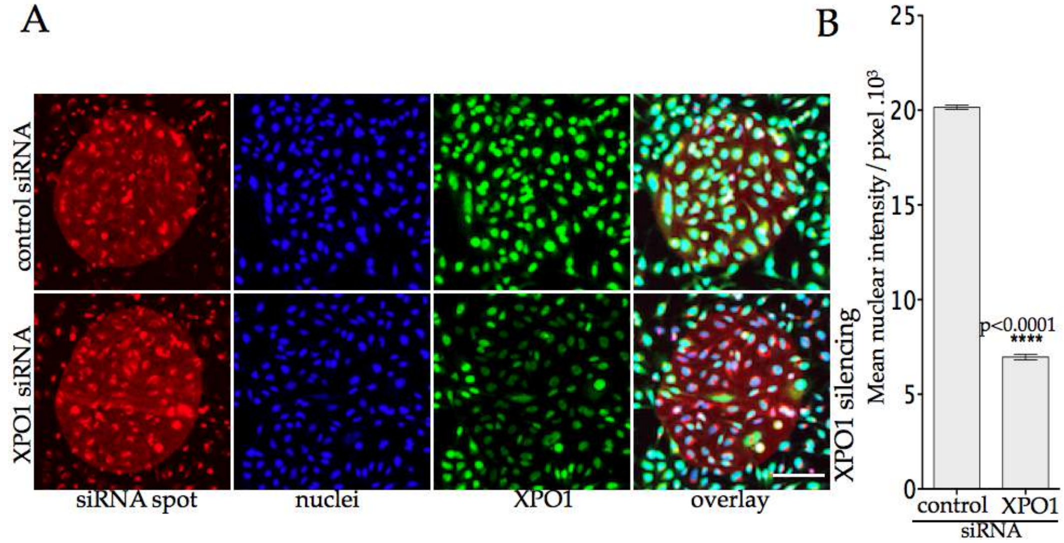


Figure 4.4: Silencing of XPO1 directed siRNA printed spot.

[A] HeLa cells were overlaid onto printed XPO1 and non-targeting siRNA spots. Cells were reverse transfected for 48 hours. For immunofluorescence, HeLa cells were stained with CRM1 (exportin) rabbit polyclonal antibody and the nuclei were counter-stained with 1 μM DRAQ5. [A] Images were acquired in three channels: column 1 highlighting the printed siRNA spots (red), column 2 showing the nuclei (blue), column 3 the immunofluorescence detection of XPO1 and the last column showing the 3 color overlay. [B] Quantification of XPO1 silencing after automatic cell detection and image analysis. (n=297 spot and P value <0.0001) Scale bar 100 µm.

To further demonstrate the effectiveness of gene silencing from siRNA spots, the cytokinetic regulator gene, INCENP siRNA was printed onto a glass substrate along with non-targeting siRNAs. The siRNA spots were reverse transfected into HeLa cells for 48 hours. Prior to imaging, cells were stained for immunofluorescence with p65 antibodies and the nucleus counter stained with DRAQ5 fluorescent dye. Images were acquired in three channels as mentioned above. As demonstrated in figure 4.5 below, polynucleated cells were observed on the INCENP siRNA spot. Polynucleated cell is a characteristic behaviour resulting from an aberrant mitotic cell division, a phenomenon triggered by knocking down the INCENP gene. The INCENP gene forms part of the mitotic regulatory complex (chromosomal passenger complex) that orchestrates chromosome alignment, histone modification, and cytokinesis (Blank and Shiloh 2007).

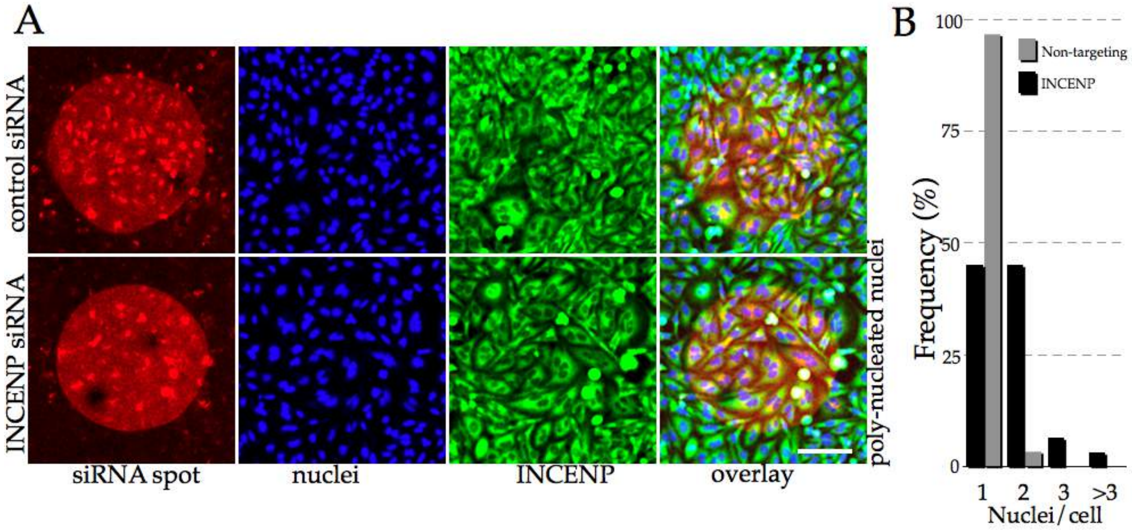


Figure 4.5: Silencing of INCENP directed siRNA printed spot.

[A] HeLa cells overlaid for 48 hours on a glass substrate containing INCENP and non-targeting control siRNA spots. Cells were immuno-stained with p65 antibody and nuclei counter-stained with 1 μM DRAQ5. The upper panel represents the non-targeting siRNA printed spot overlaid with HeLa cells while the lower panel shows the INCENP siRNA printed spot overlaid with HeLa cells. Aberrant mitotic cell divison is observed with the HeLa cells exposed to the INCENP siRNA spot. [B] Frequency histogram of nuclei/cell distribution for the non-targeting and the INCENP siRNA printed spots. (n=350 spots, twelve independent print runs) Scale bar 100 μm.

4.2.4 Gene silencing is localized to the spot

For siRNA printing to be an effective tool in screening, the deposited siRNA must remain on or close to the printed spot to avoid cross activity resulting from the different nucleic acids printed. To verify that the printed siRNA are localized to the printed spot, siRNA directed against p65 were printed as single spots on a glass chip. The siRNA spot was overlaid with U20S cells for 48 hours to permit reverse transfection and gene silencing. After immuno-fluorescent staining of p65 expression, the chip was imaged (Figure 4.6).

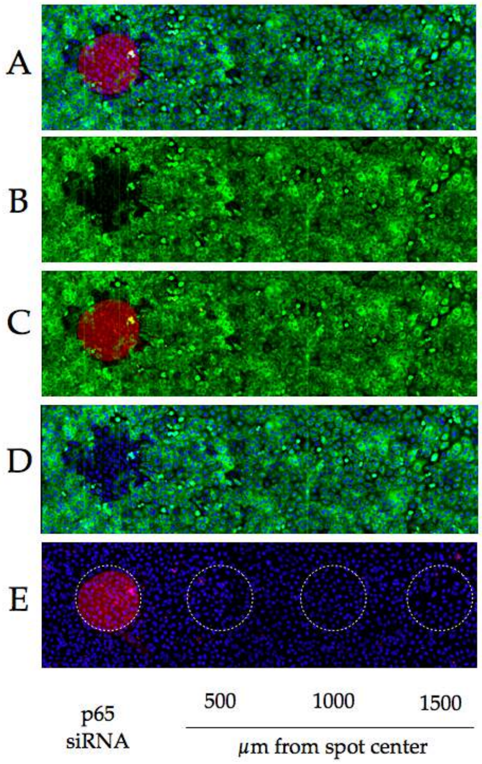


Figure 4.6: Gene silencing is restricted to the siRNA printed spot.

Gene silencing was measured using a p65 siRNA spot and three equivalent areas out from the center at 500 μm interval. U20S cells were reversed transfected with p65 siRNA spot for 48 hours. Cells were fixed, permeabilized and stained with p65 primary antibody. The cells' nuclei were counter-stained with 1μM of DRAQ5 nuclear stain. Images were acquired in three channels (red, green and blue) using a confocal imaging system. [A] Shows the overlaid image of all three channels, [B] image of the immuno-labeled cells (green) without the p65 siRNA spot and the nuclei, [C] an overlay image of the p65 siRNA spot (red) and the immuno-labeled cells green), [D] an overlaid of the nuclei (blue) and the immuno-labeled cells (green) and [E] showing a confocal images of U2OS cells overlaying the p65 spot and the equivalent areas simulating neighboring spots (dotted lines).

Human p65 expression was automatically quantified by detecting the stained cell nuclei (Figure 4.6 E blue) within the regions of interest. A cell recognition algorithm was used to detect the cell cytoplasm surrounding the nuclei and then quantify for p65 expression (staining intensity).

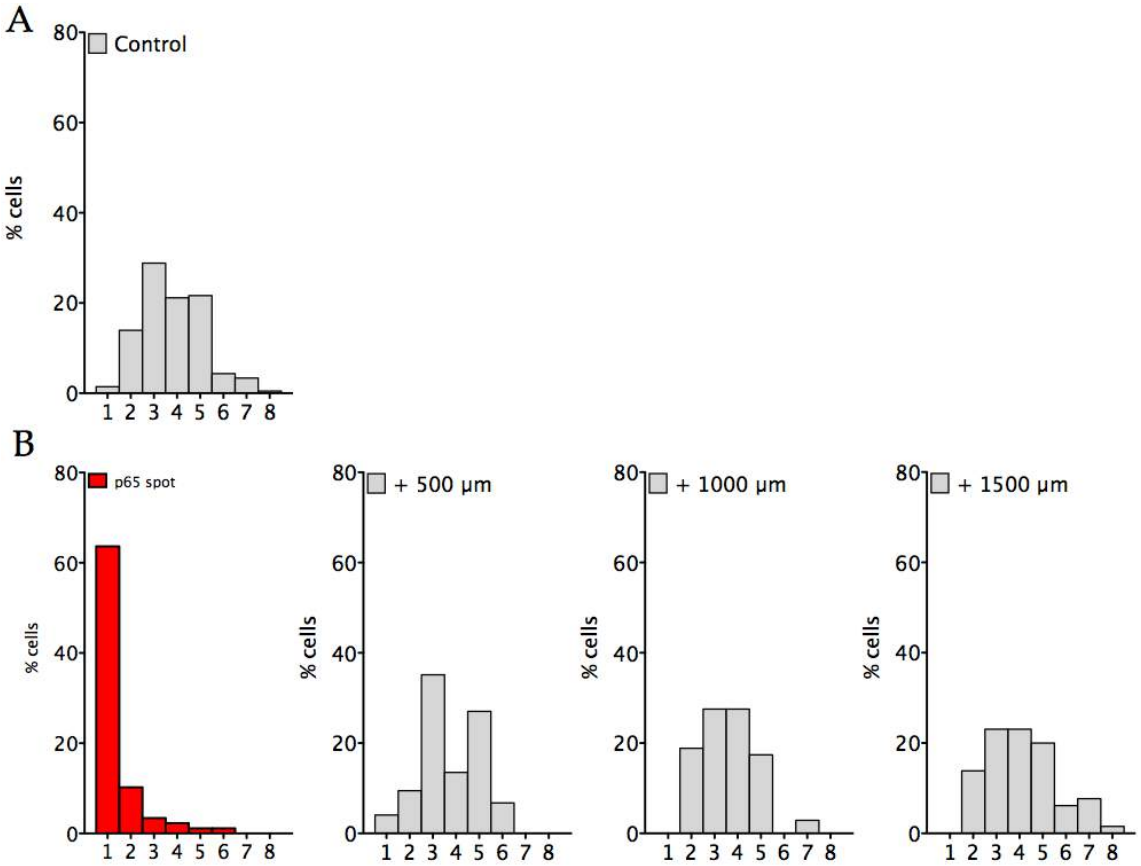


Figure 4.7: Frequency distribution histogram of p65 cytoplasmic intensity/cell for the p65 spot and three simulated spots at 500, 1000 and 1500 µm from the p65 spot center.

The p65 and control populations are well separated. At these levels, the number of cells having a silencing phenotype in the areas simulating spots was the same as cells from arrays never exposed to p65 siRNA and represents inaccuracies in imaging, cell permeabilization, labeling and detection (n>20 spots>200 cells).

As observed from images (Figure 4.6 A-D), the p65 siRNA spots demonstrated that silencing was exclusively localized on and within 100 µm of the spot boundary. Gene silencing is observed as low intensity/dark cells on and around the siRNA spot since in the absence of p65 there is little p65 specific antibody binding and little fluorescent signal. Data for all cells in the four ROI were then exported for statistical analysis.

Gene silencing was quantified from cells on the spot, and at areas equivalent to a spot, 500, 1000, 1500 µm from the siRNA. This was achieved by creating regions of interest around the siRNA spot then over the same area at increasing intervals from the siRNA spot center (Figure 4.6 E dotted lines). These three additional simulated siRNA spots were then used to determine the extent of siRNA spread into cells, by detecting any silenced cells.

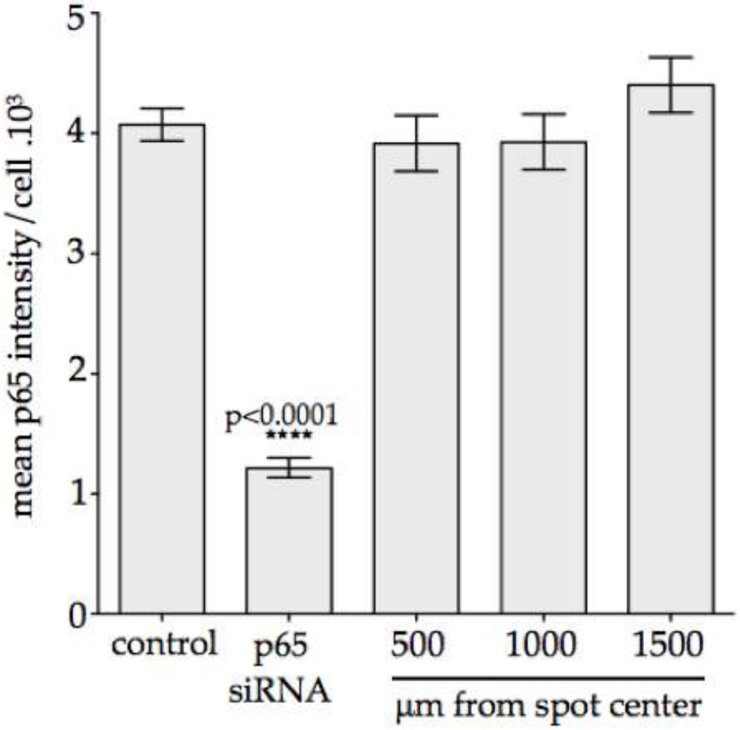


Figure 4.8: Statistical quantification of mean intensity of p65 and three simulated spots at 500, 1000 and 1500 µm from the p65 spot center.

The p65 silencing/cell/spot was quantified with >70% reduction in p65 expression compared to control within the p65 spot with a $p<0.0001$ between p65 and the +500 µm spot. In contrast, cells from the +500 to +1500 µm simulated spots were statistically indistinguishable from control cells ($p=0.5564$ (+500 µm), $p=0.5926$ (+1000 µm), $p=0.2268$ (+1500 µm).

Experimental chip data (Figure 4.7 A) and a spot area from a control chip that had no printed siRNA are shown. Cells exposed to p65 siRNA show a distribution of p65 expression that is lower than the control and simulated spots (Figure 4.7 B). Most cells on the p65 spot are silenced and the number of cells scored with a silencing phenotype in the simulated spots was the same as cells from arrays never exposed to p65 siRNA and represents inaccuracies in imaging, cell permeabilization, labeling and detection. P65 silencing was also demonstrated after a statistical analysis of the mean intensity of p65 and three simulated spots at 500, 1000, 1500 µm (Figure 4.8).

4.2.5 siRNAs array validation

The technology used for printing enables the production of many copies of the same siRNAs as arrays. To determine glass chip quality throughout a production cycle, we quantified gene silencing from a set of production cycles. Each cycle comprised more than one thousand contiguously printed chips from one load of the printer.

For assessing production, alternating columns of siRNA spots were printed to chips, depositing either p65 or non-targeting control siRNA. The printed array comprised of alternating columns of either p65 or non-targeting control siRNA across the entire array surface. This gave a chip of 75 columns of 35 spots containing 2625 individual experiments, identical but for the siRNA sequence. In three independent print cycles, >1400 arrays were printed in less than 2 hours in one continuous cycle (Figure 4.9).

Sets of measurements were applied to determine print quality, which were chosen based on the known outputs expected from RNAi array experiments. Criteria included the number of printed spots, dried spot intensity, shape and efficiency of p65 silencing when overlaid with cells.

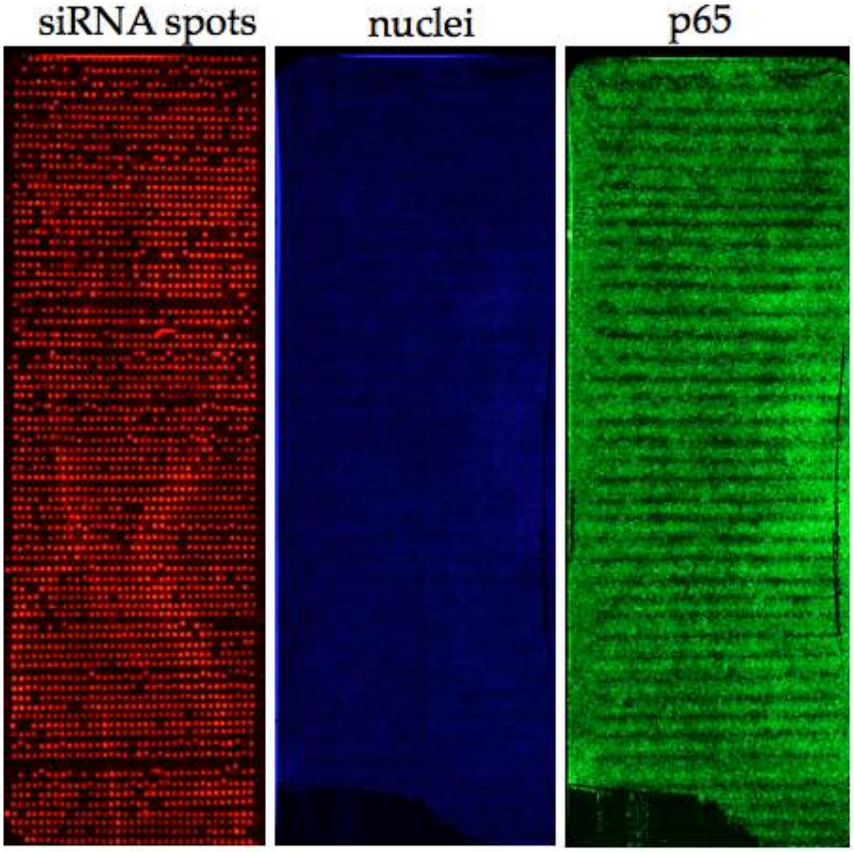


Figure 4.9: High density microarray comprising of 2625 siRNA spots (alternate p65 and non-targeting siRNA spots) overlaid with HeLa cells.

HeLa cells were reversed transfected for 48 hours on the high-density array. Cells were fixed, permeabilized and immuno-stained for p65. Column 1 shows the siRNAs printed spots, column 2, the imaged nuclei after cells' nuclei were counter stained with 1μM DRAQ5 and column 3, HeLa cells stained with p65 antibody.

To determine the fraction of printed spots and spot quality before use in cell based experiments, dry arrays were acquired on an automated confocal microscope then automatically fitted to identify spots, then spot images extracted using custom scripts for quantifying the spot characteristics and dimensions (Figure 4.10 A-C). Each spot contained a red fluorescent siRNA enabling identification of the spots during imaging.

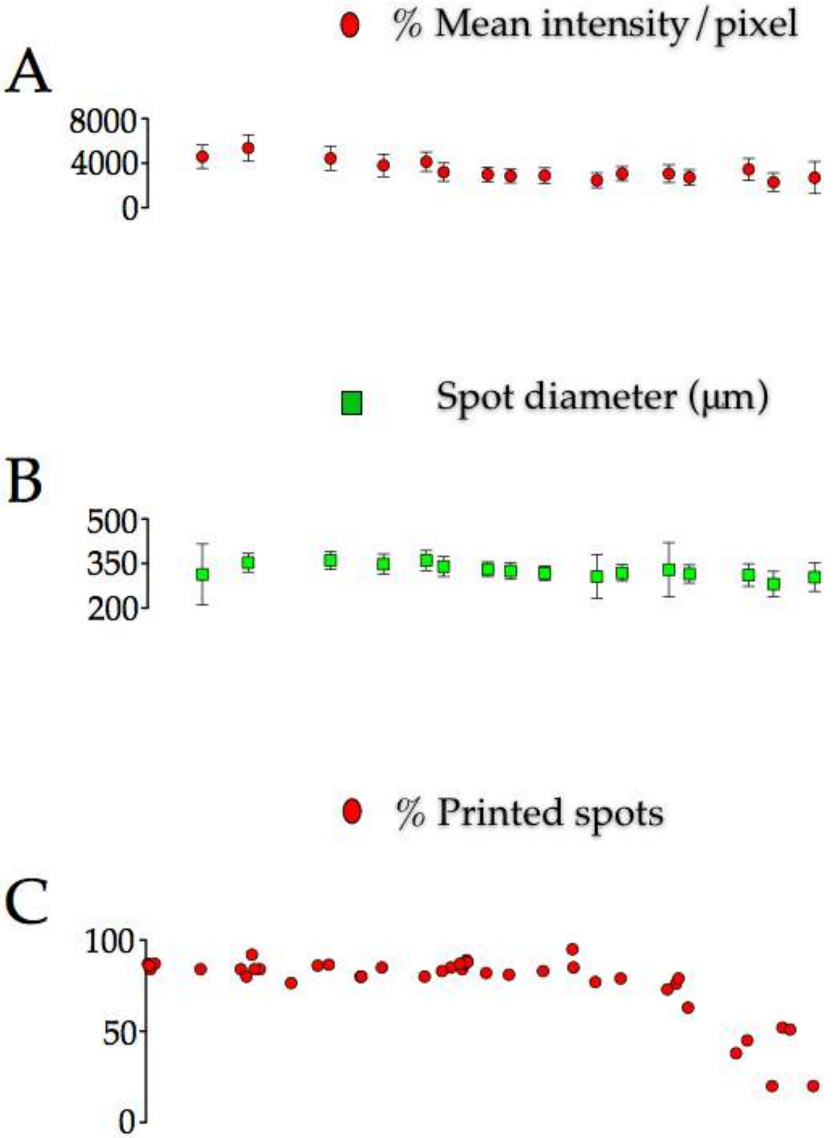


Figure 4.10: Quantification of large-scale siRNA/RNAi array printing process, print quality and gene silencing from a contiguous production of 1207 arrays.

Spot intensity measured using the red fluorescent siRNA. (n=>300 spots/point for A-C). [B] Spot diameter quantified using a MetaXpress inbuilt algorithm. [C] Fraction of printed spots compared to array 1. Spot number drops as a result of print element running out of print solution (n= 24 arrays).

Spot areas were sampled from an entire array and the same spots across the print cycle were extracted for quantification. Figure 4.10 A and B demonstrate computational quantification of the spot mean intensity (n> 300 spots) and spot average diameter across the entire cycle of the array production. Spot intensity and spot average diameter was constant, as observed during the print cycle. We then went on to quantify a more critical parameter, the fraction of printed spots on the chip over the cycle. In figure 4.10 C, a series of spots (n= 24) were selected and similar spots followed across the print cycle.

The number of printed spots started decreased after a certain point (>1000) of the print cycle. Decrease in spot number however could be attributed to the printing element running out of print solution or unable to deposit the siRNA spot due to mechanical stress encountered from the repeated up and down motion of the printer Z-axis.

Another important aspect taken into account in this quantification was to determine how effectively the nucleic acids within the printed spots could function in gene silencing. This was achieved by measuring p65 silencing after arrays comprising of alternate p65 and non-targeting siRNAs were overlaid with HeLa cells. To measure for gene silencing across the print production, arrays for analysis were taken at different time point in a single print cycle. Here, p65 expression was detected with indirect immuno-labeling, array imaging, image extraction and analysis from fitted arrays (Figure 4.11 A and B). The p65 siRNA spots (n> 1000) were sampled from the individual arrays in the production cycle. As depicted in figure 4.11 B, p65 silencing was maintained during production but began to decline after the 750th round of the print. The decrease in p65 silencing could be attributed to the print element running out of siRNA. For robust and reproducible results, it was therefore reasonable to establish arrays 1-750 as the ideal chip set for screening.

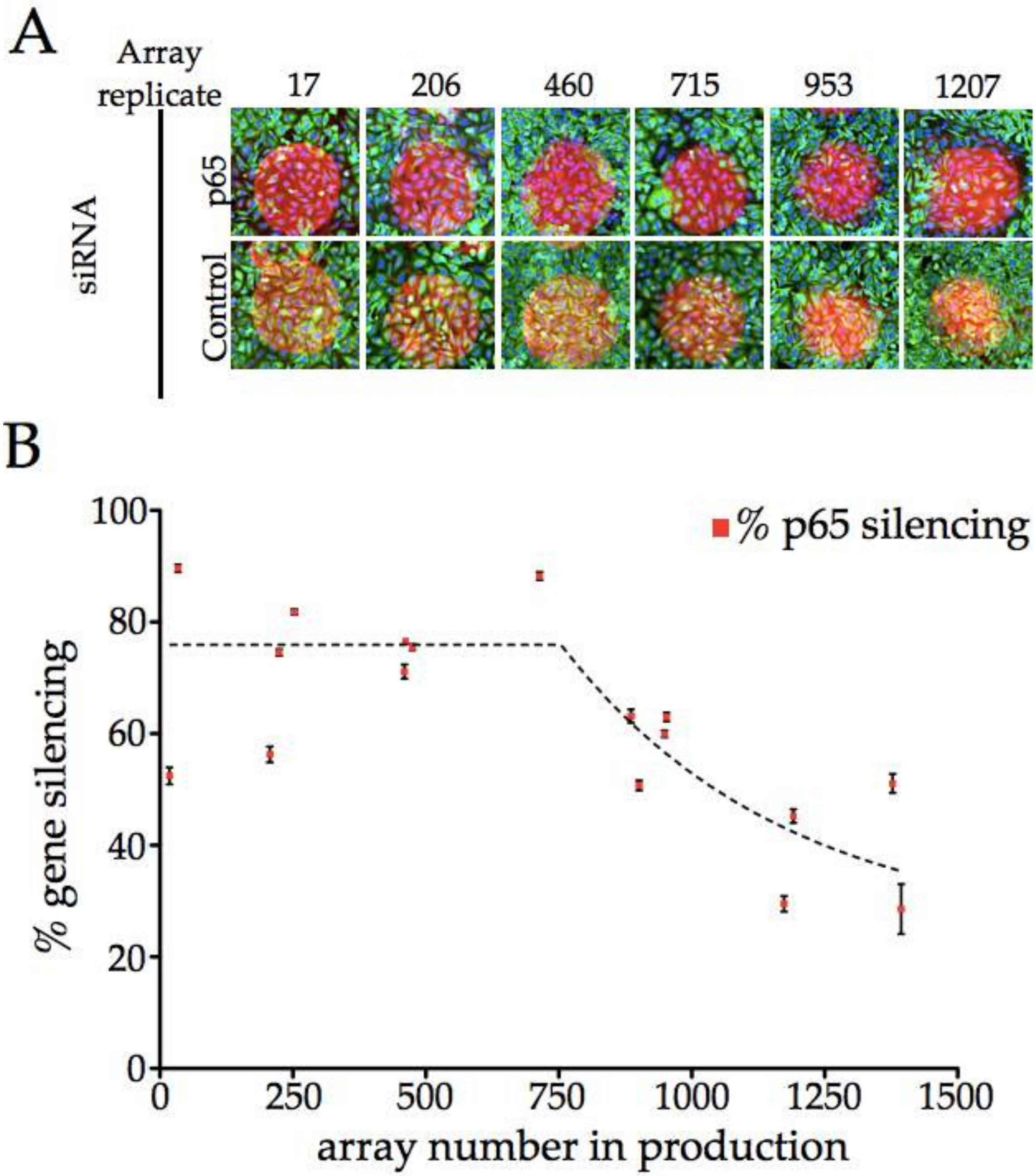


Figure 4.11: Quantification of p65 silencing across array production.

[A] Shows an immunofluorescence of HeLa cells overlaid on printed p65 and non-targeting siRNA spots. HeLa cells were reverse transfected for 48 hours, fixed and stained with p65 polyclonal antibody. Images were acquired using an automated confocal microscope. The upper panel shows p65 silencing on cells overlaid on p65 siRNA spots while the lower panel with non-targeting siRNA spots depicts no p65 silencing. [B] Quantification of p65 silencing in arrays overlaid with HeLa cells for 48 hours after grid-fitting, spot extraction and image analysis of p65 expression (n > 1000).

4.2.6 Conclusion

In the past, the use of genome-wide analysis to study loss-of-function phenotypes on a genome-wide scale was very difficult. However, with the introduction of the siRNA technology by Andrew Fire and Craig Mello (Fire, Xu et al. 1998), researchers are now using this method as a powerful approach to silence the expression of genes and to analyze loss-of-function phenotype in intact living cells.

Here, the feasibility of a technology stemming from the initial reverse transfection microarray technology is demonstrated. This was introduced by Ziauddin and Sabatini (Ziauddin and Sabatini 2001), in which they demonstrated that cells grown on a glass wafer could take up DNA-lipid complexes that had been mechanically deposited on the glass substrate. Using this basic principle, we demonstrate a flexible approach for the production of high-density microarrays, which could be used with great ease for highly parallel phenotype analysis in living cells.

Here we show that thousands (>2700) of siRNAs in an encapsulation mixture could be mechanically spotted in parallel onto a microarray glass substrate and subsequently be used to analyze for gene silencing studies in HeLa cells. To assess print quality of the siRNA microarray, three measurements based on the known outputs expected from RNAi array experiments were applied: siRNA spots should be printed, spot size and mean intensity should be fairly constant during production, gene silencing be measurable and spot-to-spot cross contamination should be minimal. It was demonstrated that printed siRNA spots were relatively regular and circular with diameter of 351 μm ± 0.5 μm across the array print cycle. Spot intensity was maintained for a given range of printed arrays. Gene silencing was shown to be measurable with silencing efficiencies >70%. Printed p65 siRNAs were effective in knocking down the expression of the target genes. Cell morphology was unaltered on spotted control siRNA compared to unspotted regions of the array with 50-120 cells/spot. It was therefore concluded that for consistent image acquisition and data interpretation, it was vital arrays used for all future experimental screens be selected from a specific group of arrays from the entire array production (between array 1 and 750).

This validation subsequently led to the printing of 3150 siRNA spots comprising of a subset of the human siRNA library covering kinases, phosphatases, ubiquitinylases and proteases. For array quality control, a series of control siRNAs that span the entire array were included in the printed arrays: p65, XPO1, and non-targeting siRNAs.

CHAPTER 5

IDENTIFICATION OF HUMAN HOST FACTORS INVOLVED IN TNF-α INDUCED P65 (NF-κB) NUCLEAR TRANSLOCATION

ABSTRACT

Implementing a cell-based assay for screening requires some in-depth understanding of the endpoint measured, the correlation with cell viability, and the limitations of the assay chemistries. Here in order to use the high-density chip to screen for genes implicated in the *Listeria* infection pathway, it was reasonable to validate the array using a simpler cell based assay model before moving into a more complex system such as that of *Listeria*. It was expected that if using this simpler model on the microarrays yielded suitable outcomes in identifying key players in a pathway, it will therefore be reasonable to use the array in a more complex model.

Three different biological nuclear translocation assays were evaluated: EGF (epidermal growth factor) induced ERK (extracellular signal regulation kinase) nuclear translocation, anisomycin induced Phospho-c-jun nuclear translocation and TNF-α (tumor necrosis factor-alpha) p65 (NF-κB) nuclear translocation. After statistical quantification of half-maximal activity concentration (AC50's) and Z' (30) for each of the above assays, the TNF assay was chosen for the validation since it gave a Z' of > 0.7, suitable for biological assays and an AC50 of 25 ng/ml corresponding to those mentioned in the literature. The NF-κB signaling pathway has over the years been reported to play a key role in the regulation of numerous cellular processes by inducing targeted gene transcription. The current large number of genes regulated by this pathway has made it a subject of intense research. Another reason why the aforementioned signaling pathway was used for the validation was because the pathway has been extensively studied and documented.

5.1 Introduction

Complex cellular processes maintain growth, differentiation and development. These cellular processes are orchestrated by the expression of a wide variety of genes tightly regulated by their respective transcription factors. These transcription factors form families such as the nuclear factor-kappaB (NF-κB) family. It is known to play a pivotal role in many cellular functions, including cell proliferation and survival, apoptosis, organ development, innate and adaptive immunity, stress and inflammatory responses (Weih and Caamaño 2003; Bonizzi and Karin 2004). The NF-κB is a dimeric transcription factor comprising of five different family members including: NF-κB1 (p50), NF-κB2 (p52), RelA (p65), RelB, and c-Rel (Rel), which can either coexist as homo-dimers (RelA/RelA, p50/p50, p52/p52) or hetero-dimers (p50/RelA, p50/c-Rel, p52/c-Rel, p65/c-Rel, RelB/p50 and RelB/p52) (Krappmann and Scheidereit 2005; Moynagh 2005; Hoffmann, Natoli et al. 2006; Oeckinghaus and Ghosh 2009).

These proteins possess an N-terminal region harboring the structurally conserved 300 amino acids Rel homology domain (RHD), comprising of the dimerization, nuclear localization signal (NLS) and DNA-binding domains (Karin, Cao et al. 2002; Lätzer, Papoian et al. 2007), (Figure 5.1 A). The NF-κB dimers are sequestered in the cytoplasm of unstimulated cells in a latent inactive form, non-covalently bound to the inhibitory protein IκB protein complex. The IκB inhibitory complex comprises of seven different subunits; IκBα, IκBβ, IκBδ, IκBε, BCL3, p100 and p105, all characterized by the presence of ankyrin repeats known to be the docking site for the NF-κB dimer (Moynagh 2005; Cervantes, Markwick et al. 2009; Miyamoto 2011), (Figure 5.1 A). Its dissociation from the NF-κB dimer is mediated by two homologous subunits, CHUK (IKKα/IKK1), IKBKB (IKKβ/IKK2) and NEMO/IKKγ (IKBKG) (Hayden and Ghosh 2008), (Figure 5.1 B).

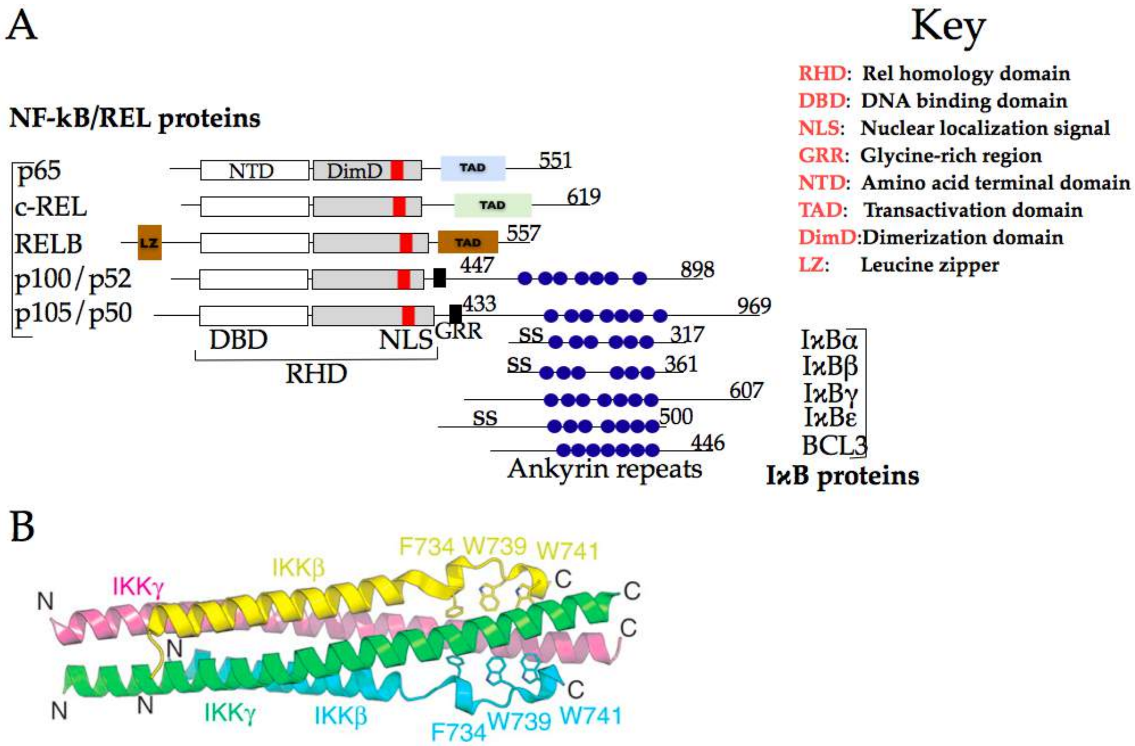


Figure 5.1: The NF-κB, IκB Family and NEMO/IKKγ ribbon structure.

[A] Domain organization of NF-κB/Rel and IκB proteins. The RHD consists of two folded domains, the amino-terminal domain (NTD) or DNA binding domain and the dimerization domain (DimD) that are joined by a short flexible linker and a carboxy-terminal flexible region that contains the nuclear localization signal (NLS). Three members of the NF-κB, p65, c-Rel, and RelB, contain a transcription activation domain (TAD) at their C-terminal ends while the other two, p100/p52 and p105/p50 together with all IκB subunits contain the ankyrin repeats at their C-terminal. RelB contains a predicted leucine zipper motif (LZ) amino-terminal to its RHD. [B] Illustrates the ribbon structure of the IKK2/IKKβ:NEMO/IKKγ complex. Individual polypeptides are labeled, as well as some of the conserved hydrophobic amino acid side chains from IKK2/IKKβ. Adapted from Karin, Cao et al. 2002 and Hayden and Ghosh 2008 with modifications.

Hundreds of stimuli including cytokines (such as the Tumour Necrosis Factor Alpha; TNF-α, interleukin-1; IL-1), growth factors, lymphokines, oxidant-free radicals, viral and bacteria products have been reported to mediate the activation of the IκB complex (Oeckinghaus and Ghosh 2009). Upon stimulation with the TNF-α ligand, the tumor necrotic factor receptor (TNFR) engages proximal adaptor proteins TRADD (TNF receptor-associated death domain protein), RIP (receptor-interacting protein) and TRAF2 (TNF receptor-associated factor 2) that mediate the ubiquitination/activation of of the IKK complex (Miyamoto 2011). This leads to the activation of the NF-κB/IκB complex by signal-induced phosphorylation of the IκB at specific serine residues (ser-32 and ser-36). This process triggers the ubiquitination of the IκB and its subsequent degradation from the NF-κB complex exposing the DNA and NLS domains of NF-κB which enables the p65 (NF-κB) to be translocated to the nucleus where it binds to a specific DNA to regulate targeted gene transcription (Figure 5.2).

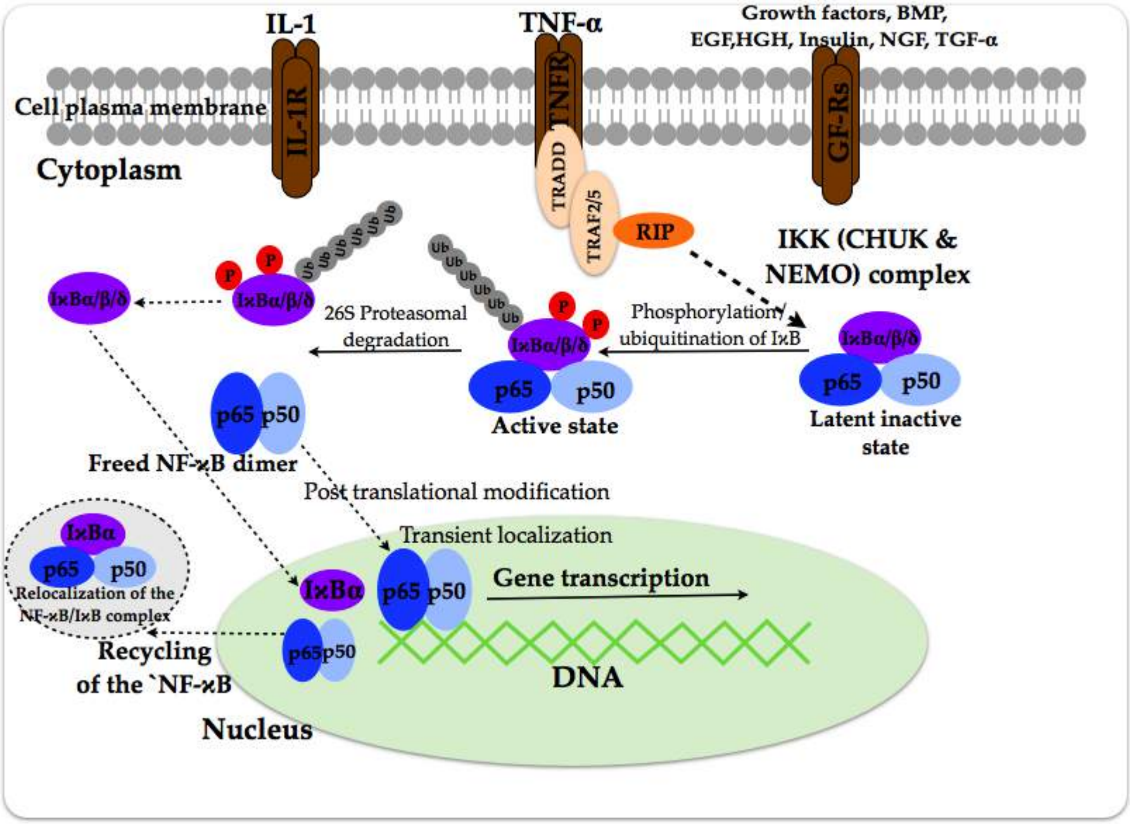


Figure 5.2: TNF alpha induced p65 (NF-κB) signaling pathway (Canonical pathway).

Tumor necrotic factor receptor (TNFR) is stimulated by TNF-α, to signal receptor bound proteins TRADD, RIP and TRAF2 to activate the IKK complex. IKK activation induces the phosphorylation and ubiquitination of the NF-κB complex leading to the dissociation of IκB from the complex. This dissociation enables freed NF-κB to be translocated into the nucleus for targeted gene transcription. Dash arrows represent unknown steps.

Researchers are currently exploiting the RNAi technology to better understand the involvement of the respective upstream proteins in the regulation p65 (NF-κB) nuclear translocation (Takaesu, Surabhi et al. 2003; Chiang, Berghman et al. 2009).

Implementing a novel high throughput RNAi array screening technology, the goal here was to use the well characterized NF-κB signaling pathway to validate the technology and also to identify key genes that may potentially be involved in the regulation of the NF-κB signaling pathway.

5.2 Results and discussion

5.2.1 TNF-α induced p65 (NF-κB) nuclear translocation assay development

A cell based assay for TNF-α induced p65 nuclear translocation was automated and validated in 96-well plates using HeLa cells. This cell line has been well characterized and has been used as a model cell line in the validation of many biological assays (Chong, Swiss et al. 2009; Balestrino, Hamon et al. 2010; Stavru, Bouillaud et al. 2011). The cell line was used to validate the assay in which p65 was translocated into the nucleus upon stimulation of the cell with TNF-α ligand.

Briefly, the p65 protein co-exists as a dimeric protein coupled to c-Rel (p65/c-Rel). The dimeric protein is sequestered in the cytoplasm of unstimulated cells in a latent inactive form non-covalently bound to the IκB protein complex (Nishikori 2005; Oeckinghaus and Ghosh 2009). Stimulating the cells with TNF-α activates the complex by signal-induced phosphorylation of IκB, releasing p65 from the IκB complex. The freed p65 is then translocated to the nucleus where it binds to a specific DNA site to regulate targeted gene transcription (Nishikori 2005).

To validate this assay, cells were cultured in 96-well plates for 24 hours. Prior to TNF-α stimulation, cells were serum starved for 1 hour and then exposed to a TNF titration (0 to 200 ng/ml) for 1hour. Cells were then fixed, permeabilized and p65 stained by immuno-fluorescence. Imaging used an automated confocal imaging system (Figure 5.4). Nuclear localization of p65 was quantified by analyzing images using existing off the shelf industrial standard algorithms to quantify for nucleus-to-cytoplasm average intensity difference of p65 within the cells. Assay conditions were optimized to give maximal p65 transport into the nucleus. Figure 5.3 is a schematic representation of the approach used in developing the TNF assay.

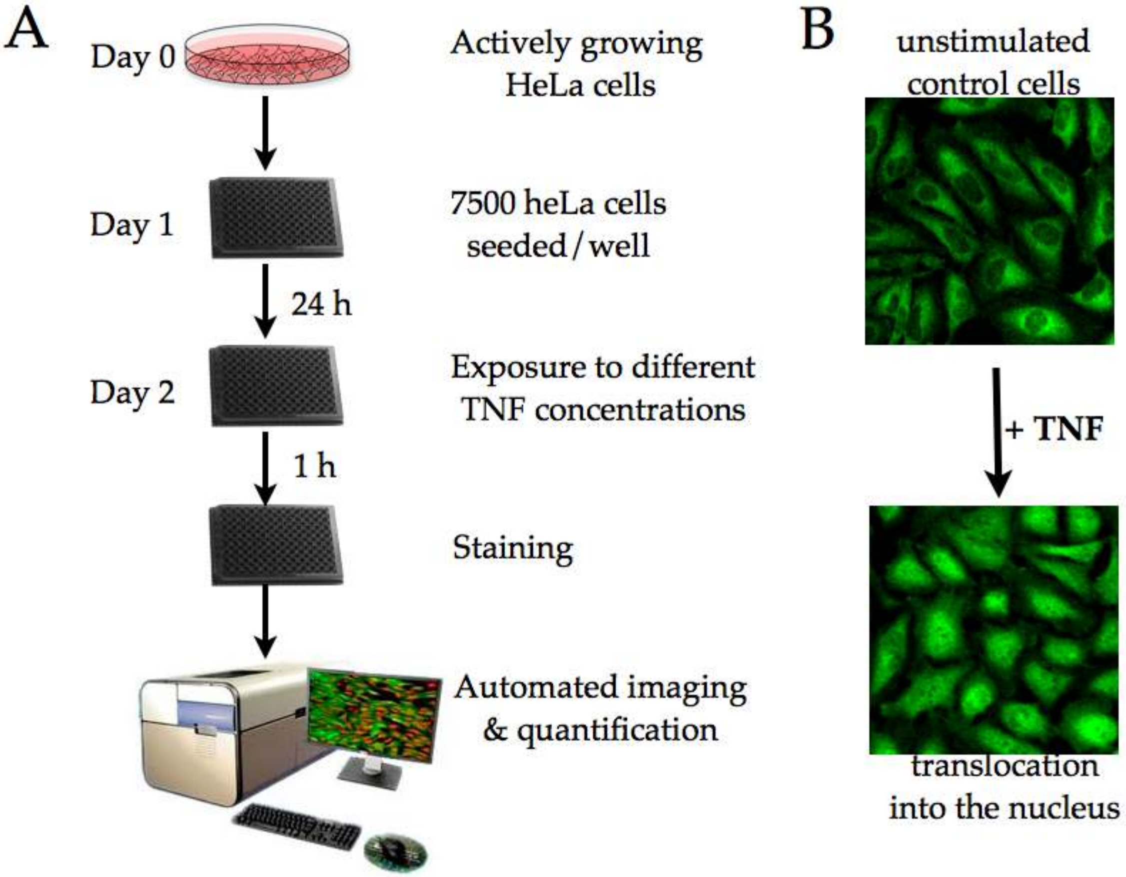


Figure 5.3: TNF pilot assay.

[A] Illustration of TNF-α induced p65 nuclear translocation experimental approach. [B] Expected outcome when cells are stimulated with TNF-α.

Visual inspection of the immunofluorescent images illustrated an increase in nucleus-to-cytoplasmic staining intensities when HeLa cells were exposed to an increasing titration (0 to 200 ng/ml) of TNF-α (Figure 5.4). To statistically validate this methodology, the nuclear and cytoplasmic staining intensities were compared to give the nuclear-to-cytoplasmic, cytoplasmic-to-nuclear ratio and the correlation coefficient as a relative measure of p65 nuclear localization (Figure 5.5 A-C). Here, the nuclear and cytoplasmic data were first normalized for the total number of data points included in the analysis and then comparison was made of the nucleus-to-cytoplasm average intensity of p65 within the unstimulated and stimulated cells. With an R^2 value of > 0.9 for each dose response curve, existing data were subsequently used to compute for the Z' and the AC50 for the assay (Figure 5.5 D).

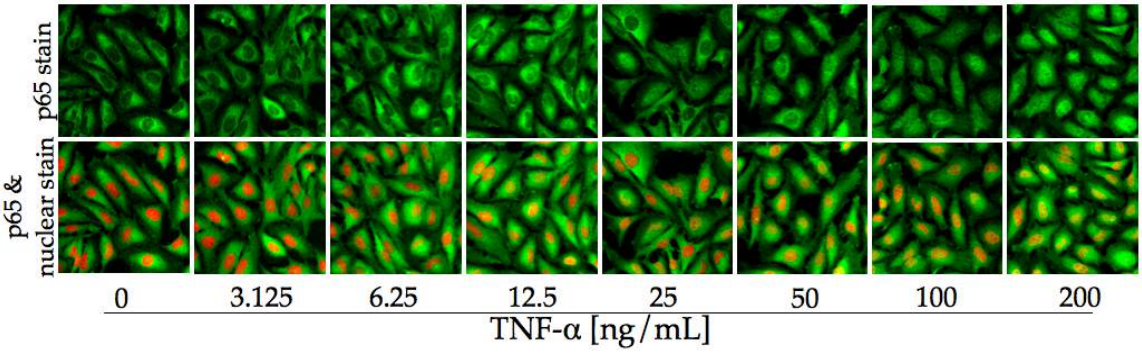
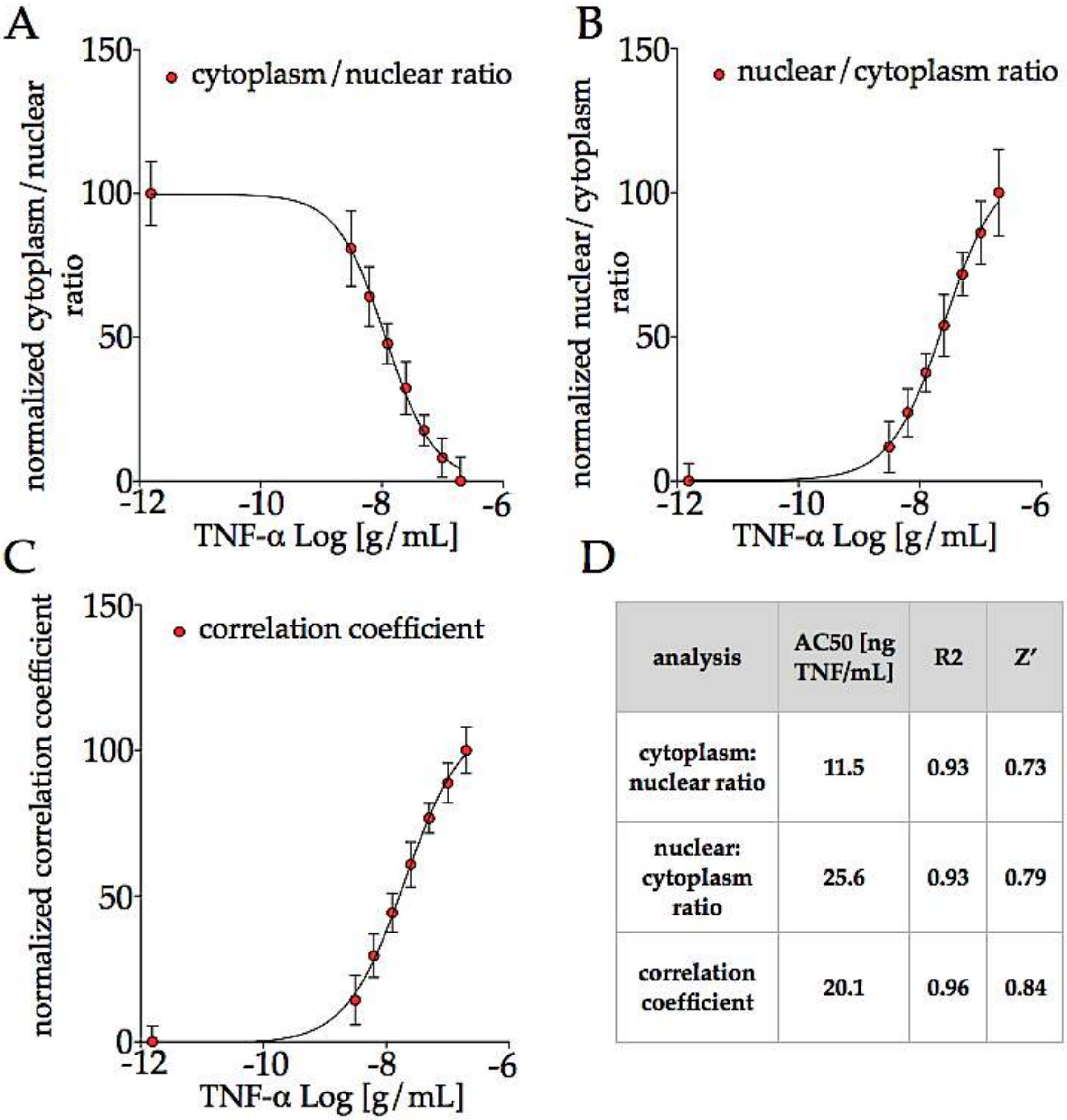


Figure 5.4: TNF-α induced p65 (NF-κB) nuclear translocation.

Immunofluorescence of HeLa cells stimulated with an increasing concentration of TNF-α for 60 minutes. HeLa cells were cultivated for 24 hours and later exposed to different concentrations (0 to 200 ng/mL) of TNF-α for 1 hour. Cells were fixed then stained with p65 polyclonal antibody (green) and 1μM DRAQ5 (red). Images were acquired using a confocal imaging system. The upper panels show p65 immuno-labelling of the cells and the lower panel shows cells labeled with both the p65 antibody and DRAQ5 nuclei stain. Images depict (from left to right) increase in p65 accumulation in the nucleus as the TNF-α concentration was steadily increased.

Results obtained from the assay yielded a statistical effect size of (Z') > 0.7 and an AC50 of 25 ng/ml. corresponding to those mentioned in the literature (Warhurst, Hopkins et al. 1998). The Z' value indicates a robust assay for high throughput screening (Zhang, Chung et al. 1999) (Figure 5.5). It was then predicted that this assay would be suitable for identifying human host factors taking part in the NF-κB signaling pathway after stimulation of the NF-κB cell surface receptors with the TNF-α ligand.

An ideal assay condition was then set as: 60 minutes stimulation with an excess of 10 times the AC50 (100 ng/mL) of TNF-α. This was to ensure maximum stimulation of the cells and nuclear import of p65 (NF-κB). This assay was then transposed to the microarrays comprising of printed siRNA covering the kinases, phosphatases, ubiquitinylases and proteases to screen for proteins taking part in p65 (NF-κB) nuclear translocation.



analysis	AC50 [ng TNF/mL]	R2	Z’
cytoplasm: nuclear ratio	**11.5**	**0.93**	**0.73**
nuclear: cytoplasm ratio	**25.6**	**0.93**	**0.79**
correlation coefficient	**20.1**	**0.96**	**0.84**

Figure 5.5: Quantification of the AC50 and Z’ for TNF-α induced p65 (NF-κB) nuclear translocation.

HeLa cells were exposed to different concentrations of the TNF-α. A-C. Determination of the AC50 of TNF-α induced nuclear import of p65 using the [A] cytoplasmic/nuclear ratio, [B] nuclear/cytoplamic ratio and [C] the correlation coefficient. [D] Z’ calculations for image analysis. Column 1: parameters used in the quantification p65 nuclear/cytoplasmic quantification. Column 2: AC50s measured for the different parameters. Column 3 shows R^2 values for the respective graphs and column 4 gives the Z’ for the measurements used.

5.3 High content RNAi screen for regulators of the TNF induced p65 nuclear transport pathway

A major challenge faced by most high throughput screening platforms is random errors resulting from poor biological assays, poor instrumentation and cell variability related influence (Gunter, Brideau et al. 2003). These factors lead to selection of false positive candidate genes (hits). However increasing the numbers of replicates used in screening have recently been shown to greatly reduced the number of false negative without increasing the number of false positive hits (Malo, Hanley et al. 2006). Incorporation of the above assay (TNF-α induced p65 nuclear translocation) and utilizing the high-density siRNA reverse transfection arrays, we demonstrate the use of high replicate screens to identify effectors of TNF-α induced p65 (NF-κB) nuclear translocation.

A set of 32 screens was performed corresponding to 32 X 3150 = 100,800 individual experiments. Array images from each individual screen were exported, grid fitted, annotated as individual images. Each individual siRNA spot was then extracted and the cells within the spot analyzed using off-the-shelf industrial standard algorithm software. The screening data was processed through principal component analysis and yielded 51 hits, of these 26 (51%) were identified as direct interactors with the TNF to NF-κB pathway including the known NF-κB kinases and regulators CHUK and IKBKB. Multi-dimensional analysis showed that the median number of screens required to robustly select from the cohort of siRNAs was 12 replicates.

These results demonstrated the arrays were optimal and could be used in screening more complex biological assays such as the *Listeria* infection assay.

5.3.1 TNF-α induced p65 nuclear translocation screen

The cell-based assay described in 5.2.1 was adapted for this screen. HeLa cells were seeded onto the arrays and reverse transfected for 48 hours. Cells were then serum starved for 1 hour and subsequently exposed to 100 ng/ml (10X its AC50) of TNF-α for 1 hour. The transfected array was fixed, permeabilized and stained with p65 antibody. Each siRNA experiment was extracted, analyzed using image recognition algorithms and the raw data exported to spreadsheets. Each image yielded data on >10 parameters including cell-by-cell data on p65 expression level, cell number, p65 distribution between the cytoplasm and nucleus per cell and nuclear morphology (elliptical form factor), taking into account the variation in phenotypes that may result from the different genes present in the siRNAs cohort. The data for each screen and each measurement was center reduced to permit cross array comparison and normalize for variance in image acquisition. All 32 screens were automatically compiled and the data expressed as the median value/siRNA for n=32 replicates.

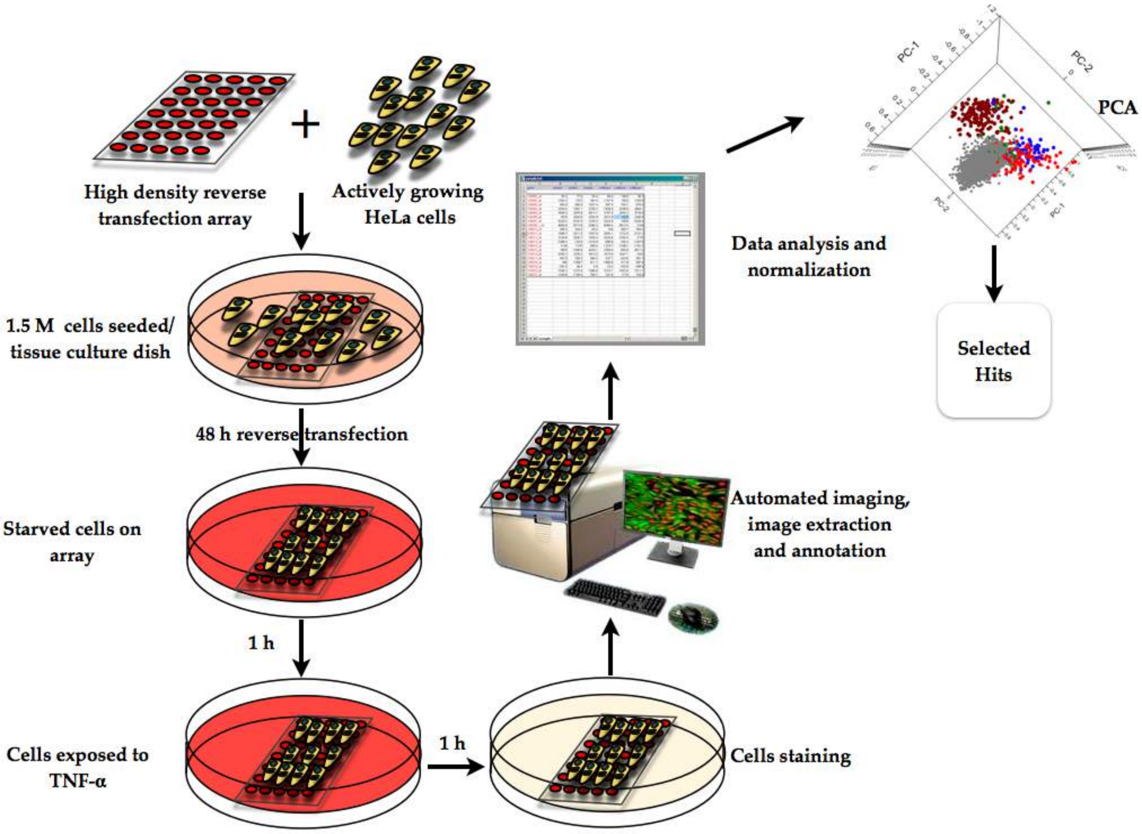


Figure 5.6: NF-κB screening approach.

Schematic representation in approach used in screening for human host factors taking part in the NFκ-B signaling pathway after TNF-α stimulation.

5.3.2 Primary hit selection

High throughput screens generate large quantities of data in a short period of time, making data mining a difficult task to perform (Hampel, Ronchetti et al. 1986). Data resulting from high throughput screens if not properly analyzed using the appropriate analytical approach, will lead to false negative and false positive results. Several analytical tools are currently being used to analyze data generated by high throughput screens (Ben-Hur and Guyon 2003). One of such analytical tools used to analyze and to exploit the information contained in gene expression is the principal component analysis (PCA). PCA is a popular and powerful operation in statistical analysis (Jolliffe 1986) and has been used in high-content screens to interpret huge data sets (Sexton, He et al. 2010; Ashcroft, Newberg et al. 2011; Kota, Eaton et al. 2013). PCA operates by constructing a correlation matrix of all variables. This is achieved by reducing the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much of the variation present in the data sets as possible (Jolliffe 1986) making data mining easier to understand and interpret.

In this study, our initial goal was to demonstrate that the inbuilt controls could be identified as different to the screen using the PCA. The two control genes XPO1 and p65 used are known players in the NF-κB pathway. It should be noted that XPO1 is the nuclear export factor for p65 and when silenced it leads to nuclear accumulation of p65 (i.e an increase in nucleus-to-cytoplasm ratio of p65). While p65 on the other hand is a nuclear factor NF-kappaB p65 subunit and when silenced leads to a 75% reduction in p65 expression within the cell (Lu, Zhang et al. 2010; Tian, Fan et al. 2012).

The control genes were targeted with XPO1 and p65 siRNAs dispersed across the array. It was of interest to determine whether the phenotypes resulting from silencing these control genes could be identified and separated into distinct phenotypic classes using PCA. Since PCA is an objective analytical method, isolating XPO1 and p65 from the screen data would indicate that the process was effective for detecting new phenotypes. As shown by the PCA results (Figure 5.7), there was an excellent separation of the control data in 3D projections.

Furthermore, some siRNAs affect cell viability and it was of interest to detect whether those gave a sufficiently differentiated phenotype to detect with PCA. Since cell number was used in the analysis, we anticipated that spots with a cell number less than 2 standard deviation (STDV) from the median of the entire population could be defined as ‘toxic,’ if the PCA was sufficiently discriminating (Figure 5.7).

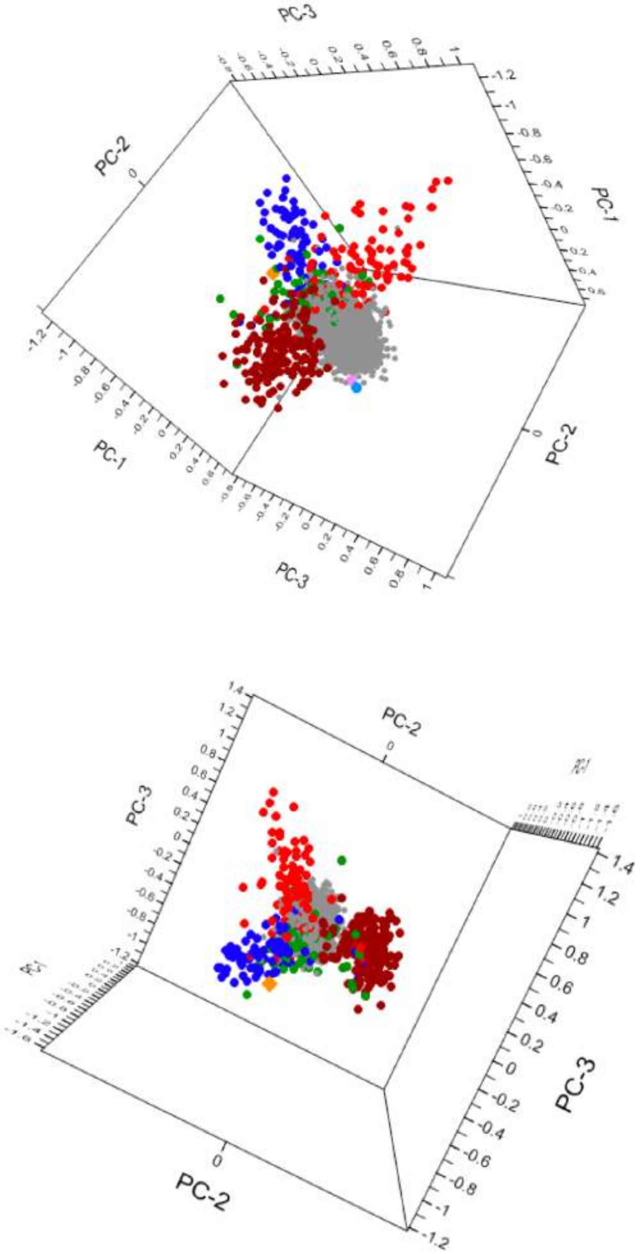


Figure 5.7: 3D projections of control siRNAs from the p65 (NF-κB) screen.

3D projection from different 3D angles showing excellent separation of the p65 (blue), XPO1 (red), non-targeting (green) and toxic siRNAs.

As shown in figure 5.7, PCA gave a robust separation of the control XPO1 (red), p65 (blue) and toxic (brown) classes of phenotypes from the body of the screen.

The human kinome consist of the RELA gene, which encodes for p65 (Nolan, Ghosh et al. 1991). So to further examine the effectiveness of the PCA in data separation, we looked for the relative position of the RELA the gene within the screen.

Silencing the RELA gene will result in a decrease in p65 expression within the cell allowing the PCA to project the gene in the same cloud as the p65 control used in the screen. As predicted by the PCA plot, the RELA gene from the kinome library was located together with the control p65 used in the screen (Figure 5.8).

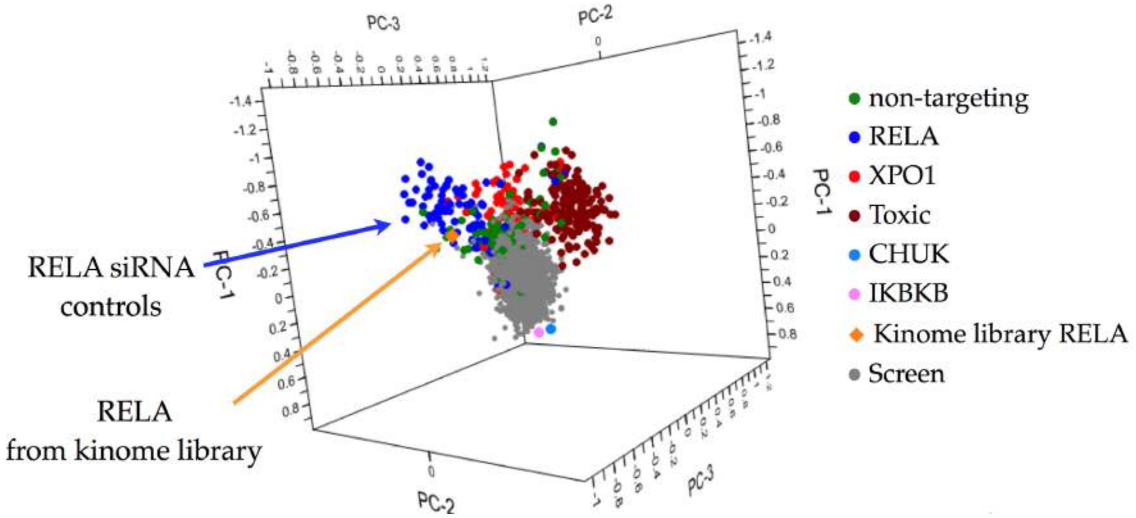


Figure 5.8: PCA plot depicting the relative position of RELA (p65) siRNA within the TNF-α induced p65 nuclear translocation screen.

HeLa cells were transfected with a cohort of siRNAs comprising of kinases, phosphatases, ubiquitinylases, and proteases. Cells were subsequently stimulated with TNF-α, stained, imaged and analyzed using the PCA. PCA separates screen data into different clusters based on the cellular phenotypic expression after gene knockout. RELA (orange star) from the library, which encode for p65 was located within the p65 control cloud.

These results therefore indicated the siRNA screen, the analysis algorithms and the PCA itself were a viable platform to use for candidate genes selection. The PCA resolved three axes heavily influenced by the cell biology of the assay. PC-1 showed the greatest dependence on the import measurement as shown by finding the XPO-1 controls at one end of the axis. The other two axes were strongly influenced by RELA (p65) expression and cell number/spot. We therefore selected “hits” as genes that gave the strongest import block phenotype on PC-1 and little/no effect on the other axes. The prediction was to find genes that block import but do not decrease RELA expression nor kill cells. The group statistically furthest from the screen center of mass is labeled as yellow below.

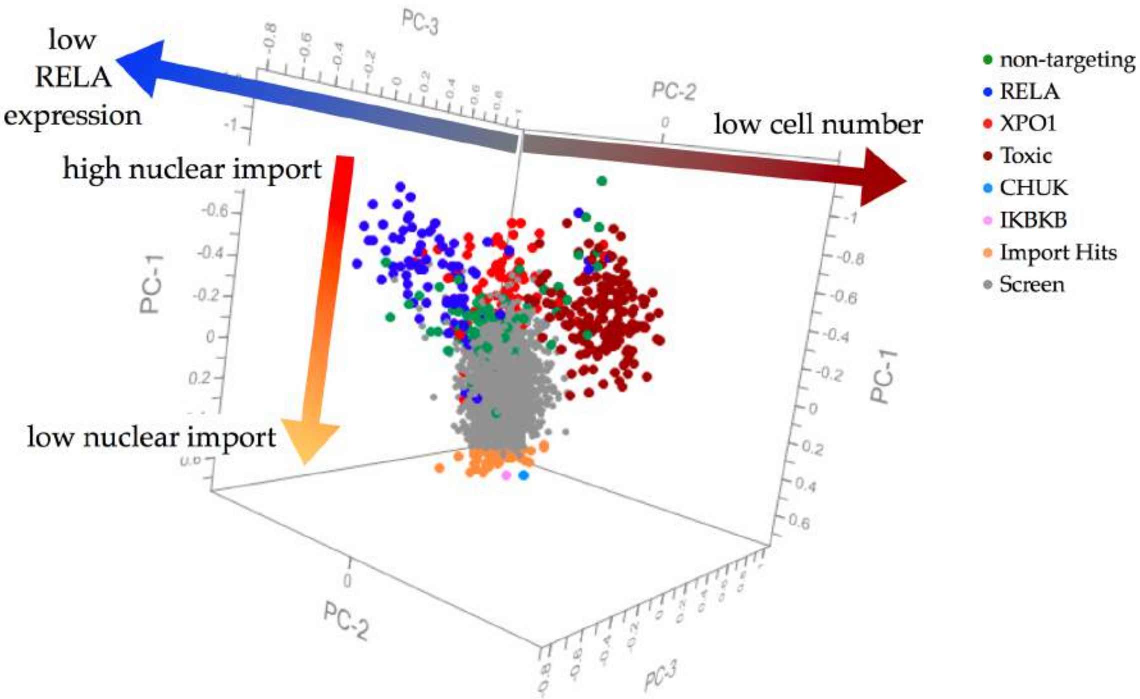


Figure 5.9: 3D projection used for hit selection.

3D projection showing the strongest influences on the PCA axes. The colour coded siRNAs are p65 (blue), XPO1 (red), non-targeting (green), toxic siRNA (brown), CHUCK (sky blue), and IKBKB (pink). Orange dots highlight region where targeted genes were selected.

Interestingly, two known genes that influence p65 nuclear transport were found in this group. Suppressing the activity of these two genes will result in p65 accumulation in the cell cytoplasm (Gewurz, Towfic et al. 2012). These were CHUK [Conserved Helix-Loop-Helix Ubiquitous Kinase] and IKBKB [Inhibitor Of Kappa Light Polypeptide Gene Enhancer In B-Cells, Kinase Beta], which are regulatory subunits of the NF-κB complex. We exploited this to seek genes blocking nuclear import. These genes are located opposite to the XPO1 cloud, as expected since XPO1 is a nuclear export factor and leads to increased accumulation when silenced (i.e. opposite to desired phenotype), (Figures 5.7-5.9). A total of 51 genes were empirically identified (Appendix table 1) in the hit group containing CHUK and IKBKB. The “hit" genes were visually examined to confirm phenotype in 9/10 experiments (Figure 5.10 B). Using the Pathway Studio (http://www.elsevier.com/online-tools/pathway-studio/pathway-studio-web) analytical tool, genes were subsequently interrogated for protein-protein interaction by blasting the candidate genes against the NF-κB signaling pathway using. Data obtained showed that 26 of the genes are known direct interactors with TNFR to NF-κB signaling pathway (Figure 5.10 A).

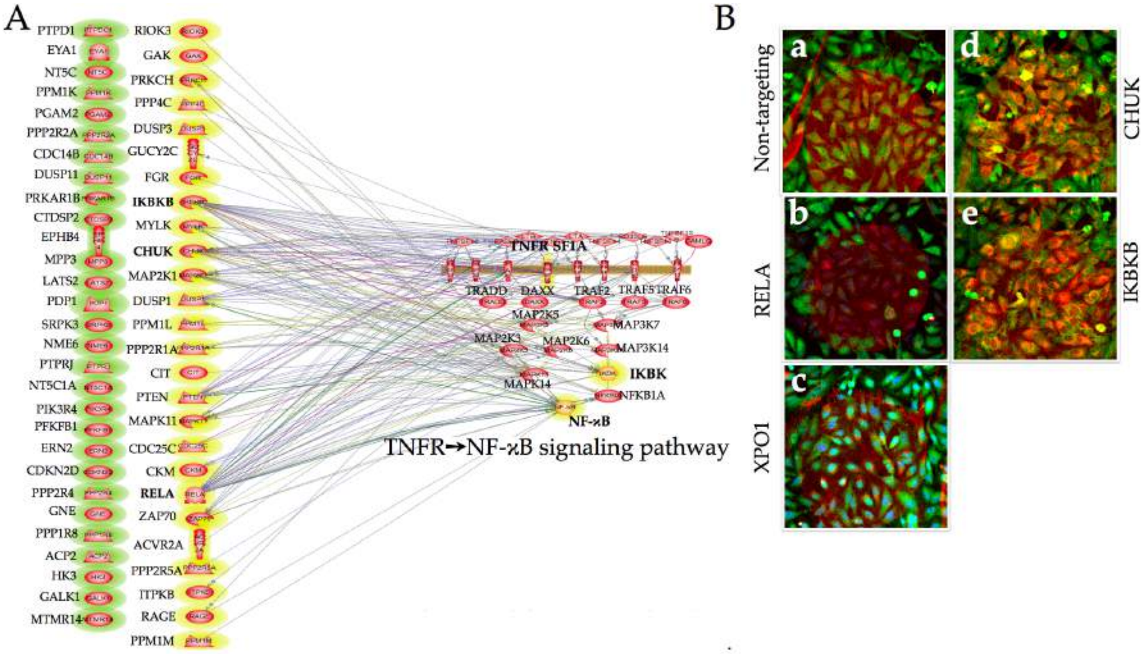


Figure 5.10: Essential interactions of hits with the NF-κB pathway: Silencing of non-targeting, p65, XPO1, CHUK and IKBKB directed siRNA printed spots.

[A] Yellow circles represent genes that have been reported by literature to directly interact with the TNFR to NF-κB signaling pathway while green circles are genes not currently reported to interact directly with TNFR to NF-κB signaling pathway. [B] Extrapolated immunofluorescent images from the screen showing HeLa cells overlaid on non-targeting, p65, XPO1, CHUK and IKBKB siRNA printed spots. Transfected cells were subsequently treated with 100 ng/ml of TNF-α and cell p65 and nuclei were counter-stained with p65 rabbit polyclonal antibody and DRAQ5 stains respectively. Images were acquired at 10X using a confocal imaging system. As shown in (a) the non-targeting siRNA had no effect on p65 nuclear translocation, (b) the p65 siRNA resulted in a decrease in p65 expression within the cell, (c) the XPO1 siRNA resulted in p65 accumulation in the nucleus while CHUK and IKBKB siRNAs (d and e) prevented the p65 from entering the cell nucleus.

5.3.2 Conclusion

Our intention in this study was to show that the high-density siRNA array described in chapter four could be used effectively in screening for human host factors taking part in biological pathways. Also, we wanted to demonstrate that by using the PCA, datasets generated from screening could be adequately separated into different phenotypic classes.

In this objective, we developed and validated a phenotypic assay of TNF-α induced translocation of p65 (RELA/NF-κB) into the HeLa cell nucleus. Genes that impact the nuclear import pathway were screened for using siRNA arrays covering the human kinome as well as all phosphatases, ubiquitinylases, and proteases.

Immunofluorescent images from the screen showed that the siRNA printed spots were localized during screening. The nucleic acid activity within the printed spots was observable as depicted by the internal p65 (RELA) and XPO1 controls siRNAs. Also the array and cell viability on overlaid printed spots (excluding toxic siRNA spots) was optimal (Figure 5.10 B).

Using a multi-dimensional analysis (PCA) we were able to separate the datasets generated from the screen into different classes based on the similarities in cellular phenotype expressed by the individual genes (Figure 5.7-5.9). To further demonstrate the effectiveness of PCA in data separation, we selected a known gene (RELA) within the library that codes for p65 and looked for its relative position within the gene cloud. As expected, the RELA gene was located within the p65 control cloud (Figure 5.8) thus confirming the effectiveness of the PCA in placing genes expressing a similar phenotype into the same phenotypic group.

Using the PCA we identified 51 genes as implicated in or controlling the TNFR to NF-κB (p65) signaling and translocation cascade. All 51-candidate genes (hits) came from a single cluster located directly opposite the XPO1 cloud. Based on the influence generated by PC-1, genes within this cluster were responsible for preventing p65 from entering the nucleus (Figure 5.10 B). Using Pathway Studio, a data mining and visualization software that rapidly connects relevant biological events based on existing literatures, we were able to demonstrate that 26 of the obtained hits were known direct interactors to the NF-κB pathway (Figure 5.10 A). Among the identified hits, two of the genes CHUK and IKBKB are known key regulators of TNFR to NF-κB (p65) signaling pathway (Connelly and Marcu 1995; Regnier, Song et al. 1997; Rothwarf, Zandi et al. 1998; Mercurio, Murray et al. 1999; Yamamoto and Gaynor 2004). Therefore, the other 25 genes could be potential new players in the NF-κB (p65) pathway.

Based on the above findings, it was reasonable to assume that the high-density array platform and data mining approach used in this screen was effective and could be used in screening for human host factors involved in more complicated biological assays such as the *L. monocytogenes* infection assay.

CHAPTER 6

IDENTIFICATION OF HUMAN HOST FACTORS MEDIATING *L. MONOCYTOGENES* INFECTION USING RNAi SCREENING

ABSTRACT

Intracellular bacteria pathogens, such as *L. monocytogenes* orchestrate their internalization by subverting cellular functions through the interaction with specific host adaptor proteins. The bacteria initiate this process by means of its virulence proteins. Several of the bacterial proteins leading to their invasion and proliferation have been established. Identifying the corresponding human receptor proteins hijacked during the bacteria invasion is currently the focus of most research studies. To address this, we use high-density siRNA microarrays and computer-assisted image analysis procedures to quantify for intracellular bacteria in HeLa cells and subsequently look for potential human hosts factors involved in *L. monocytogenes* infection.

Here we developed a GFP-*L. monocytogenes* invasion assay using a 96-well plate format. Using a computational based-approach, we quantified for the average number of bacteria within the cell as well as the morphological changes on the cell surface resulting from the bacterial infection. The assay was subsequently transposed onto a high-density siRNA microarray plate comprising of the kinases, phosphatases, ubiquitinases and proteases to screen for human host factors involved during *L. monocytogenes* infection. The screen had a total of 15 replicates and using a customized algorithm, we identified 65 candidate genes whose silencing visually impaired or enhanced bacterial invasion/proliferation. The 65 hits were further validated in a secondary screen by forward transfecting HeLa cells with the individual pooled siRNAs. Using a four standard deviation cut-off mark, we selected the 15 strongest genes whose silencing greatly enhanced bacterial invasion.

Among the hits selected were transcription factors, genes involved in protein phosphorylation, proteolysis, intracellular signal transduction and in the regulation of chromatin assembly or disassembly.

6.1 Introduction

The bacterial pathogen *L. monocytogenes* has evolved mechanisms (virulence factors) enabling it to invade the host cells, subvert and undermine the host immune system to propagate its survival. Most of the research work carried out over the past two decades to understand how this intracellular pathogen invades and proliferates within the host cells has been centered on the bacteria produced virulence factors (Agaisse, Burrack et al. 2005).

Some of the bacterial virulence factors whose functions have been largely linked to host cell invasion include; Internalins, surface protein p104, Listeriolysin O (LLO), Phospholipases (PLC), Protein p60, ActA protein, Clp proteases and ATPases. The bacteria uses its virulence factors to exploit host proteins and propagate its invasion, phagosomal escape, proliferation, host actin polymerization, and cell to cell spread (Vazquez-Boland, Kuhn et al. 2001; Lee and Liu 2008; Cossart 2011; Ireton 2013). The human host factors implicated in this process have not been systematically described.

The approach in this objective was to use the siRNA interference technology along with a reverse transfection strategy to silence all human genes in the kinases, phosphatases, ubiquitinylases, and proteases then evaluate their role, if any, in cellular infection.

6.2 Results and discussion

6.2.1 *L. monocytogenes* infection assay

The assay developed for this screen was geared at measuring the intracellular bacteria per cell after infection with *L. monocytogenes* To achieve this, two human cell lines were used in validating the assay: human epithelial carcinoma cells (HeLa) and human osteosarcoma cells (U2OS). Both cell lines have been used as model cell lines in the validation of many infectious assays (Niforou, Anagnostopoulos et al. 2008; Chong, Swiss et al. 2009; Balestrino, Hamon et al. 2010; Stavru, Bouillaud et al. 2011).

We first defined the optimal multiplicity of infection (MOI) concentration of the bacteria required for infection and an optimal time course for invasion and proliferation of the pathogen within the host cell. To enable easy identification of bacteria without the use of secondary staining, a *L. monocytogenes* strain expressing the green fluorescent protein was used (GFP-*L. monocytogenes*). Quantification of the intracellular GFP expressing bacteria after cells were exposed to a titration of the GFP-*L. monocytogenes* was performed in fixed cells counter-stained with Hoechst 33342. Hoechst 33342 stain is a blue fluorescent dye with excitation and emission wavelength of around 350 nm and 461 nm, which enabled bacteria to be detected without the need for GFP expression. Once the bacteria pathogen is internalized, its uses its virulence factor ActA to polymerize the host actin, generating an actin rocket tail which it then uses for its propagation within the cell cytoplasm. Therefore to ensure complete internalization of the bacteria, it was important to image the cell actin cytoskeleton. We demonstrated that staining the cell actin cytoskeleton with phalloidin 647 (excitation/emission: 650/668 nm) we could identify internalized bacteria that have acquired the actin rocket tail. Figure 6.1 shows a schematic diagram illustrating the various steps used in developing the infection assay.

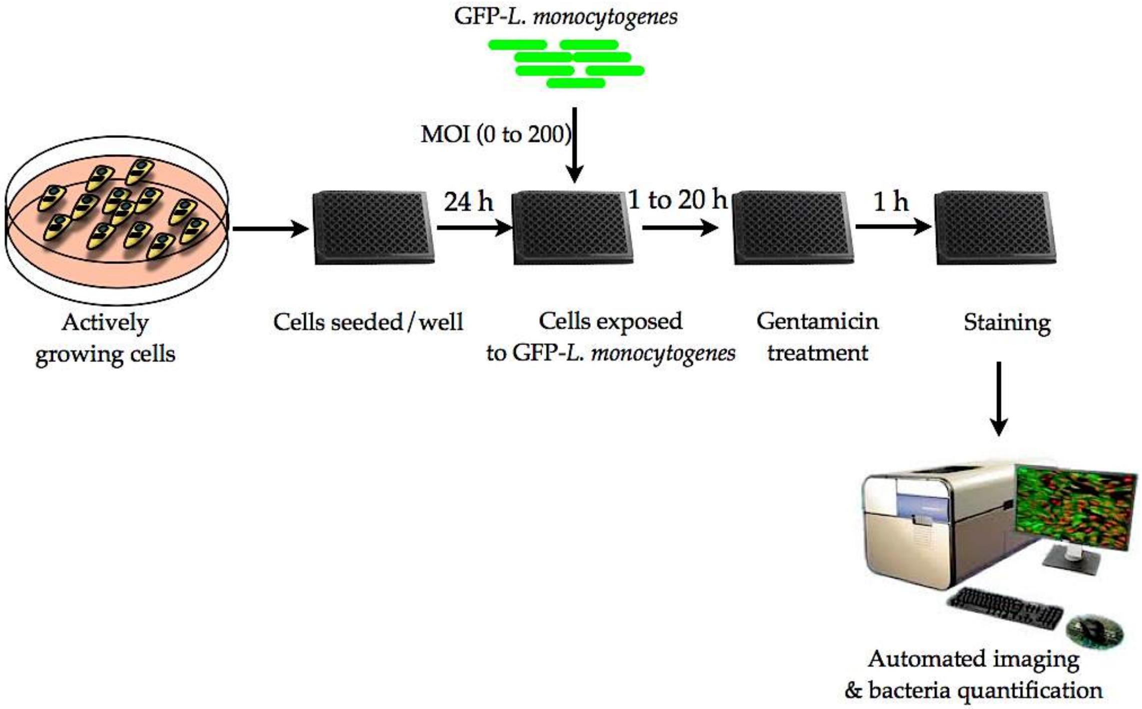


Figure 6.1: Stages involved in *L. monocytogenes* invasion assay.

Schematic demonstrating the approach used in performing the *L. monocytogenes* infection assay.

6.2.2 Image interpretation by visual inspection

It order to develop an algorithm that will robustly measure for the intracellular bacteria per cell for subsequent infected images, it was necessary to select an image where cells infected with the bacteria depicted an observable infection characteristic. To demonstrate this, HeLa cells were cultivated for 24 hours in a 96-well plate and later exposed to a titration of GFP-*L. monocytogenes* (MOI: 0, 25 and 100) for 2 hours. Cells were fixed, permeabilized and cell nuclei and cell actin cytoskeleton counter-stained with 12 ng/ml Hoechst 33342 and 0.1 μM phalloidin 647 stains respectively. Images were acquired at 10X using a confocal imaging system (Figure 6.2).

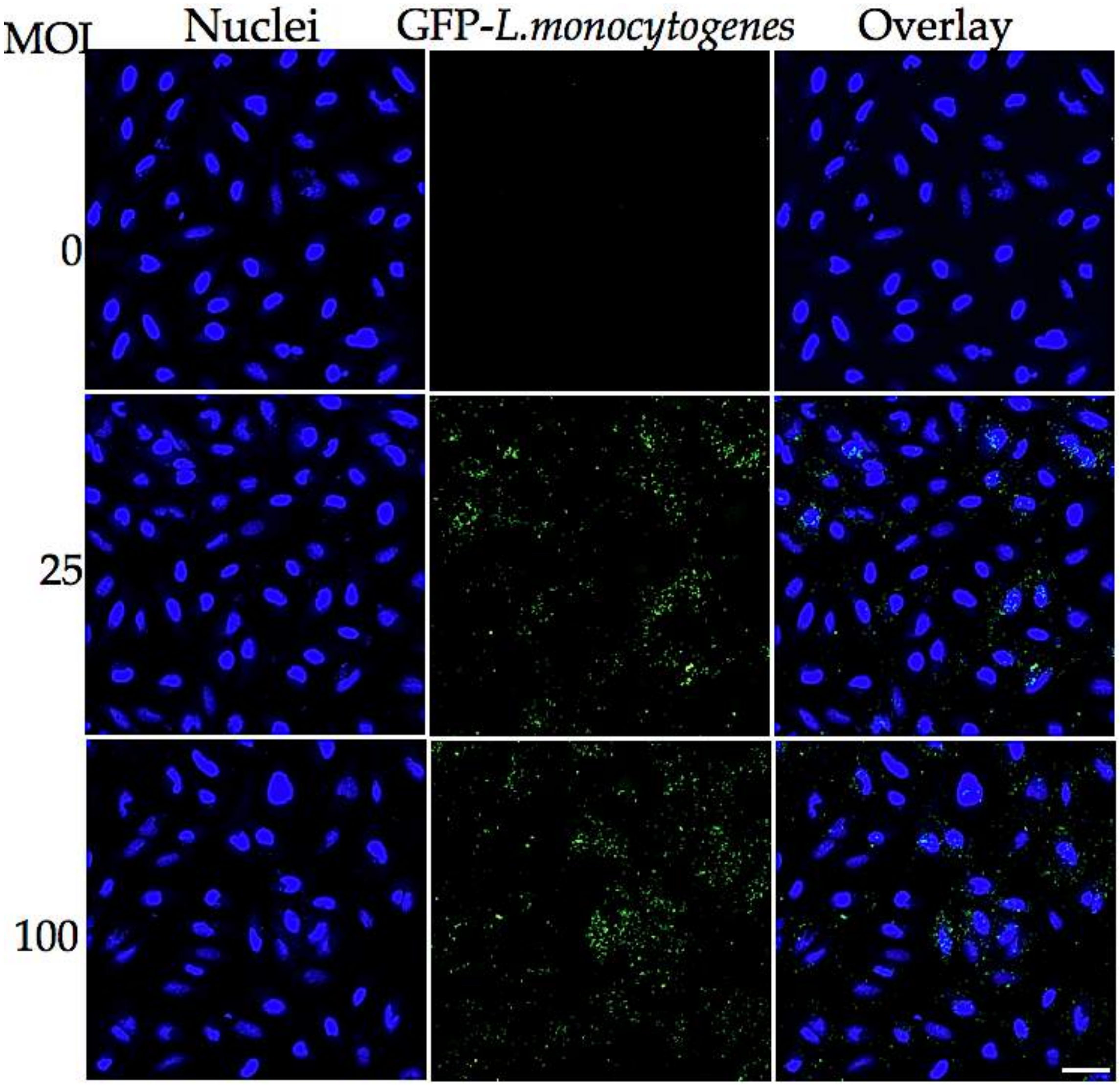


Figure 6.2: HeLa cells infected with a titration of GFP-*L. monocytogenes*.

HeLa cells were infected for 2 hours with different MOI concentrations (0, 25 and100) of the bacteria. Post-infection, cells were fixed, permeabilized and cell nuclei counter-stained with 12 μg/ml Hoechst 33324. Images were acquired at 10X using a confocal imaging system. Column 1 represents the cell nuclei (blue); column 2 the GFP-*L. monocytogenes* (green) and column 3 represents overlay images of both the nuclei and GFP-*L. monocytogenes*. Scale 25 μm.

In this experiment we further illustrated that the bacteria were not lying on the cell surface (that may affect intracellular bacteria quantification), but had invaded the cells and used the cell actin network to generate actin rocket tails, which they use to propel themselves within the cell cytosol (Figure 6.3).

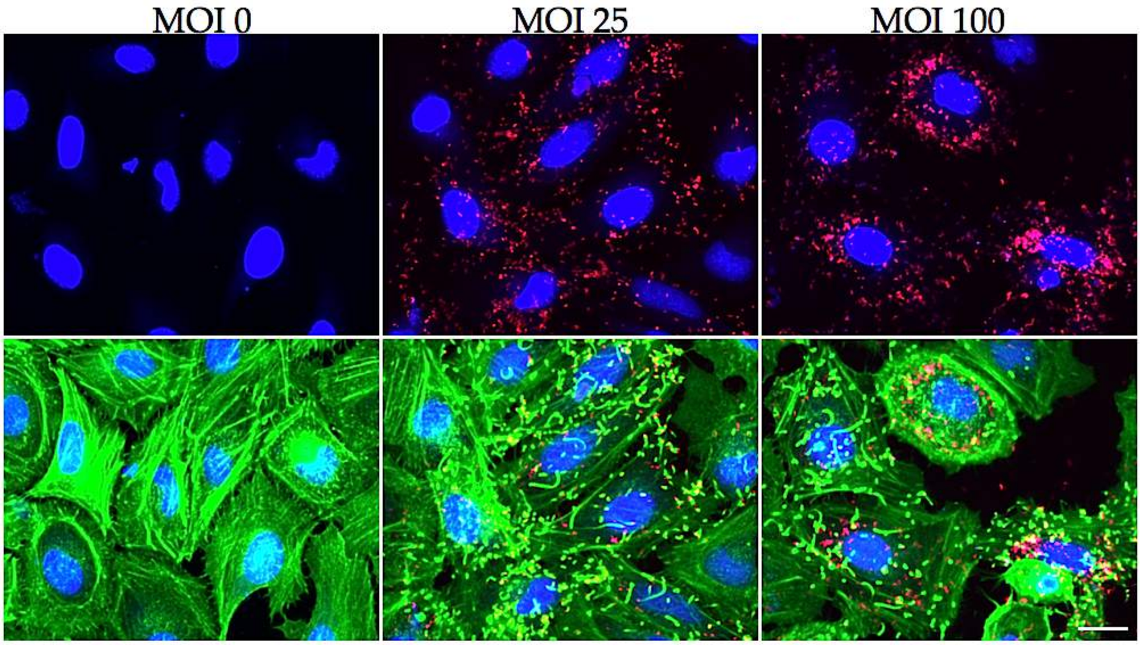


Figure 6.3: Actin stained images of HeLa cells infected with different concentrations of GFP-*L.monocytogenes*.

HeLa cells infected with an increasing MOI (0, 25 and 100) of GFP-*L. monocytogenes* for 3 hours. Post-infection, cells were fixed and permeabilized. The actin cytoskeleton and cell nuclei were counter-stained with 0.1 µM phalloidin 647 and 12 μg/ml Hoechst 33324 respectively. Images were acquired at 40X using a confocal imaging system. Top panel shows overlay images of the Hoechst (blue) and GFP-expressing bacteria colour coded as red while the lower panel represents the phalloidin stained images. Bacterial actin rocket tails are represented as tiny green fiber within the cell cytosol of the phalloidin stained images. Scale bar 40 μm.

6.2.3 Quantification of intracellular bacteria post infection

This assay was designed to measure the intracellular bacteria per cell after exposure of cells to the bacteria. Inconsistency in intracellular bacteria quantification will lead to erroneous data, thus affecting the end results of the screen. It was therefore necessary to design an algorithm that will robustly quantify for the actual number bacteria per cell. To achieve this, the intracellular bacterial concentration was measured by image analysis. This analysis was geared at obtaining an optimal MOI for infection and infection time for subsequent assays.

The aim here was to generate a bacteria quantification software capable of [i] automatically finding cells and [ii] quantifying how many bacteria were present per cell.

To demonstrate the approach used in quantifying for the intracellular bacteria, 7.5 X 10^3 HeLa cells/well were cultivated in 96-well plates for 24 hours. Cells were later infected with the GFP-*L. monocytogenes* for 2 hours post infection; extracellular bacteria were removed by gentamicin treatment. Infected cells were subsequently fixed, permeabilized, and cell nuclei counter-stained with Hoechst nuclear stain before imaging (Figure 6.4).

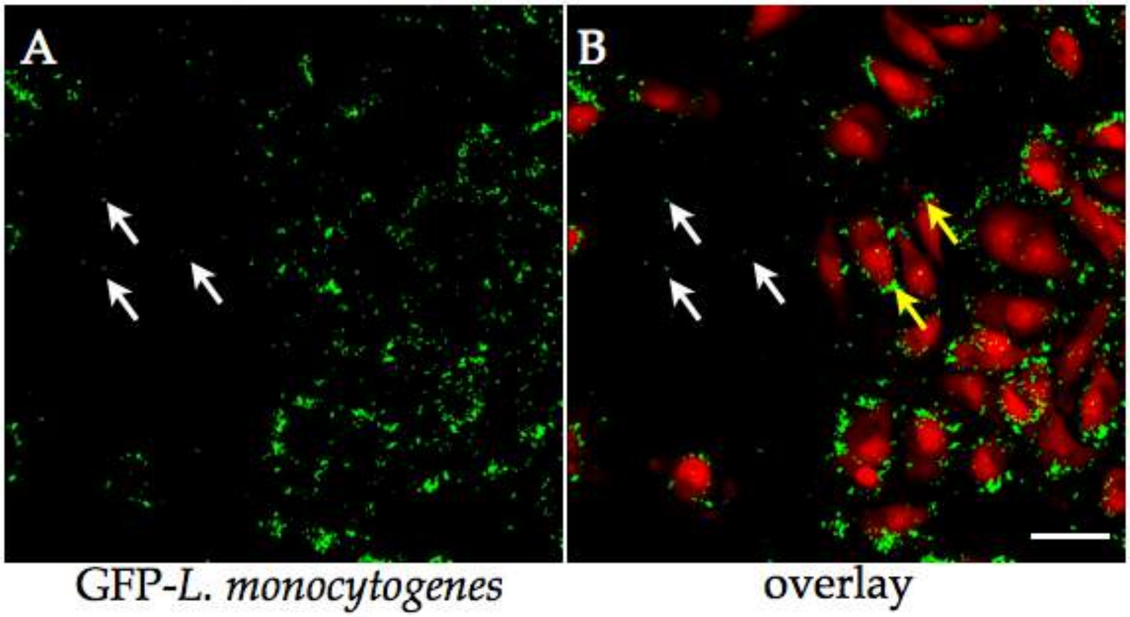


Figure 6.4: Image for intracellular bacteria quantification

HeLa cells infected with GFP *Listeria*. [A] Shows the GFP-labeling and [B] the overlay with Hoechst staining. Intracellular bacteria are labeled with yellow, extracellular with white arrows. Scale bar 16 μm.

A critical challenge was to be able to find the infected cells and exclude any extracellular bacteria from entering analysis. Extracellular bacteria, if not excluded from the analysis, will result in misleading data. This was addressed by using an adaptive algorithm to first identify cell and nuclear boundaries in images (Figure 6.5 a). We then expanded the area recognized using a cell recognition algorithm to create a mask corresponding to the footprint of the cells in the image (Figure 6.5 b). Automated thresholding (Figure 6.5 c) of the cell footprint segmentation image then automatically generated a binary mask (Figure 6.5 d) and an inverted mask (Figure 6.5 e) of the cell footprint. This detection is independent of the level of infection and not affected by it. The inverted mask was then used to isolate the cell footprint image (region of interest) from both the Hoechst and GFP labeling images (Figure 6.5 g & h) while the excluded image portion from processing is shown in figure 6.5 f, as an example. Extracellular bacteria are essentially completely excluded from image analysis and only the excised cell (6.5 g) and intracellular bacteria images (6.5 h) are analyzed from this point onward.

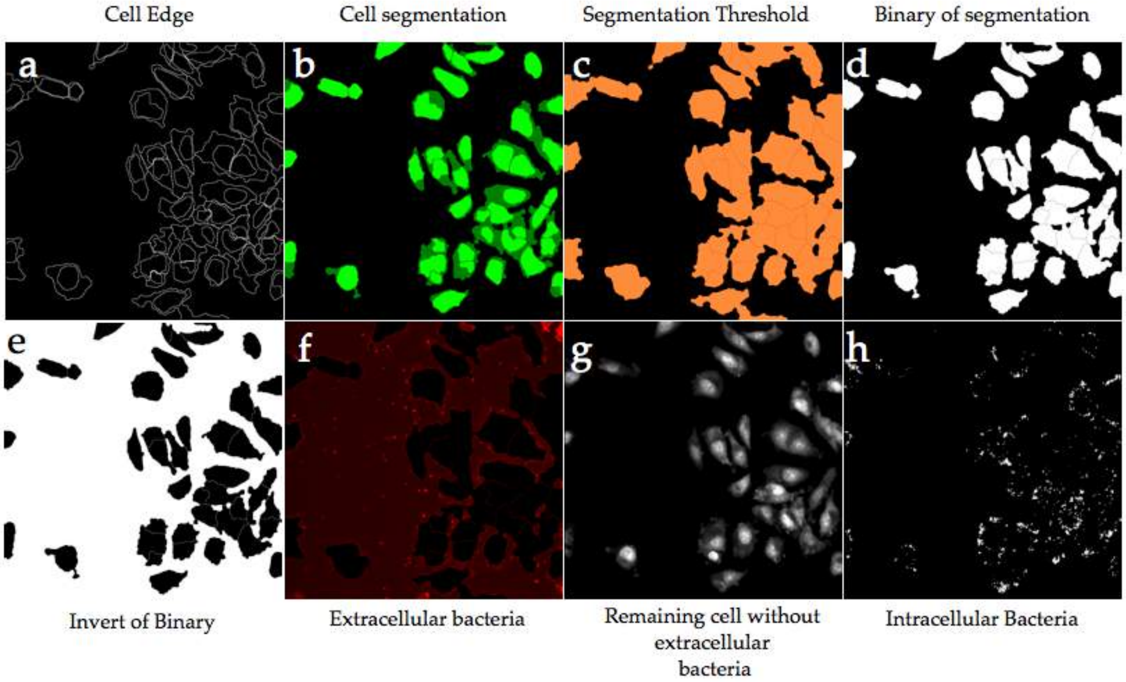


Figure 6.5: Steps used in intracellular bacteria quantification

After developing the algorithm to extract infected cells without any contribution from extracellular bacteria, the next step was to use the processed images for automated analysis and recognition of intracellular bacteria. The algorithm used was selected to detect bacteria within cells based on a predetermined size and shape. It is not particularly sensitive to bacterial particle intensity instead it exploits the 'shape' of the bacteria within the cell image. Given the use of a thin confocal section for acquisition, bacteria lying on top of the cells were effectively removed following a Z-series acquisition through the infected cells (not shown).

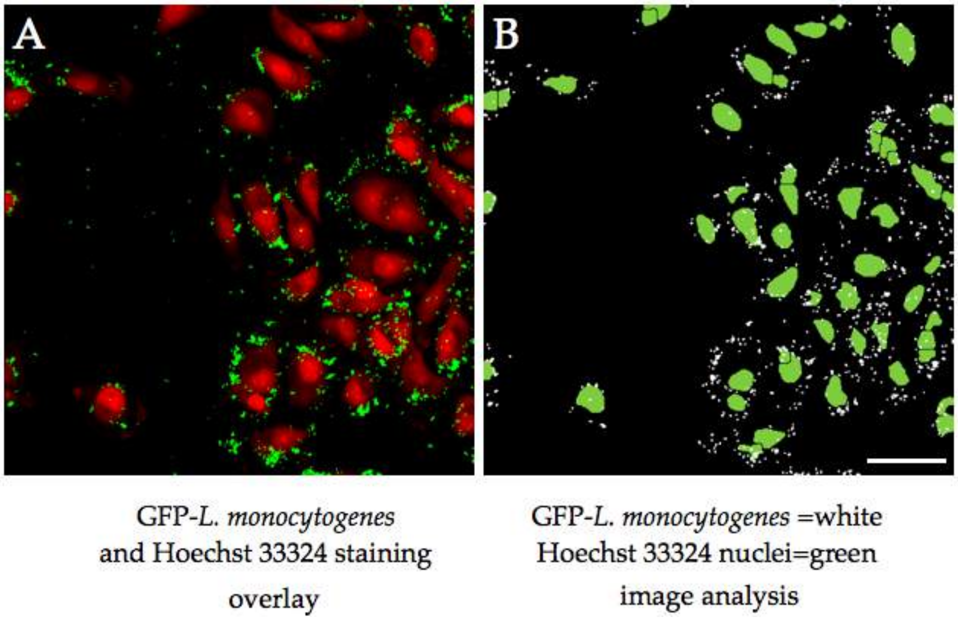


Figure 6.6: GFP-*L. monocytogenes* image analysis

GFP-*L. monocytogenes* infected cells [A] and the image recognition result [B] with the detected bacteria shown in the right panel. Scale bar 16 μm.

Identified bacteria are shown in white surrounding the automatically detected cell nuclei colour coded green (Figure 6.6 A and B). This approach robustly identifies cells, irrespective of cell number/image, and excludes extracellular bacteria because cell free regions of the image are identified and excluded (Figure 6.6). The image shown in panel B was then automatically segmented into individual cells and the number of bacteria per cell quantified and exported for analysis.

To validate the effectiveness of the software in intracellular bacteria quantification, the algorithm was used to quantify intracellular bacteria after cells were exposed to a range of MOI for the bacteria pathogen. It was predicted that the algorithm should be capable of rigorously detecting differences in infection across images from HeLa and U2OS cells infected with a titration of MOI.

Cells were infected with GFP-*L. monocytogenes* at increasing MOI (0 to 100 units). Images of cells were then analyzed using the bacterial counting algorithm 1 and 2 hours post-infection (Figure 6.7).

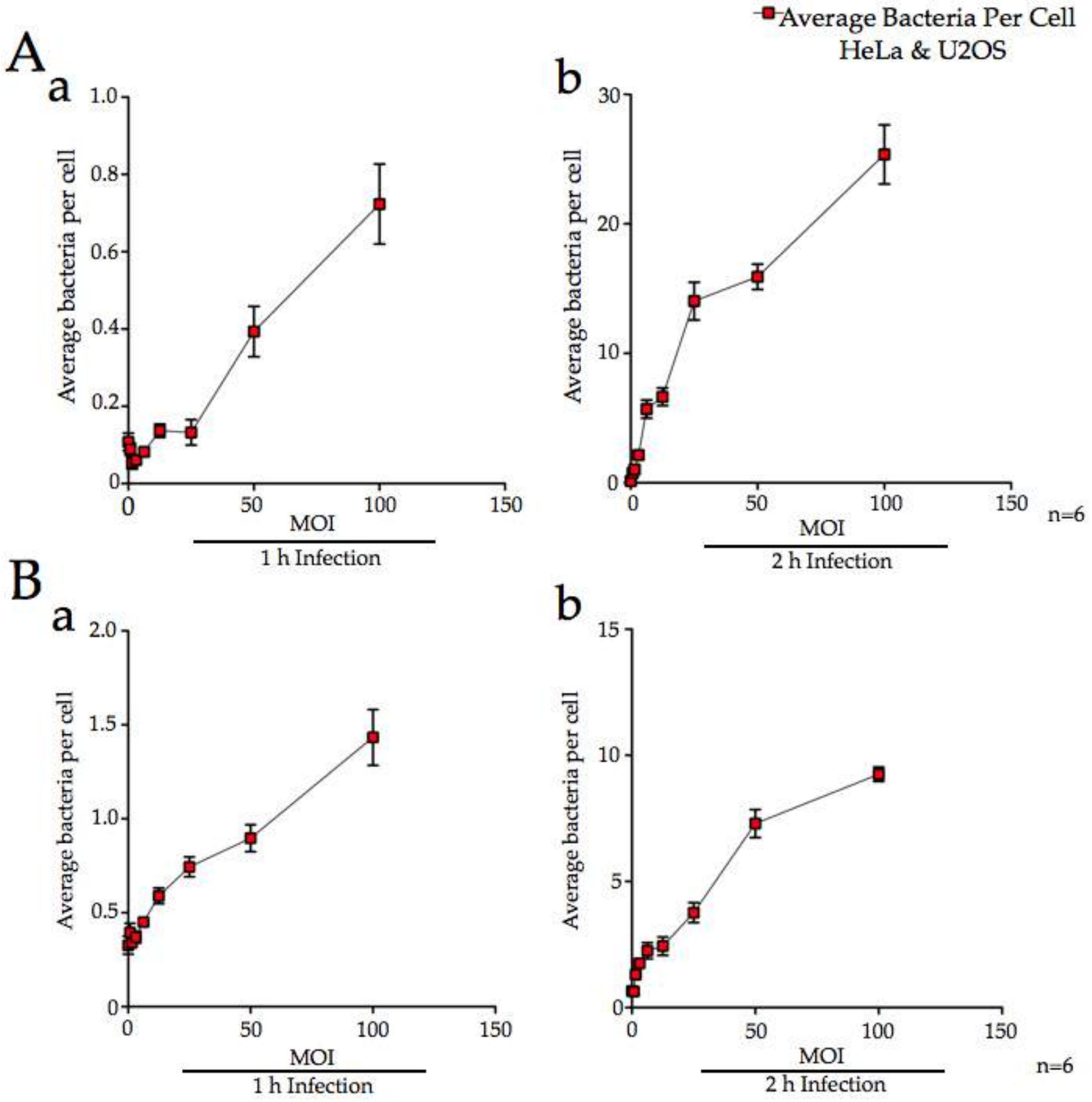


Figure 6.7: Graphical quantification of intracellular bacteria

Average intracellular bacteria per cell measurement after exposing HeLa [A] and U2OS [B] cells to a titration of GFP-*L. monocytogenes* (MOI 0 to 100). (a) Quantification of intracellular bacteria per cell after 1 hour of infection and (b), the intracellular bacteria per cell after 2 hours of infection. For clarity and better graphical representation, the y-axis for each graph was scaled to depict the average intracellular bacteria per cell.

These data demonstrated a semi-linear relationship between the MOI and the average bacteria per cell and also shows a graphical representation that depicts a standard bacteria growth curve. We could then conclude that the approach set for intracellular bacteria quantification was suitable. To measure the robustness of the analysis, T-tests were used to compare the detected numbers of bacteria in both the

HeLa and the U2OS cells. The T-test was performed on experiments with three MOI titrations (i.e. 0, 25 and 100). The T-test shows significant difference (p<0.0001) between the various MOI titration used (Figure 6.8).

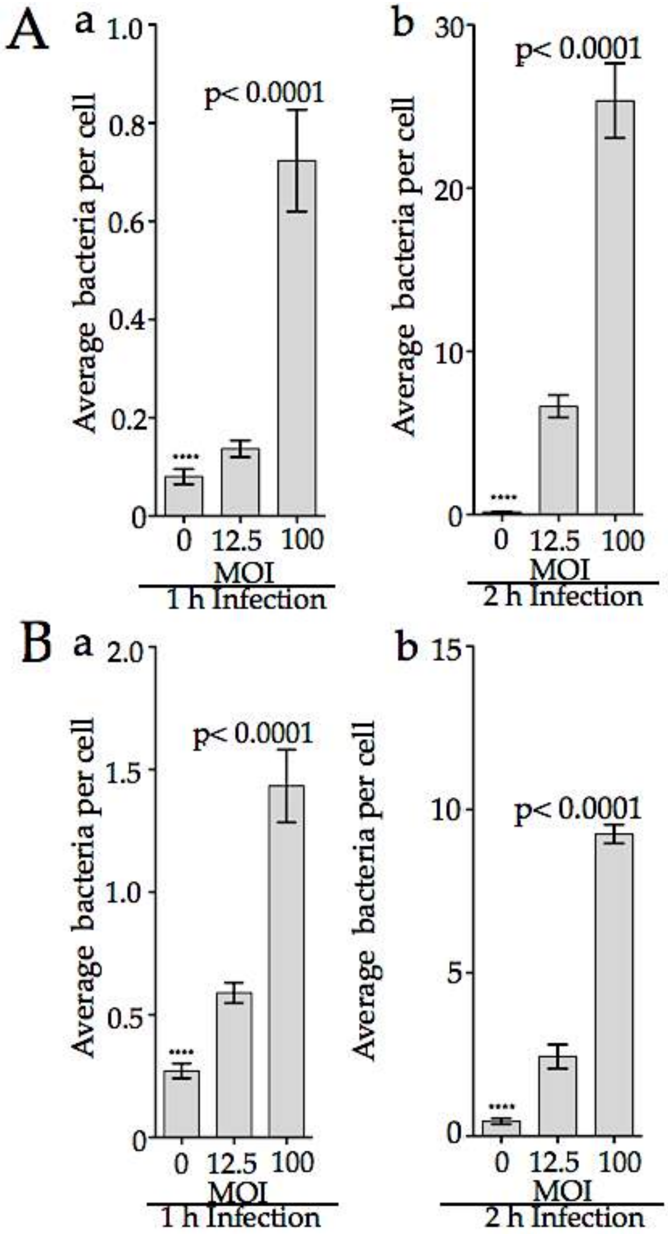


Figure 6.8: Student T-test for intracellular quantification.

Image analysis. Quantification of average the bacteria per cell after 1 hour (a) and (b) 2 hours post infection in HeLa [A] and U2OS [B] cells. Student T-test gives a significant difference between uninfected (MOI=0) and infected cells (MOI=100) (P < 0.0001, n= 6). For clarity and better graphical representation, the y-axis for each graph was scaled to depict the average intracellular bacteria per cell.

Therefore the bacterial counting algorithm used was effective in measuring the intracellular bacteria per cell in both cell lines. The bacteria counting algorithm was then implemented in all subsequent experiments to measure the number of intracellular bacteria after cells were exposed to the bacteria.

6.2.4 Infection time course

Cellular processes such as apoptosis, anti-apoptosis and inflammatory related signaling pathways are reprogrammed after bacterial infection to help the cell to survive the stress induced by the infection (Cossart and Sansonetti 2004). However, prolonged exposure of cells to pathogenic bacteria decreases cell viability, which in turn leads to cell death. Since the assay developed for this study was directed at measuring the bacteria present in cells, it was therefore reasonable that an optimal MOI and time course for infection be determined that minimized cell death. To investigate the optimal MOI and time course for infection, HeLa cells were infected with the GFP-*L. monocytogenes* at different MOI concentrations (0 to 100). Cells were exposed to this pathogen over a period of 20 hours. At different time points post infection (1, 2, 3, 4, 10 and 20 hours), cells were fixed, permeabilized, and DNA stained with 12 μg/ml Hoechst 33324. Treated cells were imaged and using the automated analysis package as described above, the intracellular bacteria and intracellular bacteria per cell were quantified. Data points were extracted and plotted as in figure 6.9.

These results demonstrated the proliferation of the GFP-*L. monocytogenes* within the cells. For clarity, arrows and dotted lines were introduced to monitor bacterial proliferation. Using an MOI of 25, the growth of bacteria was monitored at the different infection time points. It was observed that at this MOI concentration, bacteria proliferation was at an exponential growth phase (light grey arrows). However, at 4 hours post-infection at MOI 25, bacteria attained a maximum growth phase (brown arrow). After 4 hours of infection, further exposure of the cells to the bacteria resulted in depletion of the intracellular bacteria quantified (dark grey arrows). The decline in intracellular bacteria after 4 hours of cells being exposed to the bacteria could be attributed to: [i] cells losing their viability due to continuous exposure to the pathogen and [ii] bacteria lacking the necessary ingredients to continue proliferation. It was therefore reasonable to set the optimal infection conditions at a time point and MOI concentration where cells exposed to the pathogen will neither be under or over-exposed to the pathogen (that is at the bacteria exponential growth phase). The optimal infection conditions were therefore set at MOI 25 and 3 hours for infection. These infection conditions were subsequently used for all future experiments.

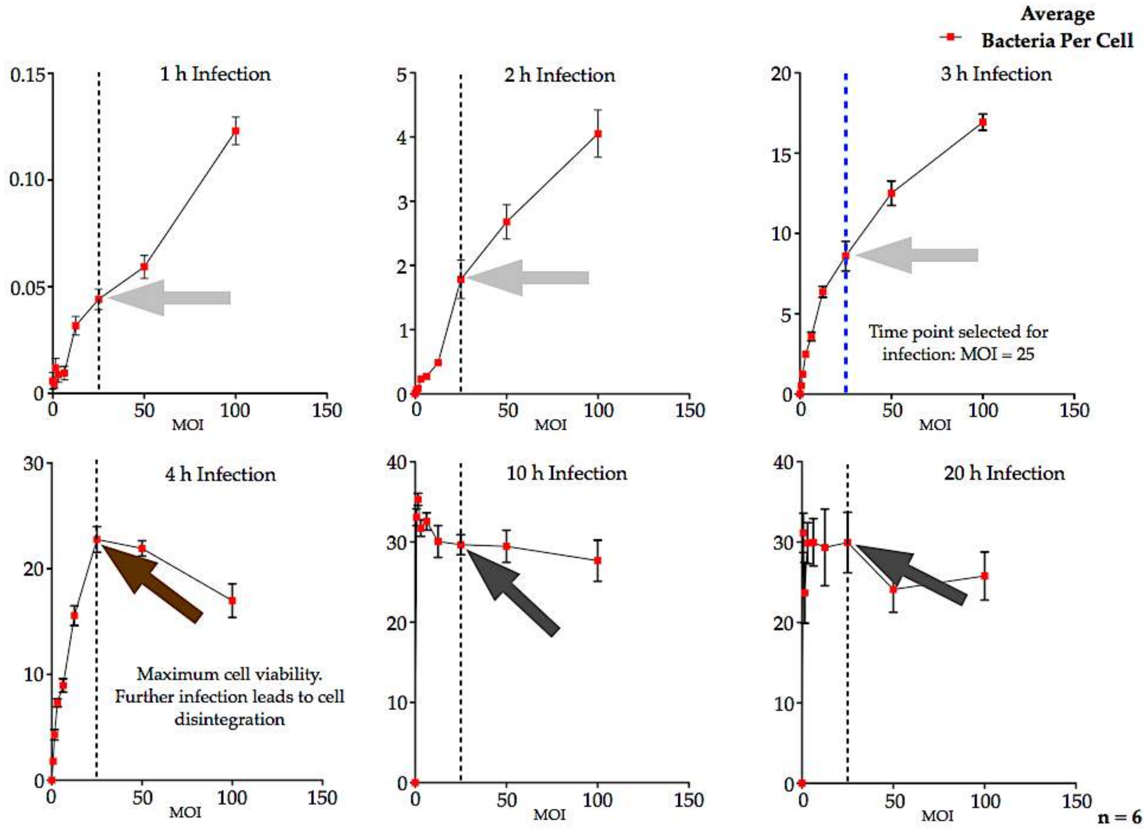


Figure 6.9: GFP-*L. monocytogenes* infection time course analyses

HeLa cells were cultured for 24 hours and later infected with a titration (MOI 0 to 100) of *GFP-L. monocytogenes*. Bacteria cell invasion proceeded at different time points (1-4, 10, 20 hours). After each time point, cells were treated with gentamicin. Post infection, cells were fixed and permeabilized. Images were acquired using a confocal imaging system. Intracellular bacteria were quantified using the image counting algorithm developed as described in 6.2.3. Each data point is a representation of six individual experiments (n = 6). Arrows and dotted lines were used to monitor bacteria proliferation within the cells. Using MOI of 25 as an arbitrary reference point, the light grey arrows depict the exponential phase of the bacteria, the brown arrow show the stationary phase while the dark grey arrows show the declined phase of the bacteria during the invasion. MOI of 25 was used for all subsequent infection assays since based on the time course experiment, at MOI 25, bacteria were generally in their actively growing phase. For good representation, the y-axis scales were adjusted for each graph.

6.2.5 Bacterial actin quantification

L. monocytogenes are known to propagate within the host cytoplasm or dissipate to adjacent cells by exploiting the actin cytoskeleton (Vazquez-Boland, Kuhn et al. 2001). Image analysis of the actin cytoskeleton has been used for quantifying *L. monocytogenes* infection (Henmi, Tanabe et al. 2011). To exploit this as a measure of infection, the actin cytoskeleton was stained with red fluorescent phalloidin. Phalloidin is a seven amino acid peptide toxin isolated from *Amanita phalloides* mushroom that specifically bind the polymerized form of actin (F-actin) with high-affinity (KD 20 nM) (http://www.cytoskeleton.com/phdg1). F-actin is a linear polymeric microfilament formed after an assembly of the G-actin monomers (Carlier and Pantaloni 2007). In this work we made use of phalloidin conjugated to Alexa 647 dye (excitation/emission wavelength 650/668 nm) to avoid crossover into imaging channels used for cell or spot staining.

Actin cytoskeleton is found in all cell types and they provide the cell with structure and shape. Because of its size, the cell actin cytoskeleton can change cell morphology simply by assembling or disassembling itself (Lodish, Berk et al. 2000). *L. monocytogenes* once in the cytoplasm of the cell, capitalizes on the actin network to polymerize actin rocket tails, which they use for mobility within the cell cytoplasm. It is therefore reasonable to believe continuous assembly of the actin rocket tail by the pathogen may alter the general stoichiometry of the actin cytoskeleton thus leading to changes in cell morphology. To access changes in cell surface morphology resulting from the latter process, parameters such as Cell Texture Index, Cell, Cellular Texture Index, Cell Gradient Index, Cell Cellular Gradient Index, Cell Laplacian Index and Cell Cellular Laplacian Index were used. The above measuring parameters mathematically calculate the local curvature of grayscale intensities in an image and serves as a method for picking up edges that is independent of edge orientation. Edges in images are areas with strong intensity contrasts.

To demonstrate the changes in cell morphology resulting from *L. monocytogenes* infection, HeLa cells were infected for 3 and 4 hours at MOI 25. At 3 and 4 hours post infection, most of the bacteria should have invaded the cells and capitalized on the host actin network to polymerize actin rocket tails which they use to move within the cell cytosol (Figure 6.3). Post infection, cells were fixed, permeabilized and the nucleus and actin cytoskeleton of the cells stained with 12 μg/ml Hoechst 33324 and 0.1 µM Phalloidin 647, respectively. Cell images were acquired using a confocal microscope. To measure for changes in cell surface morphology resulting from the infection, the cell surface measuring parameters listed above were used (Figure 6.10). We demonstrated that there was a direct relationship between the cell surface morphology and the number of bacteria (Figure 6.10: 3 hours infection). This linear relationship was however affected after 4 hours of infection at MOI 25. As shown in the figure, at 4 hours of infection there was a decline in measured parameters used in quantifying for cell surface morphology. To assess the decline in cell surface morphology, a student T-test was performed on the total cell number after cells were exposed to a prolonged culture of GFP-expressing bacteria.

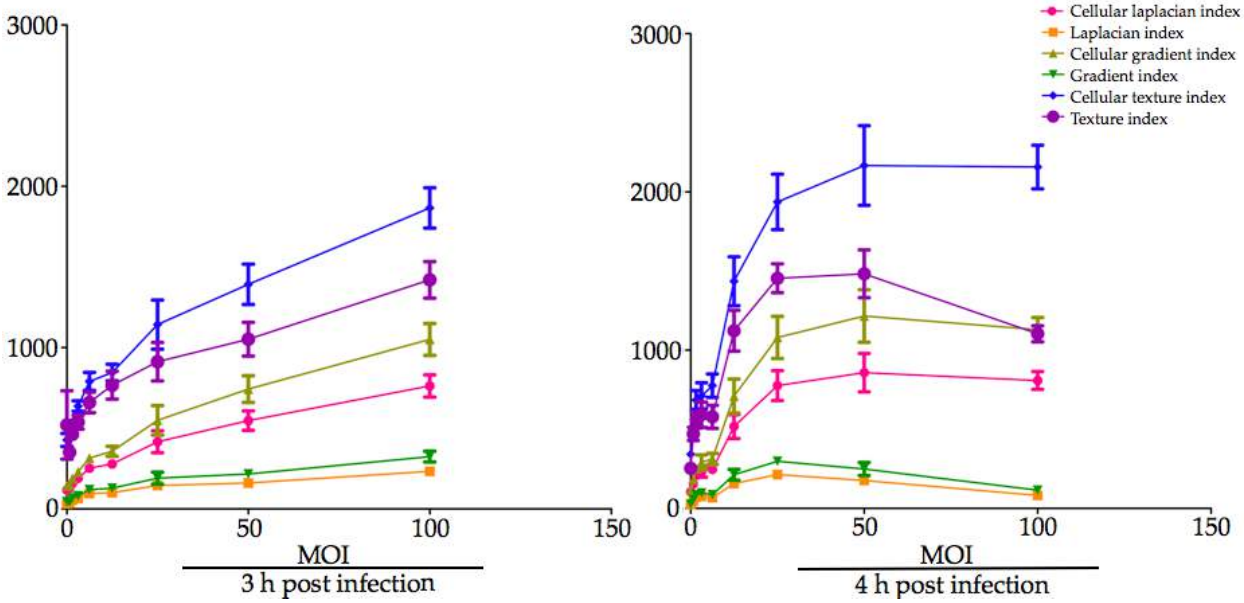


Figure 6.10: Correlation of cell actin morphology after HeLa cells were infected with GFP-*L. monocytogenes*.

HeLa cells were exposed to different MOI concentration for 3 and 4 hours. Post infection, the actin cytoskeleton was stained with 0.1µM Phalloidin 647 while the nuclei were counter-stained with 12 μg/ml Hoechst 33324. Images were acquired using a confocal microscope and changes in cell surface morphology were detected using measuring parameters such as cell texture index, cell cellular texture index, cell gradient index, cell cellular gradient index, cell laplacian index and cell cellular laplacian index.

Student T-tests were used to compare the total HeLa cell count after cells were exposed to the bacterial pathogen.

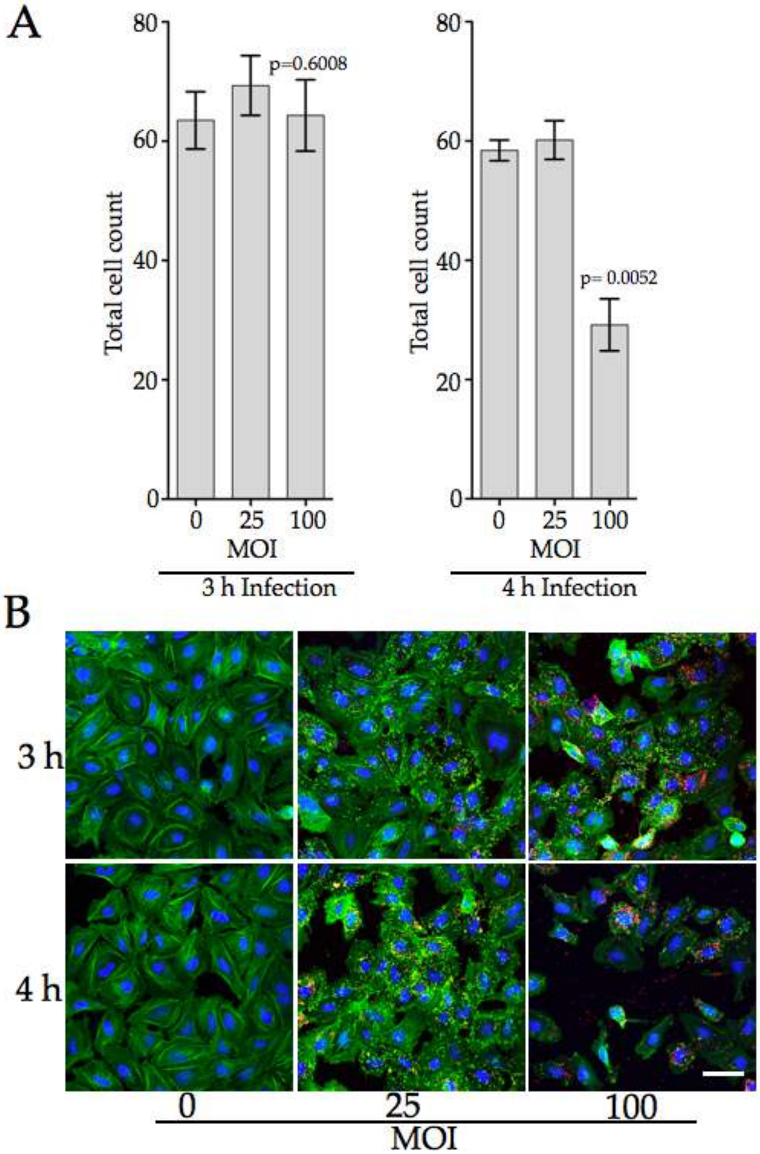


Figure 6.11: Quantification of HeLa cells post GFP-*L. monocytogenes* infection.

[A] Student T-test gives a significant difference in cell number between uninfected (MOI=0) and infected cells (MOI=100) ($P < 0.05$, n= 6) after 4 hours of infection while there was no significant difference observed at 3 hours post infection with the bacteria. [B] Immuno-labeling of HeLa cells infected with MOIs (0, 25 and 100) of GFP-*L. monocytogenes* for 3 and 4 hours. Cell nuclei and actin cytoskeleton were stained with Hoechst and Phalloidin 647 stains respectively. The upper panel shows cells infected for 3 hours while the lower panel illustrates cells after 4 hours exposure to the bacteria. Column 1, 2 and 3 show the different MOIs used during infection. Scale bar 16 μm.

The T-test was performed using three MOI titrations (i.e. 0, 25 and 100). The T-test shows significant differences ($p<0.05$) between the MOI 0 and 100 after cells were infected with the bacteria pathogen for 4 hours. However, no significant difference was observed in the total number of cells after cells were exposed to the pathogen for 3 hours (Figure 6.11 A and B). This result demonstrated that cells exposed to prolonged bacterial infection can only maintain their viability up to a certain point, after which they begin to disintegrate. This explains why there was a decline in cellular morphology quantification following prolonged exposure of the cells to the bacteria.

6.2.6 Conclusion

The goal of this objective was to develop optimal conditions for screening cells infected with GFP-*L. monocytogenes*. Results obtained identified an optimal MOI concentration and time course for infection (MOI: 25; time: 3 hours). The optimal MOI for infection was chosen within the exponential growth phase of the bacteria after cells were exposed to the bacteria for a prolonged period of time. Also demonstrated was the effect of *L. monocytogenes* invasion on cell surface morphology. It was observed that continuous exposure of the cells to the bacterial pathogen affected the cell actin assembly. Having shown that these conditions were optimal for intracellular bacteria quantification, we then transposed the set conditions onto siRNA microarrays to identify human proteins implicated in *L. monocytogenes* infection.

6.3 *L. monocytogenes* screen

6.3.1 Approach used in screening for human host factors involved in *L. monocytogenes* infection

The intracellular pathogen *L. monocytogenes* orchestrates its internalization, proliferation and cell-to-cell spread within the host cell by subverting the host defense machinery. Using the high-density microarrays validated for the TNF-α induced p65 nuclear translocation screen, we screened for genes implicated in *L. monocytogenes* infection. A total of 15 replicate screens were performed using arrays of 3150 siRNAs. This corresponded to 15 X 3150 = 47 250 individual experiments. To demonstrate the screening protocol used, HeLa cells were seeded onto the arrays and reverse transfected for 48 hours. Cells were then infected with the GFP-*L. monocytogenes* (MOI 25) for 3 hours. To remove external bacteria, infected cells were subsequently treated for 1 hour with gentamicin and cells were fixed, permeabilized and the cell actin cytoskeleton and nuclei counter stained with 0.1 μM Phalloidin 647 and 12 μg/ml Hoechst 33324 respectively. Images were acquired using an automated confocal imaging system. Images were exported, grid fitted and each siRNA spot was extracted as a single image. Using the same image counting algorithm and cell surface morphology quantification approach from the assay, analysis of the individual images were performed.

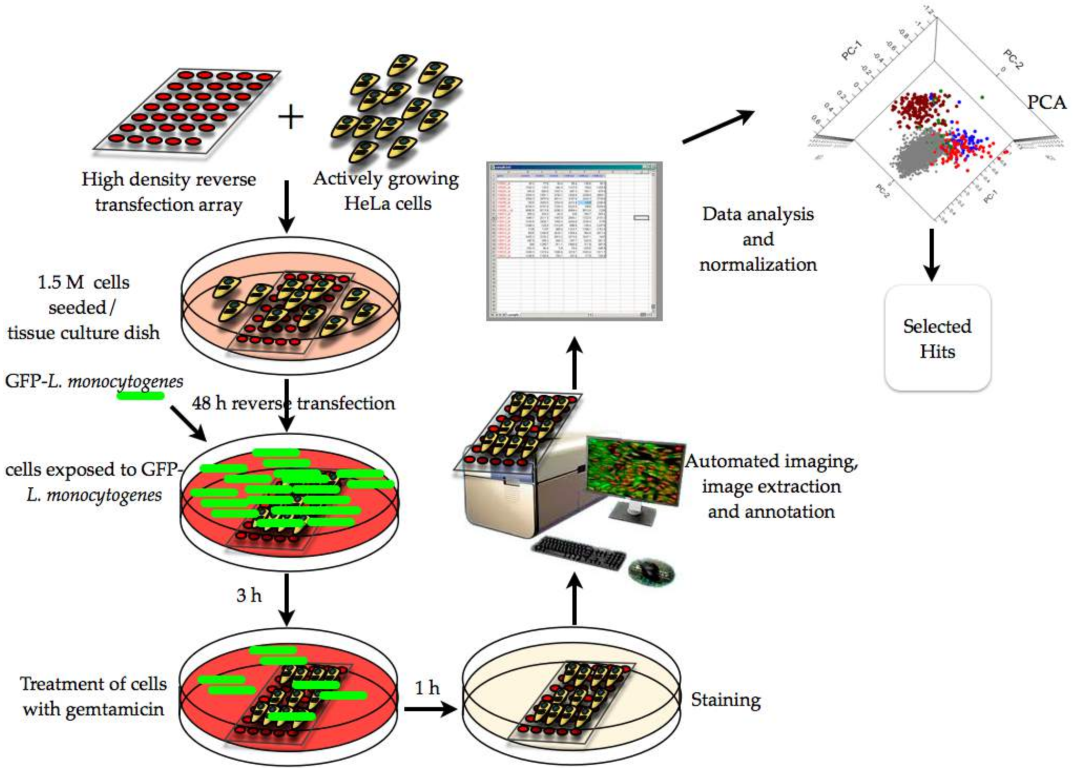


Figure 6.12: High content screen workflow for human host factors involved in *L. monocytogenes* infection

A high-density siRNA array was placed into a tissue culture dish and actively growing HeLa cells were added. After 48 hours incubation, cells were infected with GFP-*L. monocytogenes* for 3 hours. Extracellular bacteria were removed by gentamicin treatment and infected cells were fixed, permeabilized and stained with 12 μg/ml Hoechst 33324 and 0.1 µM Phalloidin 647. Images were acquired at 10X using an automated confocal imaging system. Data analysis was performed using an off the shelf algorithmic software.

The following measuring parameters were incorporated in the analysis: spot area, total nuclear count, cell shape, cell elliptical form factor, cellular bacteria, bacteria per cell, total bacterial area, mean bacterial area, average bacterial intensity, cell texture index, cell gradient index, cell cellular gradient index and cell cellular laplacian index. These measuring parameters were incorporated to account for any variation in phenotype resulting from the different siRNAs present within the screen array. Data obtained from the analysis were exported then processed by center reduction of each data set, that is, dividing each data set by its mean value. This transformation was to enable array-to-array comparison. The median value for each data point per spot obtained from the transformed data was used in the PCA.

6.3.2 Data processing

Data resulting from high throughput screens can be noisy, and using clustering algorithms may not effectively extract all the classes contained in the data (Ben-Hur and Guyon 2003). One approach shown to increase reliability in data is by increasing the number of screening arrays used when conducting screening assays (Yeung and Ruzzo 2000). Another rate limiting step in screening is the sheer amount of data sets generated during the process. Analyzing such data sets requires an analytical tool capable of reducing the dimensionality of the large variables present in the data set while retaining as much of the variation present in the data as possible (Jolliffe 1986).

In this study, we used a total of 15 replicate screens in conjunction the PCA to identify human host factors implicated in *L. monocytogenes* infection.

Each screen array comprised of all the kinases, phosphatases, ubiquitinylases, and proteases and for quality control p65, XPO1, and non-targeting siRNAs were included in the array. To generate a statistically correct background, data obtained from each screen were transformed by mean center reduction. The median values of each corresponding data set were projected into the PCA for analysis.

To examine the effectiveness of the PCA and the reliability of the arrays, we compared the PCA projections of the control siRNAs obtained from *L. monocytogenes* screen to that of the TNF-α induced p65 nuclear translocation screen previously described. Figure 6.13 below shows 3D PCA projections of the control siRNA generated from both the *L. monocytogenes* and TNF-α induced p65 nuclear translocation screens.

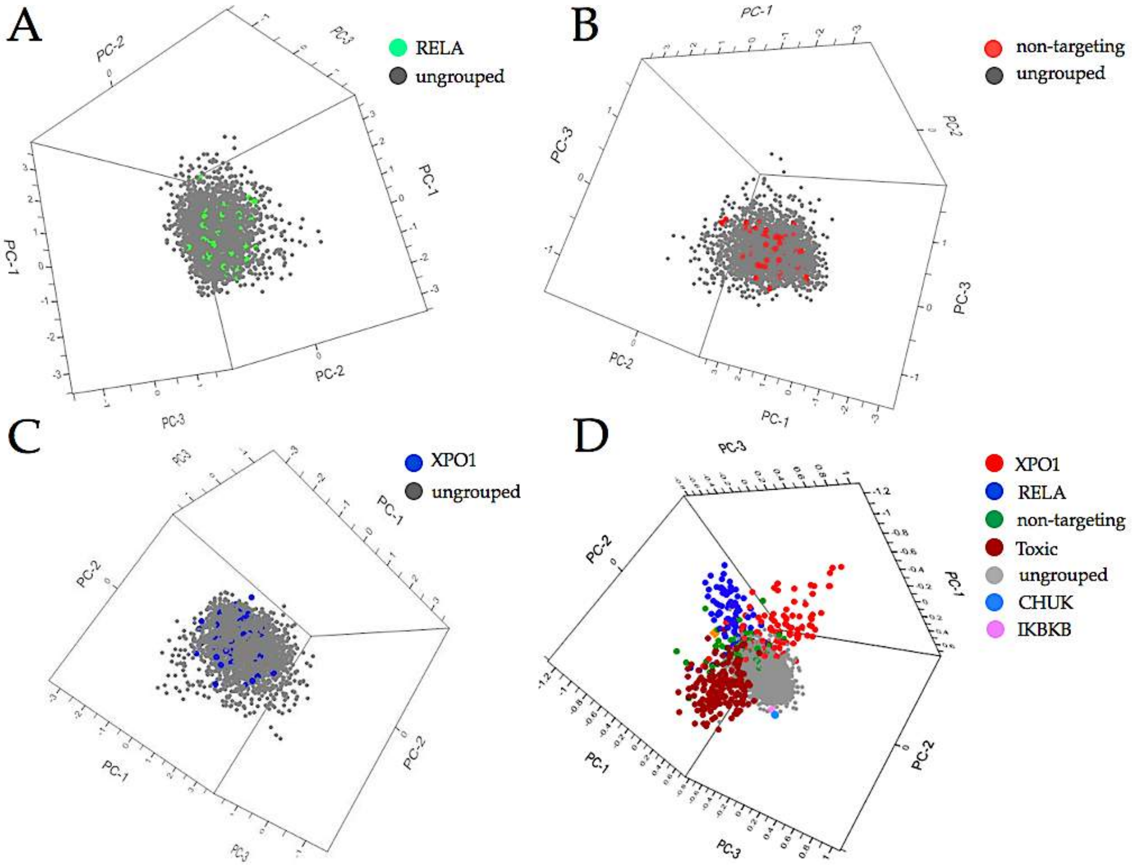


Figure 6.13: 3D projections comparing the distribution of the control siRNAs between the *Listeria* and NF-κB screens

3D projections of the *L. monocytogenes* screen [A, B and C] and the TNF-α screen D, showing the distribution of the control siRNAs from the rest of the screen. As represented in A-C, the control siRNAs are more evenly distributed within the ungrouped siRNAs population while D depict a more reasonable separation of the control siRNAs.

As expected from the 3D projections resulting from the *L.monocytogenes* screen (Figure 6.13 A-C), the control siRNAs distribution was more evenly distributed within the screen while the 3D projection from the TNF-α screen (Figure 6.13 D) showed an excellent separation of the control siRNAs from the body of the screen. The observable differences in siRNA distribution from both screens resulted from the different assay designs used in both screens. For the TNF-α induced p65 nuclear translocation screen, the assay was designed to look for proteins involved in p65 nuclear translocation. Since p65 and the XPO1 are both components of the p65 pathway, it was therefore expected for the PCA to group these control siRNAs into separate classes based on the set of measurements used. However, p65 and XPO1 are not components of the *Listeria* pathway thus there was no statistical influence generated on the p65 and XPO1 axes by the PCA. This then led to an even distribution of the p65 and XPO1 control siRNAs within the body of the screen. This result demonstrated that using the PCA, data with similar patterns to each other could be grouped together into classes and there is no systemic bias toward the internal controls eliminating system error, as was the case with the TNF-α screen described in chapter 5.

6.3.3 "Candidate gene (Hit)" classification

Using the non-linear iterative partial least squares (NIPALS) algorithm and a maximum of seven principal components from the PCA, candidate genes from the screen were identified by ranking the hits according to the absolute values of the Z-scores and selecting those resolved from the body of the screen. Using the Z-score of ±2 Standard deviation (STDV) from the mean of the median values, we classified genes resulting from the screen into four major classes: [i] cells with low bacteria infection, [ii] cells with high bacteria infection, [iii] low cell count and [iv] genes affecting cell actin cytoskeleton (Figure 6.14).

However, since this project was geared at identifying candidate genes involved in propagating or prohibiting the invasion of the pathogen, we therefore used the first two classes for hit identification. Two approaches were implemented for hit selection [i] statistical interpretation of screen data and [ii] image-based triage (using visual observation) of hits resulting from the statistical interpretation hit lists. Image triage compensated for errors in acquisition and analysis where extraordinary phenotypes existed (damaged spots, contamination of the spot and aberrant cell growth). Based on this system of classification, we identified 65 candidate genes, which may potentially interact with the *L. monocytogenes* invasion pathway (Appendix table 2). Since we performed a total of fifteen independent screens, it is important to keep in mind that for a gene to be considered a hit; it should have a 2/3 pass score after visual observation of the different screens. All genes that attained or exceeded the above set pass score threshold were considered to be involved in *L. monocytogenes* infection.

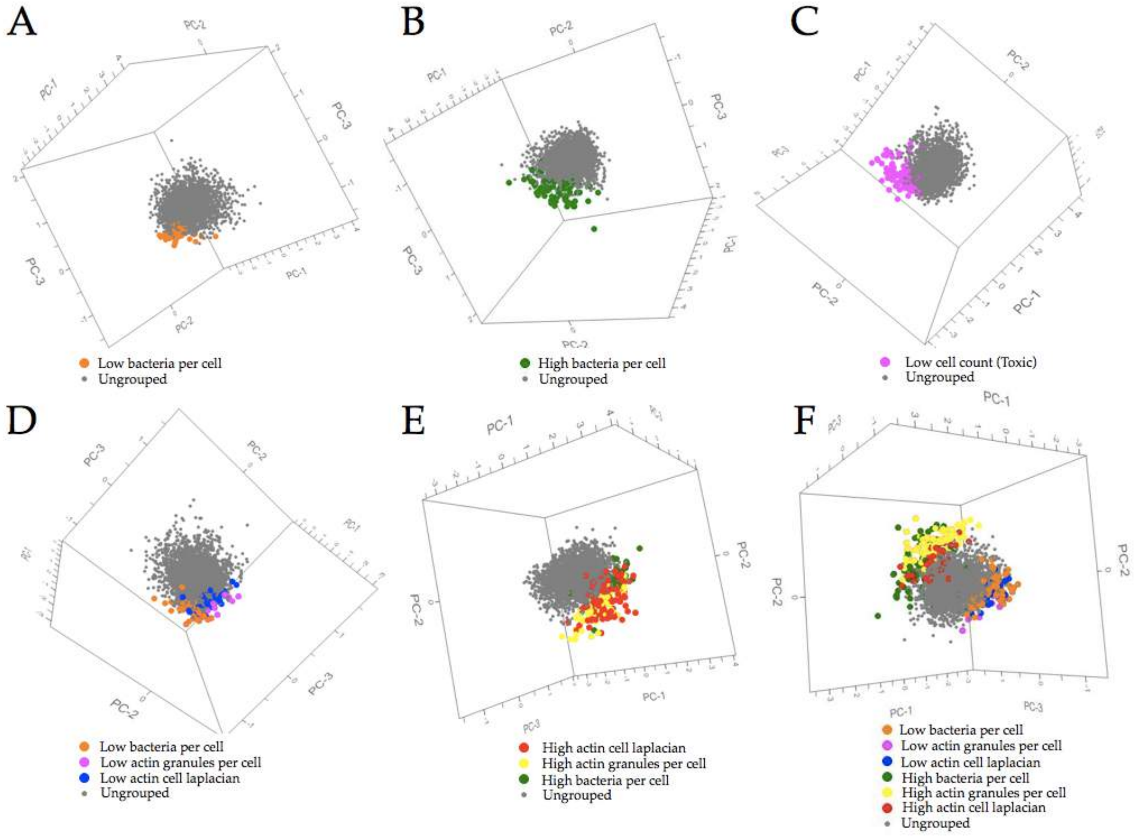


Figure 6.14: 3D projections depicting groups of selected hits

3D projections from the PCA depicting the different classes used in hit selection. [A] siRNAs resulting in low bacteria infection per cell. [B] siRNAs resulting in high bacteria infection per cell. [C] siRNA toxic to the cells (low cell count). [D] and [E] siRNAs affecting cells surface morphology (cell actin cytoskeleton). While [F], is a combination of all the different phenotypic groups.

6.3.4 Gene cluster classification

The phenotypic groups created by the PCA as mentioned ealier were classified in four major classes. Class one was defined as high bacteria infection per cell. Genes within this cluster were those that scored >2 STDV more bacteria/cell compared to the negative control (non-targeting siRNA spot). Class two were those siRNAs spots harboring cells with low bacteria infection. Genes in this class were those that scored a > 2 STDV fewer bacteria/cell compared to the control. Class three was defined as cytotoxic siRNAs. These were spots with > 2 STDV fewer cells/spot compared to the negative control (non-targeting spot). Class four were genes affecting cell actin cytoskeleton. Genes within this class all scored > 2 STDV low cell laplacian compared to the negative control.

The subsequent immunofluorescent images illustrate the different classes used in the hit selection. For the control class, we use images of the distribution of the GFP-*L. monocytogenes* and the cell actin cytoskeleton after HeLa cells were reverse transfected with a non-targeting siRNA spot. As observed from the image (GFP-labeling), the GFP-*L. monocytogenes* (green) were evenly distributed within and outside of the siRNA spots. This result demonstrated that the non-targeting siRNA had no influence on how the cells reacted to the *L. monocytogenes* infection. Also shown in this figure is the cell actin cytoskeleton after infected cells were counter stained for actin with fluorescent Phalloidin 647. These images were used as a baseline control to select positive hits (2/3 pass score) by visual observation after screen data were statistically quantified using PCA.

6.3.4.1 Class one: High infection

The first phenotype we defined was high *Listeria* infection and is defined as cells with > 2 STDV more bacteria/cell compared to the negative control (non-targeting control) (Figure 6.15 A and B). The figure illustrates two genes TLK1 (tousled-like kinase 1) and LATS2 (large tumor suppressor, homolog 2). TLK1 is known to regulate processes involved in chromatin assembly while LATS2 is involved in protein phosphorylation. As depicted in the figure, the first panel comprises of the non-targeting siRNA spot showing a > 2 STDV fewer bacteria/per cell compared to panels 2 and 3 where we see examples of two siRNA spots whose silencing visually enhanced bacterial invasion. Column 1 shows the GFP-bacteria distribution, column 2 the nuclei, column 3 the siRNA spots and column 4 overlay images of all three channels.

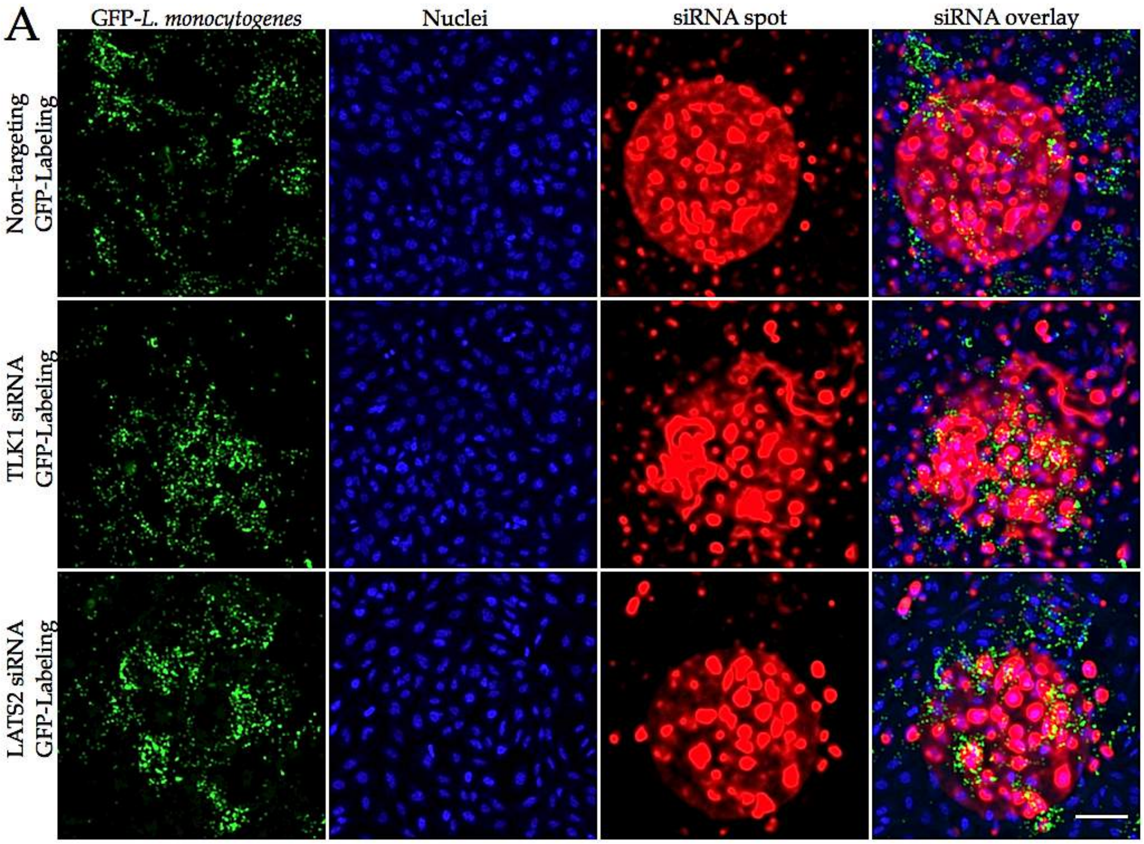


Figure 6.15 A: GFP-Labeling: Silencing of TLK1 and LATS2 directed siRNA printed spots

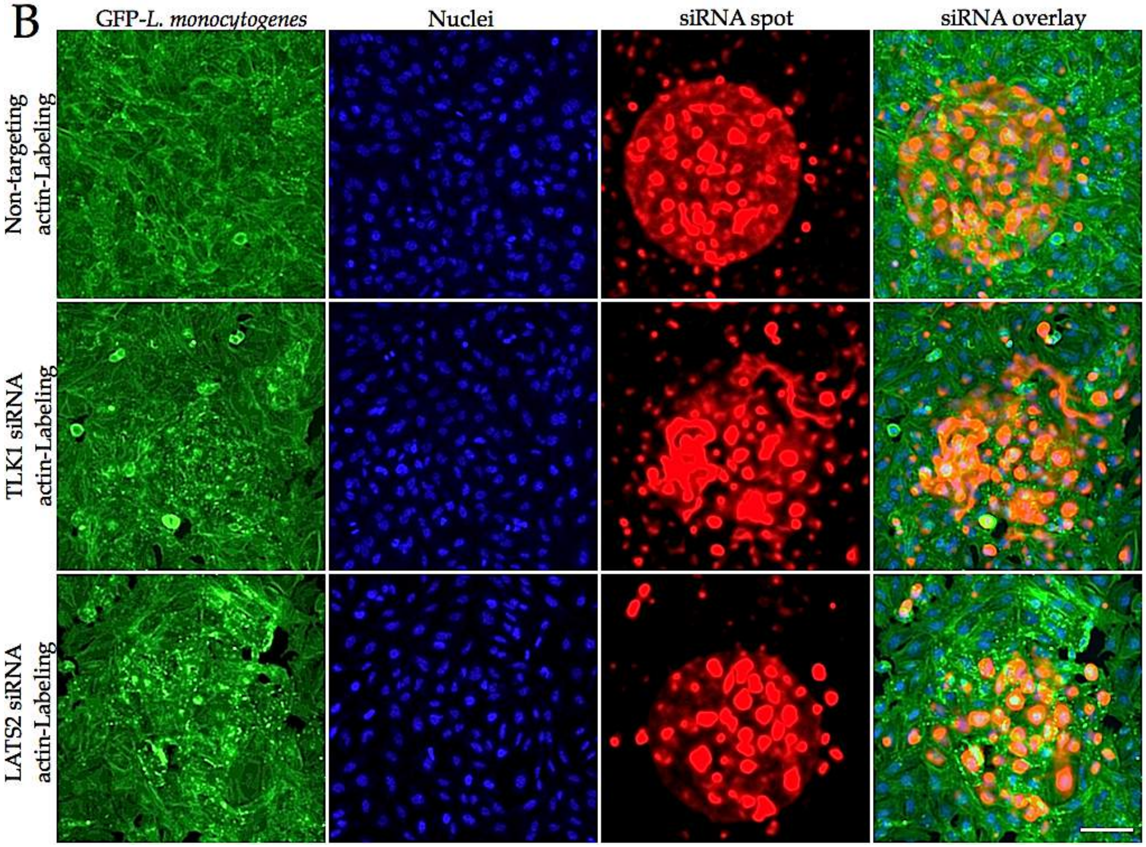


Figure 6.15 B: Actin-Labeling: Immunofluorescent images of the cell actin cytoskeleton following treatment of HeLa cells with TLK1 and LATS2 siRNAs

Immunofluorescent images of HeLa cells exposed to GFP-L. monocytogenes. HeLa cells were reverse transfected with human siRNAs for 48 hours. Transfected cells were exposed to the bacteria (MOI: 25) for 3 hours and later treated with gentamicin for 1 hour. Post infection, cells were fixed, permeabilized and the cell nuclei and actin cytoskeleton were counter-stained with 12 ng/ml Hoechst 33473 and 0.1 μM Phalloidin 647 respectively. Images were acquired at 10X using a standard confocal imaging system. [A] GFP-labeled images demonstrating high bacteria/cell (TLK1 and LATS2) compared to the non-targeting control (panel 1). Column 1 of each image shows the GFP-bacteria, column 2 the nuclear images, column 3 the siRNA spots while the last column overlays images of the respective channels. [B] The corresponding actin-labeled images of the non-targeting control, TLK1 and LATS2 siRNAs. Column 1 of each image shows actin-stained images; column 2 the nuclear images; column 3 the siRNA spots while the last column represents the overlay images. Scale bar 100 μm.

6.3.4.2 Class two: Low infection

The second phenotype we defined was low *Listeria* infection and is characterized as cells with > 2 STDV fewer bacteria/cell compared to the negative control (non-targeting control) (Figure 6.16 A and B). Here we demonstrate two siRNA spots PPP6C (protein phosphatase 6, catalytic subunit) and PON1 (paraoxonase 1) whose silencing visually hindered bacteria entering.

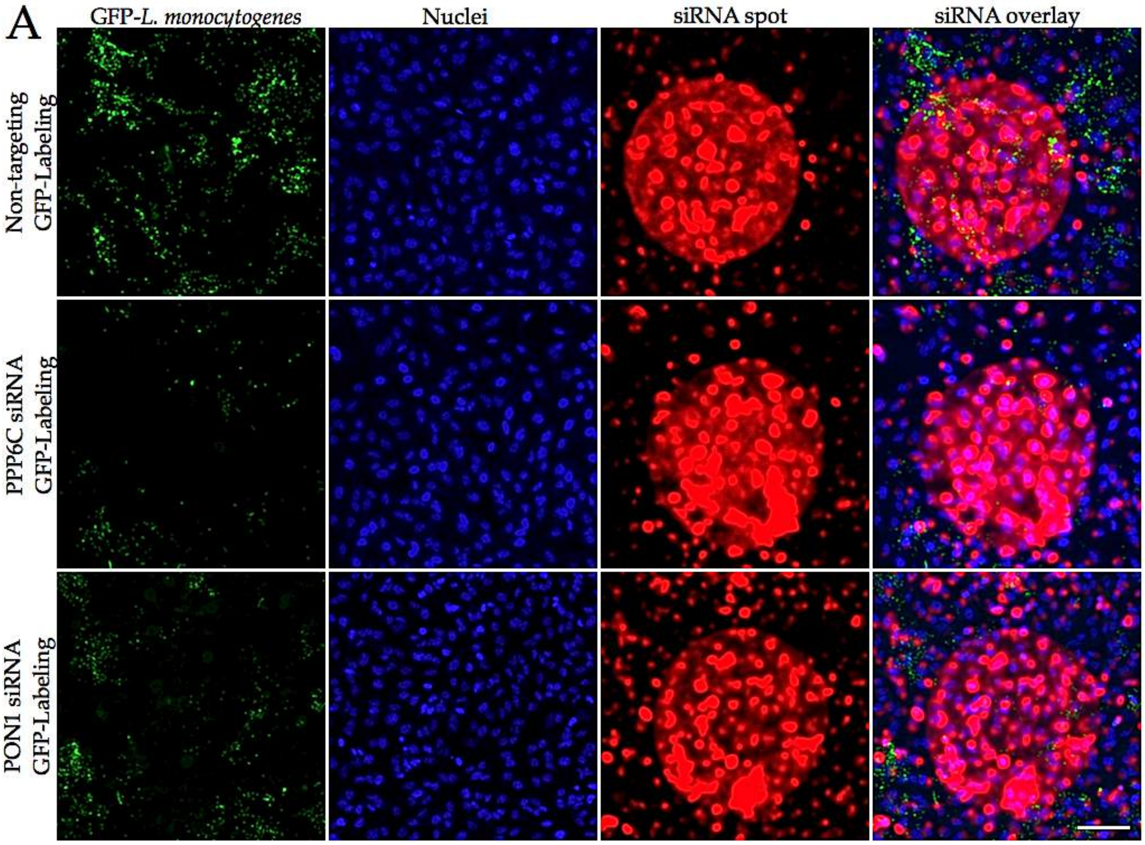


Figure 6.16 A: GFP-Labeling: Silencing of PPP6C and PON1 directed siRNAs printed spots

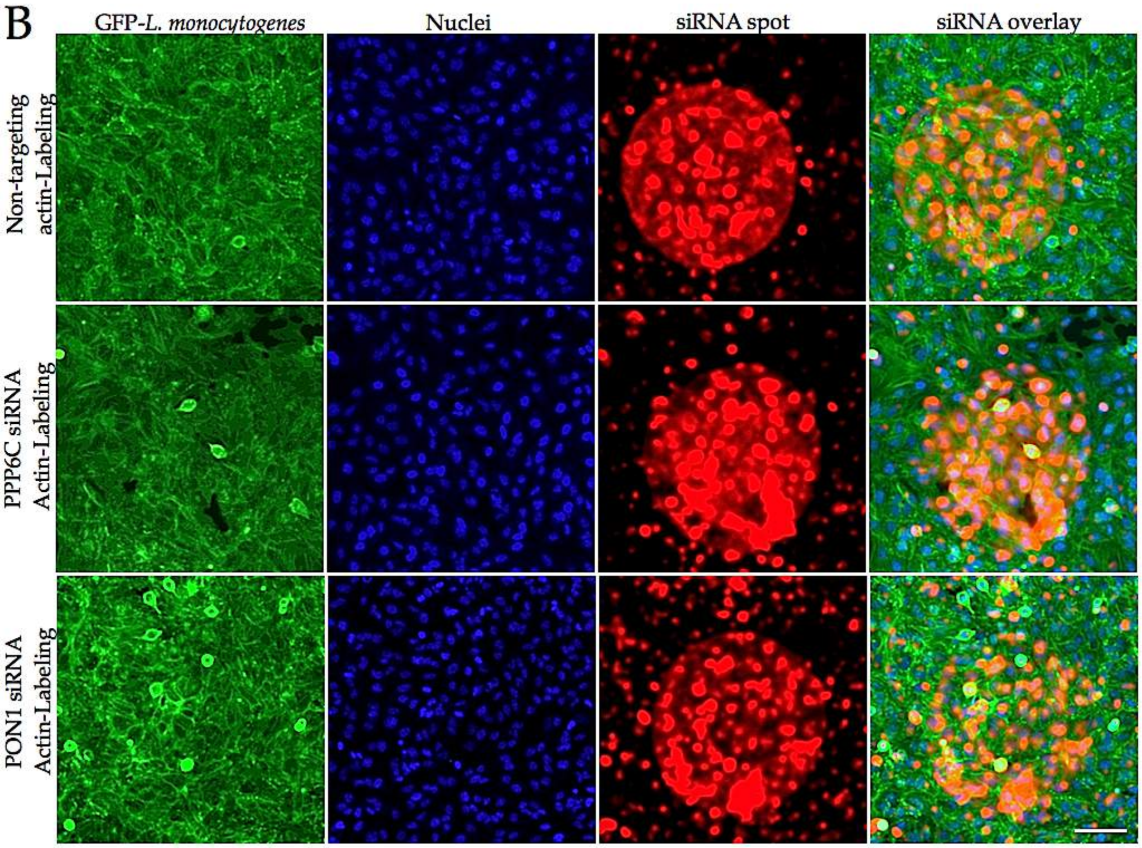


Figure 6.16 B: Actin-Labeling: Immunofluorescent images showing the cell actin cytoskeleton following treatment of HeLa cells with PPP6C and PON1 siRNAs

Immunofluorescent images of HeLa cells exposed to GFP- L. monocytogenes. HeLa cells were reverse transfected with human siRNAs for 48 hours. Transfected cells were exposed to the bacteria (MOI: 25) for 3 hours and later treated with gentamicin for 1 hour. Post infection, cells were fixed, permeabilized and the cells' nuclei counter-stained with 12 ng/ml Hoechst 33473 and 0.1 μM Phalloidin 647 respectively. Images were acquired at 10X using a standard confocal imaging system. [A] GFP-labeled images demonstrating fewer bacteria/cell (PPP6C and PON1) compared to the non-targeting control (panel 1). Column 1 of each image shows the GFP-bacteria, column 2 the nuclear images, column 3 the siRNA spots while the last column overlays images of the respective channels. [B] Corresponding actin-labeled images of the non-targeting control, PPP6C and PON1 siRNAs. Column 1 of each image shows actin-stained images; column 2 the nuclear images; column 3 the siRNA spots while the last column represents the overlay images. Scale bar 100 μm.

6.3.4.3 Class three: siRNA induced cell death

The third class of phenotype was siRNAs that induced host cell death. We scored this class as > 2 STDV fewer HeLa cells/siRNA printed spots compared to the non-targeting control. The gene selected from this class as an example is the potassium voltage-gated channel (KCNB1), which is a shab-related subfamily, member 1. Members of this family have been shown to play critical roles in cellular signaling processes regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume regulation (Shieh, Coghlan et al. 2000). As illustrated by some research work, changes in K+ channel function have been linked with apoptosis, oncogenesis and cardiac hypertrophy (Shieh, Coghlan et al. 2000). Here we demonstrated that there was a significant (> 30%, $p < 0.0002$) reduction in cell number compared to the negative non-targeting control after cells were treated with the KCNB1 siRNA (Figure 6.17 B).

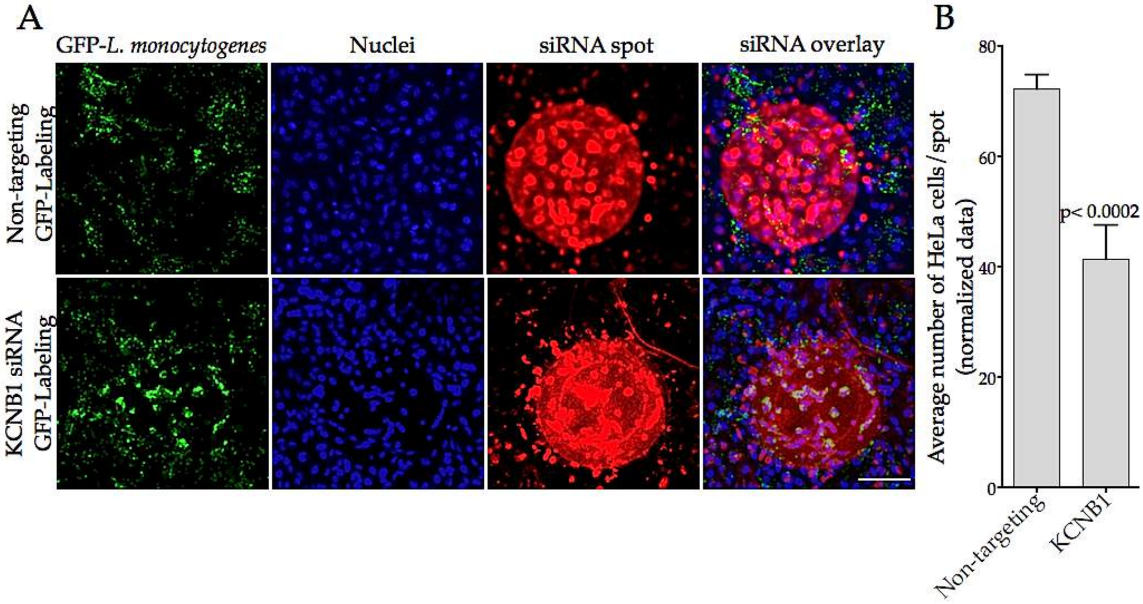


Figure 6.17 A and B: GFP-Labeling: Silencing of KCNB1 directed siRNA printed spot

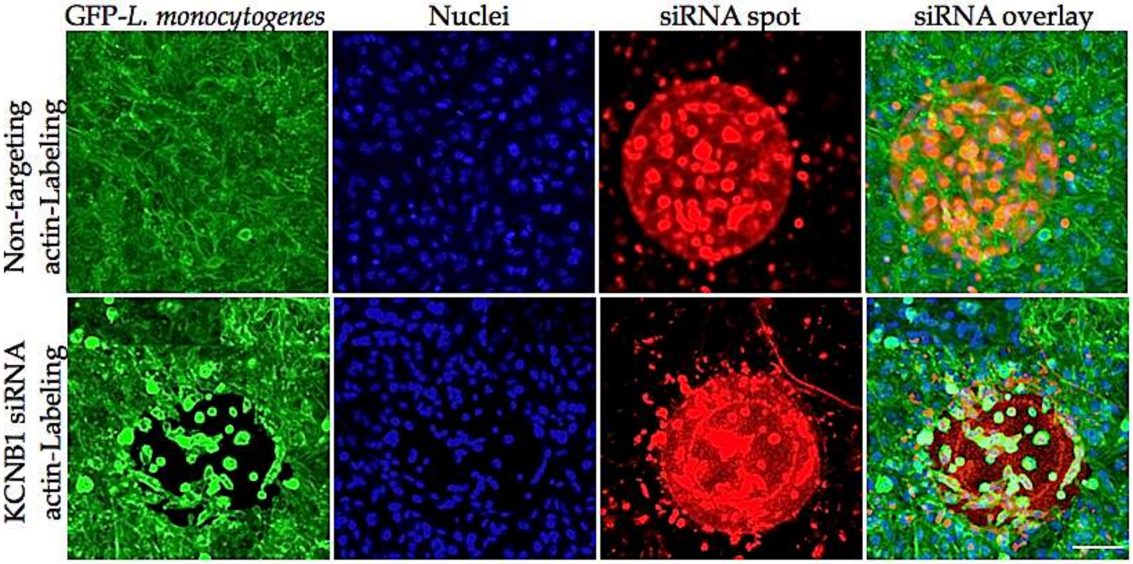


Figure 6.17 C: Actin-labeling: Phenotype induced by KCNB1 directed siRNA printed spot.

HeLa cells were reverse transfected with human siRNAs for 48 hours. Transfected cells were exposed to the bacteria (MOI: 25) for 3 hours and later treated with gentamicin for 1 hour. Post infection, cells were fixed, permeabilized and cells' nuclei and actin cytoskeleton counter-stained with 12 ng/ml Hoechst 33473 and 0.1 μM Phalloidin 647 respectively. Images were acquired at 10X using a standard confocal microscope. [A] GFP-labeled images depicting the distribution of the bacteria on the KCNB1 siRNA spot. The upper panel represents the non-targeting images while the lower panel shows the KCNB1 images. Column 1 of the image shows the GFP-bacteria. Column 2, shows the cell nuclei images. Column 3, the siRNA printed spots and the last column shows the overlay images of the three channels. [B] Unpaired student T-test showing the relative significant in cell number between the non-targeting control and the KCNB1 siRNAs ($p < 0.0002$). n= 180 for non-targeting and n=10 for KCNB1. [C] Actin labeled images demonstrating the effect of KCNB1 on the cell actin network. Upper panel represents the non-targeting control siRNA while the lower panel represent the KCNB1 siRNA. Scale bar 100 μm.

6.3.4.4 Class four: Genes affecting the cell actin cytoskeleton

One probable phenotype are genes whose silencing disrupts the cell actin cytoskeleton and as a result affects infection (Figure 6.18). Interestingly, one of the genes (ras homolog family member A; RHOA) within this cloud belongs to the GTPase family. Members within this family are generally known to regulate the actin cytoskeleton (Sit and Manser 2011; Blajecka, Marinov et al. 2012). The phenotypic expression resulting from the RHOA siRNA spot may have resulted in membrane actin disintegration, which in turn compromised cell viability. This may have played a key role in slowing down the bacteria proliferation within the cell cytoplasm as observed with the ditribution of bacteria within the RHOA siRNA printed spot (Figure 6.18 A).

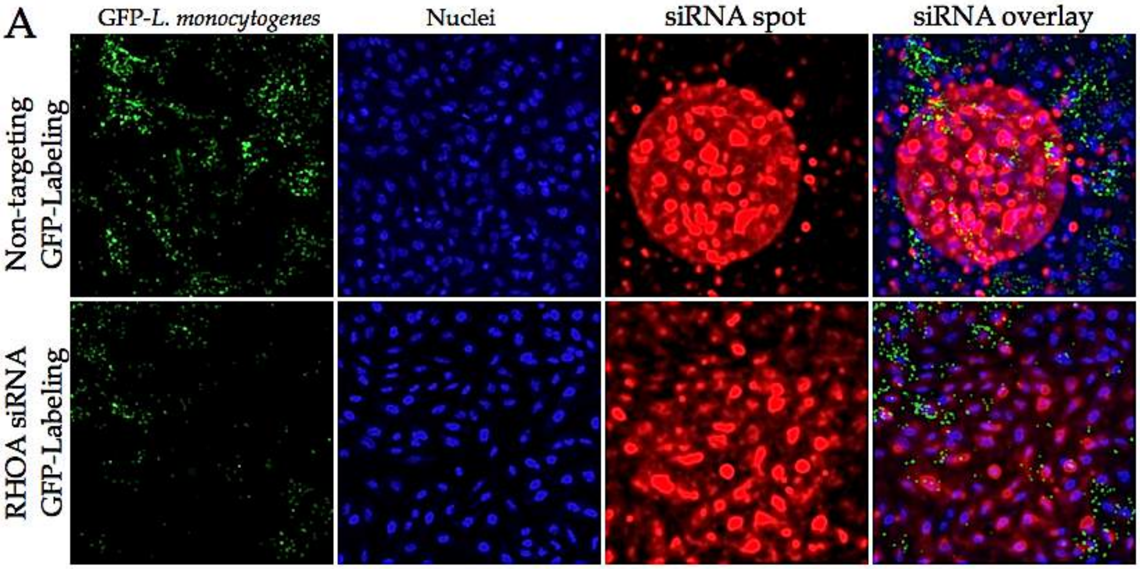


Figure 6.18 A: GFP-Labeling: Silencing of RHOA directed siRNA printed spot.

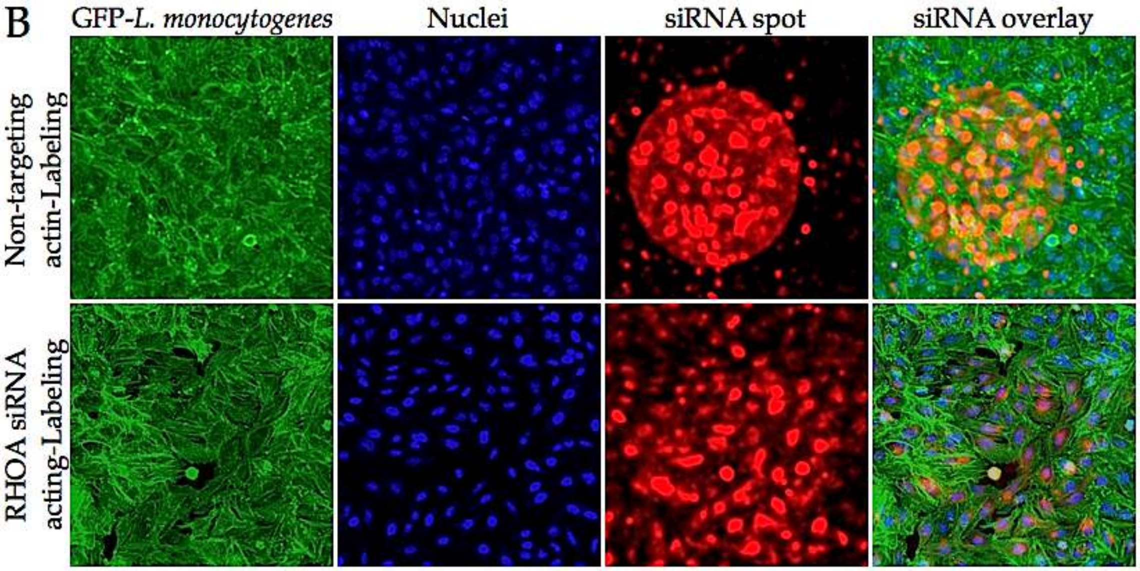


Figure 6.18 B: Actin-Labeling: Silencing of RHOA directed siRNA printed spot.

HeLa cells were reverse transfected with human siRNAs for 48 hours. Transfected cells were exposed to the bacteria (MOI: 25) for 3 hours and later treated with gentamicin for 1 hour. Post infection, cells were fixed, permeabilized and cells' nuclei and actin cytoskeleton were counter-stained with 12 ng/ml Hoechst 33473 and 0.1 μM Phalloidin 647 respectively. Images were acquired at 10X using a standard confocal imaging system. [A] The upper panel shows the distribution of bacteria on and round a non-targeting siRNA spot while the lower panel shows the distribution of bacteria on and around the RHOA siRNA spot. Column 1 of each siRNA image shows the GFP expressing bacteria. Column 2 shows the nuclear images. Column 3 the siRNA spots while the last column overlay images of the respective three channels. [B] Shows the corresponding actin stained images for the RHOA siRNA printed spot. HeLa cells within the RHOA siRNA spot look aberrant that is when compared to the HeLa cells localized within the non-targeting siRNA control spot. Scale bar 100 μm.

6.3.5 Identification of essential protein interactions using Pathway Studio software

To obtain an initial understanding on the possible protein interactions taking place, the human host factors (proteins) identified from the screen were employed to construct a protein interaction network using the Pathway Studio. Pathway Studio is a pathway analysis tool which harvests information from deposited literature in PubMed and other public sources and provides us with an understanding of the possible interactions taking place between proteins and other small molecules both at the molecular and cellular level (Nikitin, Egorov et al. 2003; Hosseinpour, HajiHoseini et al. 2012). This analytical tool was then used as a scaffold to establish comprehensive information on the possible interactions existing between the potential genes identified. Of the 65 proteins projected into the software for protein-protein interaction, 23.07 % of the genes had direct interaction with one or more members from the cohort (Figure 6.19). Table 6.1 provides the protein-protein interactions between the genes as well as the PMID as reference where these genes have been shown to interact with each other.

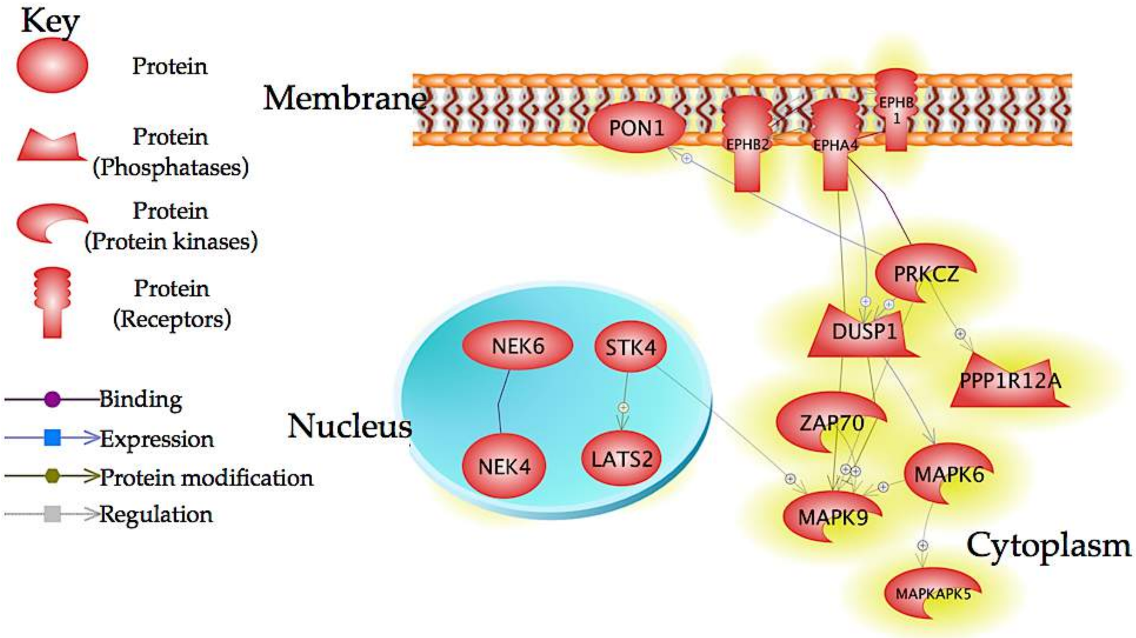


Figure 6.19: Identification of essential proteins interactions using Pathway Studio.

A cohort comprising of 65 genes was projected into the Pathway Studio for possible protein-protein interaction. As shown, 23.07% of these genes interacted directly with one or more members of the group. Genes were localized in the nucleus and cytoplasm as well as in the cell membrane.

Table 6.1: Primary hits protein-protein interaction and their respective PMID

Genes	Interactions	Reference, PMID
NEK4---NEK6	Binding	21605211
STK4---LATS2	ProtModification	18640976, 17442573, 19878866, 22898666, 22898666, 22898666
STK4---MAPK9	Regulation	10699464
MAPK6---MAPKAPK5	Regulation	17947239, 16971392,
MAPK6---MAPK9	Regulation	20409730
DUSP1---MAPK6	Expression	19789277
EPHA4---MAPK9	ProtModification	10617634, 9334169, 9545354, 9553058, 10575005
ZAP70---MAPK9	Regulation	10521707
PRKCZ---MAPK9	Regulation (positive)	15313379, 15313379
DUSP1----MAPK9	ProtModification (negative)	8626452, 20708668, 20708668, 15855046, 21350193, 11139479, 21255694, 15059639, 20708668

PRKCZ---PPP1R12A	Regulation (positive)	19011165
PRKCZ---PON1	Expression (positive)	15380450, 20092859
EPHA4---DUSP1	Expression (positive)	9280068
EPHA4---EPHB1	Binding	12084815
EPHA4---MPHB2	ProtModification	11754835
EPHB2---EPHB1	Regulation	11136979, 10600686
PRKCZ---EPHA4	Binding	10962008, 15164763
EPHB2---EPHA4	Binding	19875447, 20633976
EPHB1---EPHB2	Regulation	9499402, 19303107, 12971890, 10600686

6.3.6 Conclusion

The objective of this study was to identify human host factors (proteins) implicated in *L. monocytogenes* infection. Using arrays of 3150 siRNA spots comprising of all human kinases, phosphatases, ubiquitinylases, and proteases we screened for genes whose silencing may either hinder or enhance bacterial invasion of cells. A total of 15 replicate screens were performed. Data were centrally reduced using the mean value of the negative control (non-targeting siRNA). Using a multi-dimensional analysis (PCA) we separated the genes into four different classes based on their similarities in expressed cellular phenotype:

- genes whose silencing led to an increase in intracellular bacteria,
- genes whose silencing led to a decrease in intracellular bacteria,
- siRNAs that may result in cellular toxicity or whose gene silencing led to cell death,
- genes whose silencing may affect the cell actin cytoskeleton.

Implementing an image based triage approach we were able to identify 65 candidate genes, which may potentially be involved in *L. monocytogenes* infection. Using the Pathway Studio, we also demonstrated that 23.07 % of the genes had a direct interaction with one or more other genes from the cohort of gene.

6.4 Secondary screen validation

To eliminate internal errors that may have resulted from the high-density array screening, a new set of siRNA pool targeting the 65 genes identified in the primary screen were validated against the existing values of the array. The secondary screen assay was validated in 96-well plates, a robust and well-established approach regularly used for screening. As was previously established for the high-density microarray screen, this 96-well format also had to be validated before proceeding with the screening. We performed a total of six replicate screens for this assay and the screen gave rise to fifteen strongest hits some of which have already been implicated in *L. monocytogenes* infection.

6.4.1 Quantification of silencing efficiency by forward transfection

To confirm the 65 candidate genes identified during primary screening, we performed a secondary screen validation assay in 96-well plates by forward transfecting HeLa cells with the individual siRNA pool. The process of forward transfection involves exposing the cells to nucleic acids after seeding the cells. This validation was performed to ensure that phenotypic gene silencing resulting from forward transfection was the same as that obtained from the microarray technology in which reverse transfection was used. It is important to note that if inappropriate siRNA concentrations as well as inadequate time of transfection are used, there will be an increased risk of misidentifying off-target genes as candidate genes during data analysis after the secondary screening. The goal here was to establish an siRNA concentration and time of transfection that will give similar silencing efficiency as in the microarray screen where spotted nucleic acidsw were reversibly taken up by the cells.

To validate this format, HeLa cells were seeded in 96-well plates 6 hours prior to siRNA treatment. Non-adherent cells were aspirated and 100 μl of Opti-mem media pipetted into each well. 20 μl of lipofectamine transfection mixture containing a titration of the p65 siRNA (0 to 50 nM) was added in to the respective wells. Transfection was allowed to proceed for 48 hours at 37^0C in a humidified atmosphere. Post transfection, cells were fixed, permeabilized, and p65 expression and cell nuclei were immunofluorescently labeled using rabbit polyclonal antibody (1:1000) and 1μM DRAQ5. Imaging was performed using a confocal imaging system and cytoplasmic p65 expression was measured by quantifying the intensity of the p65 staining.

As demonstrated by the immunofluorescent image (Figure 6.20 A and B), there was a significant increase in cytoplasmic to nucleus p65 expression after HeLa cells were exposed to the non-targeting siRNA control. On the other hand, there was an even distribution of p65 expression between the cytoplasm and the nucleus after Hela cells were treated with the XPO1 siRNA. The latter observation was as a result of p65 being unable to relocate back into the nucleus since the nuclear export factor responsible for the process was silenced. Another important observation was the decrease in cell number as the concentration of the p65 siRNA was increased (Figure 6.20 C). This result depicts that at high concentrations, the siRNAs are toxic to the cells.

Based on this findings, the ideal assay concentration for the siRNA and time of transfection was set at 20 nM and 48 hours respectively to ensure the same level of silencing as on the microarray screening (Figure 6.20 B and C).

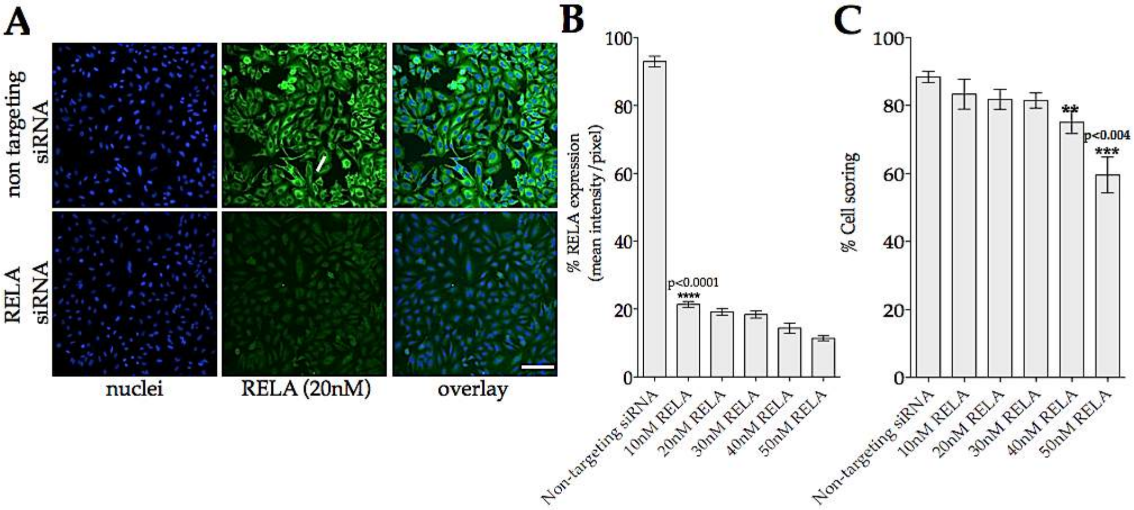


Figure 6.20: Validation of the secondary sceening

Forward transfection of HeLa cells with a titration of p65 siRNA. [A] Non-targeting and RELA siRNA duplexes [10-50 nM] were forward transfected into HeLa cells for 48 hours and RELA (p65) expression imaged and [B] Quantified as % RELA expression relative to control non-targeting siRNA [RELA intensity/pixel/cell]. [C] Student T-test performed on the number of viable cells following siRNAs treatment expressed as % cell scoring: (10 nM: $p=0.3$, 20 nM: $p=0.8$, 30 nM: $p=0.03$, 40 nM; $p<0.005$, 50 nM: $p<0.0004$). Scale bar 15 µm.

To ensure the effectiveness of the chosen concentration, two different control siRNAs (INCENP and XPO1) were tested (Figure 6.21). Results obtained depicted both a poly-nucleate phenotype following silencing of the cytokinesis regulator INCENP (middle panel) and a reduced nuclear to cytoplasmic p65 translocation (lower panel) after the silencing of the nuclear export gene XPO1. These assay conditions were later used to further validate the candidate genes obtained from the primary screen.

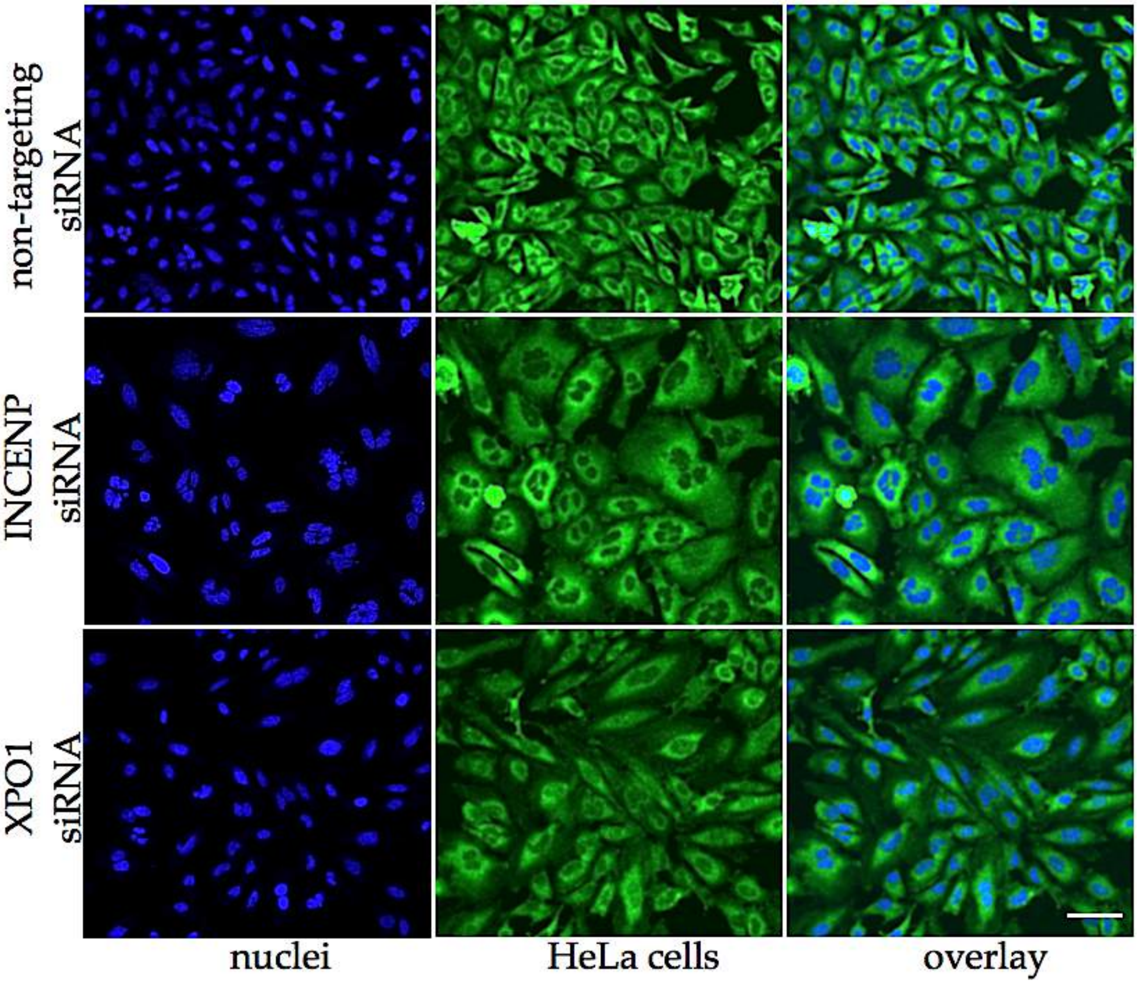


Figure 6.21: Silencing of INCENP and XPO1 directed siRNAs.

HeLa cells were cultured in a 96-well plate for 6 hours prior siRNA treatment. Cells were forward transfected with non-targeting, INCENP and XPO1 siRNAs for 48 hours at 37^{0}C. Transfected cells were treated for immunofluorescence and imaged at 10X using a standard confocal microscope. The upper panel shows the immunofluorescent image of cells treated with the non-targeting siRNA, the middle panel shows a poly-nucleate phenotype following treatment of the cells with INCENP siRNA while the lower panel shows a reduced p65 cytoplasmic expression following treatment of cells with XPO1. n = 6, Scale 20μm.

6.4.2 Secondary screen validation for genes involved in *L. monocytogenes* infection

Adopting the experimental approach described in section 6.4.1, a new batch of the 65 candidate genes identified from the primary screen was validated against *L. monocytogenes* infection. Here, HeLa cells were seeded in 96-well plates 6 hours prior to siRNA treatment. Non-adherent cells were aspirated from each well, washed and 100 μl Opti-mem media added to each well. Twenty micro-liters of the lipofectamine tranfection mixture containing the different siRNAs at concentration of 20 nM (final) were added in to the respective wells. Cells were cultured for 48 hours at 37^0C in a humidified atmosphere. Transfected cells were exposed to GFP-bacteria for 3 hours and subsequently treated with gentamicin to exclude extracellular bacteria. Post infection, cells were fixed, permeabilized, and the cell actin cytoskeleton and cell nuclei were counter-stained with 0.1 µM Phalloidin 647 and 12 μg/ml Hoechst 33324 stains respectively. A total of four images were captured per well using an automated confocal imaging system. It should be noted that for adequate silencing quantification, p65, XPO1, INCENP and non-targeting siRNAs were introduced as internal controls.

Figure 6.22 shows a layout of the screen plate and the positions of the different siRNAs used during screening. Briefly, the white circles represent the potential siRNAs from the primary screen whose involvement in human cells either enhanced or impaired bacteria entry. The colour coded circles represent the control siRNAs; the blue circles represent wells containing cells that were neither siRNA treated or infected with the GFP-bacteria, the green circles comprise of GFP-bacteria infected cells but no siRNA treatment, while the pink, yellow, red and brown colour coded circles represent the internal p65, non-targeting, XPO1 and INCENP control siRNAs respectively.

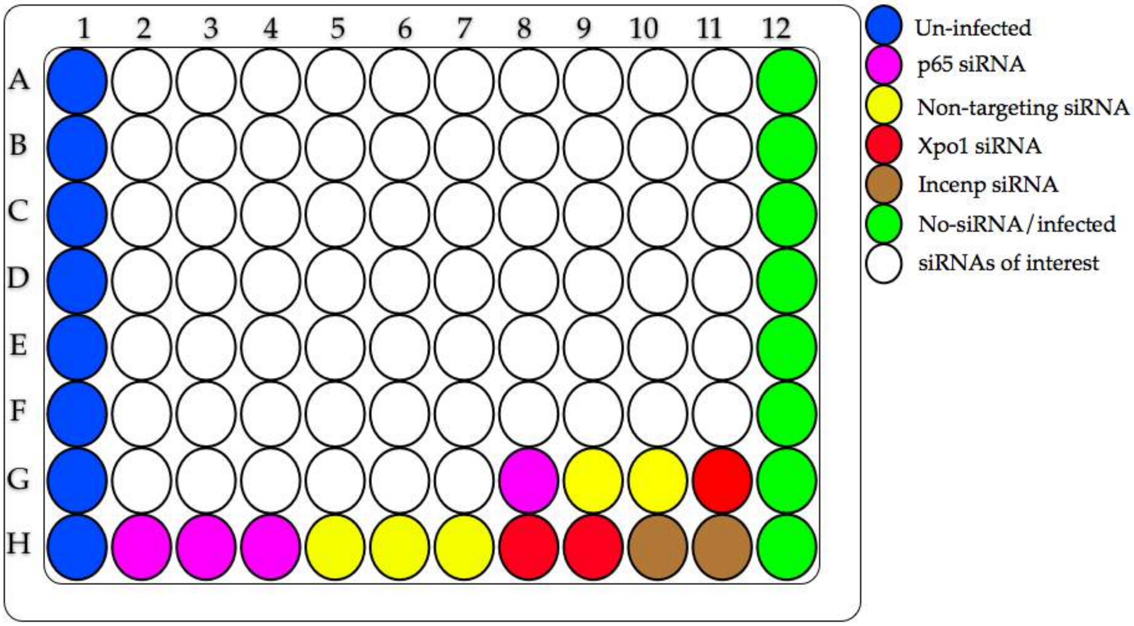


Figure 6.22: Secondary screen-plate layout.

Screen plate depicting the orientation of the different classes of siRNAs used during screen. White colour coded circles represent the siRNAs obtained from the primary screen while pink (p65), yellow (non-targeting), red (XPO1) and brown (INCENP) circles present the Internal controls which were added to account for variations in phenotype resulting during screening. The blue circles represent uninfected and no siRNA treated wells while the green circles are siRNA treated but no bacterial infection.

Images were analyzed by quantitative confocal microscopy and resulting data from the screen were computed using a computational based approach. A total of 6 replicated screens were performed and the number of bacteria in each well was expressed as the average bacteria per cell. For data interpretation, data normalization was performed by averaging the total number of bacteria in each well from the 6 replicate screens. The targeted siRNAs were grouped into different clusters depending on the variation in the average number of bacteria expressed in each well relative to the number of bacteria in the non-targeting siRNA treated wells. Data are shown as mean with standard deviation (Figure 6.23). The dotted and the solid black lines represent 1, 2, 3 and 4 STDEV from the control.

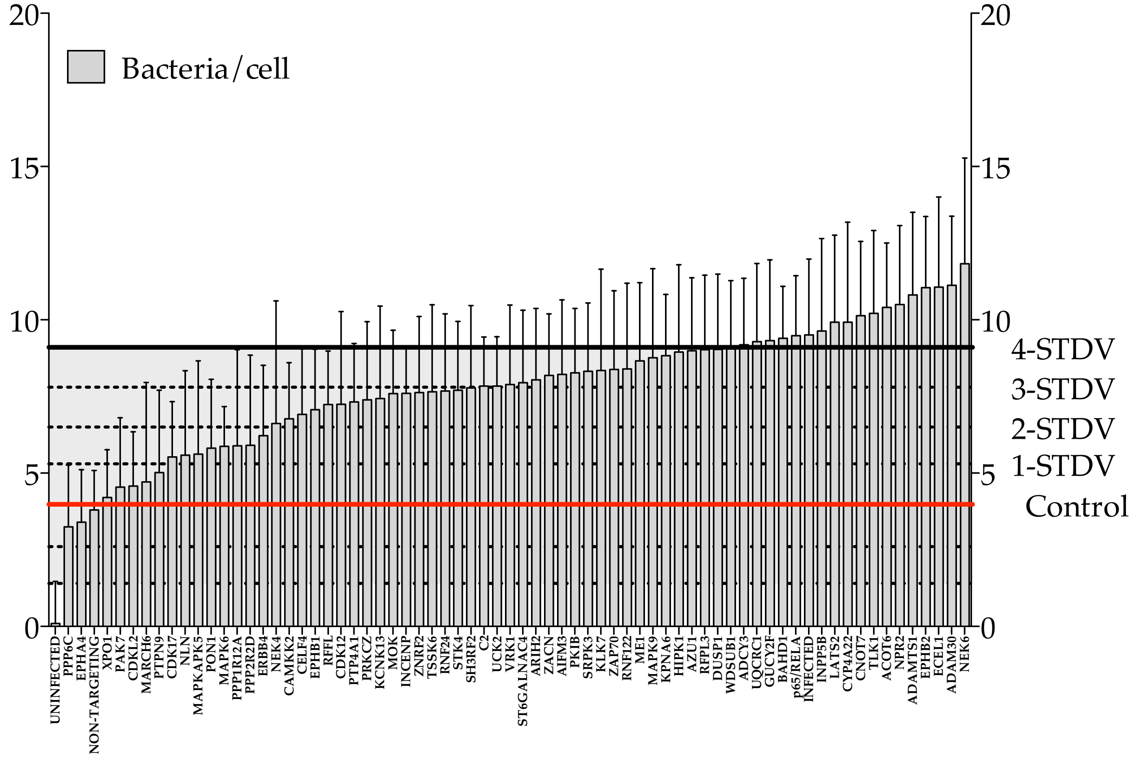


Figure 6.23: Statistical quantification of secondary screen data.

Targeted siRNAs were grouped into different clusters depending on the variation in the average number of bacteria expressed in each well relative to the number of bacteria in the non-targeting siRNA treated wells. Data are shown as mean with standard deviation. Dotted and the solid black lines represent 1, 2, 3 and 4 STDV from the non-targeting control. Selected hits were all from the cluster with > 4 STDV from the control, n = 24.

Figure 6.24 shows an extrapolated image view resulting from the screen after HeLa cells was treated with NEK6 siRNA. This result depicted that silencing the NEK6 gene will result in high bacteria internalization.

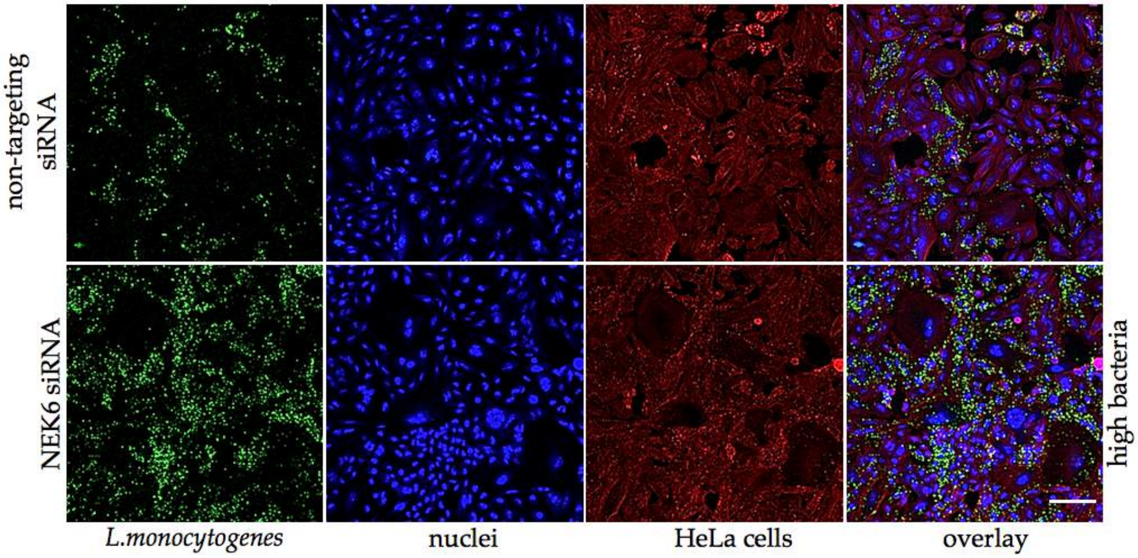


Figure 6.24: NEK6 siRNA induced silencing.

The siRNA transfected HeLa cells were exposed to GFP-*L.monocytogenes* for 3 hours. Cells were fixed, permeabilized and cell nuclei and actin cytoskeleton counter-stained with 0.1 µM Phalloidin 647 and 12 *µ*g/ml Hoechst 33324 stains respectively. Imaging was performed at 10X using a confocal imaging system with four fields acquired per well. The upper panel represents cells treated with non-targeting siRNA control while the lower panel represents cells treated with an on-target (NEK6) siRNA. The intracellular GFP-bacteria, cell nuclei and cells actin cytoskeleton are represented as green granules, blue and red colours respectively. Scale bar 25 µm. n = 6.

For absolute certainty, we selected as the strongest hits those siRNAs whose resulting effect gave a STDEV greater than 4 (Table 6.2). To examine whether there is a significant difference between the average number of bacteria from the selected hits to those of the control, a student T-test was performed (Figure 6.25). The results obtained depicted that the average number of bacteria present in the wells following treatment of the cells with the respective siRNAs, were significantly different ($p< 0.0001$) from those of the control.

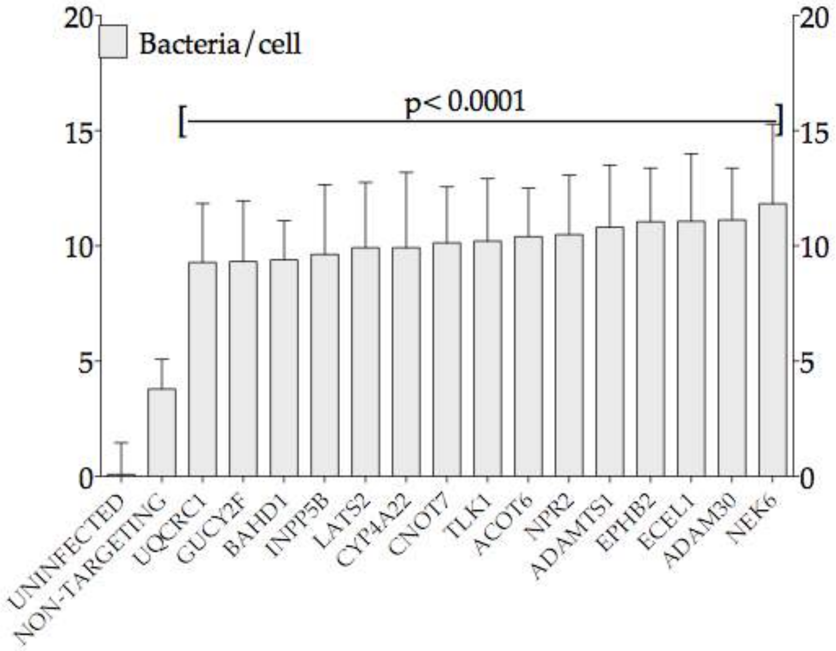


Figure 6.25: Student T-test analysis of secondary hits.

The student T-test was performed to determine the statistical difference in average bacteria per cell of the 15 strongest genes selected relative to the non-targeting control. As shown by the statistical test, the average bacteria per cell of the hits were significantly different from those of the non-targeting. *p*<0.0001. n = 24.

Table 6.2: Strongest hits resulting from the secondary screen

Gene Names	Mean bacteria/cell	STDEV	N
UQCRC1	9.28606575	2.55296335130693	48
GUCY2F	9.325791	2.62954064918008	24
BAHD1	9.3984545	1.69299501335469	24
INPP5B	9.6329645	3.01775355487085	24
LATS2	9.91962	2.83910848443451	24
CYP4A22	9.927026	3.25691156615411	24
CNOT7	10.139127	2.42269430614165	24
TLK1	10.209128	2.70786092824864	24
ACOT6	10.408104	2.09612650741346	24
NPR2	10.495962	2.57717763722927	24
ADAMTS1	10.811063	2.6994244435332	24
EPHB2	11.052552	2.32057009391403	24
ECEL1	11.06599	2.93840021939801	24
ADAM30	11.1301055	2.2458633760299	24
NEK6	11.8288085	3.45229379157522	24

Hits were selected from a single cluster (> 4 STDEV from the negative control [non-targeting siRNA]) following statistical quantification of the mean bacteria/cell. n = 24 with the exception of UQCRC1.

6.4.3 Functional characterization of secondary hits

6.4.3.1 BAHD1 role in *L. monocytogenes* infection

Intracellular bacteria use their effector proteins (virulence factors) to hijack host cellular functions to proliferate within the host cell. On the other hand, host cells respond to the latter process by activating the appropriate defense transcriptional programs (Lebreton, Lakisic et al. 2011). Recently, it has been shown that the bacterial virulence factor, LntA targets the host chromatin repressor, the bromo adjacent homology domain containing 1 protein (BAHD1) in the nucleus to activate interferon stimulated genes. This interaction in turn mediates immune responses to control bacterial colonization of the host cell (Cossart 2011; Lebreton, Lakisic et al. 2011; Lebreton, Cossart et al. 2012).

The BAHD1 gene is a host-silencing factor that orchestrates heterochromatin assembly at specific genes (Lebreton, Lakisic et al. 2011), which the bacterial pathogen uses to its advantage to control host gene expression to propagate its intracellular survival. Figure 6.26 shows the BAHD1/LntA$^{+/-}$ mediated responses.

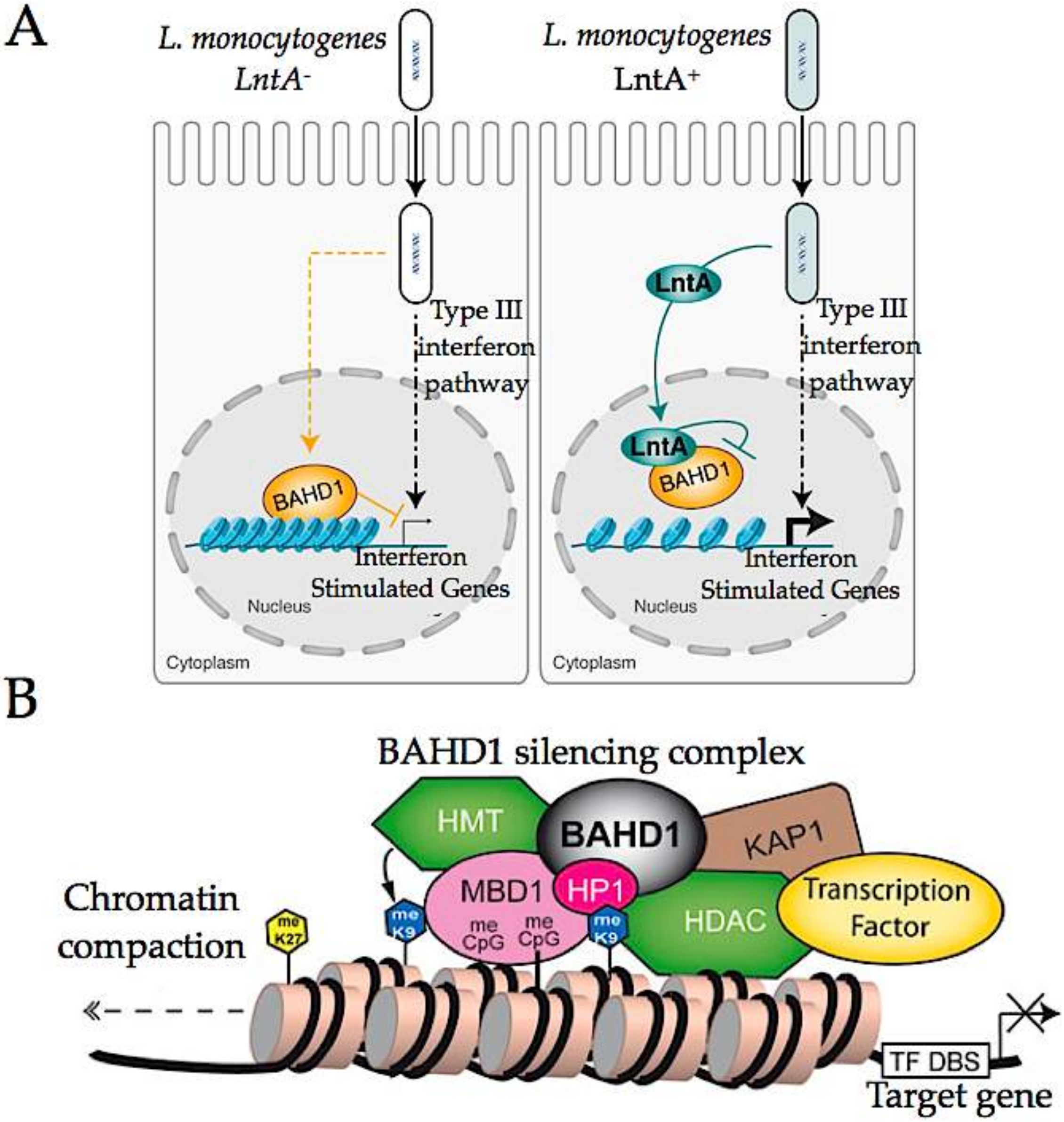


Figure 6.26: BAHD1-LntA mediated responses.

[A] Interplay between the bacterial virulence factor (LntA) and the host BAHD1 silencing complex represses a set of immunity genes induced by type III interferon in cells. [B] BAHD1 acts as a silencer by tethering chromatin regulators (HP1, MBD1, KAP1) and chromatin modifying enzymes (HMT and HDAC) to sequence-specific transcription factors, enabling local chromatin compaction at specific target genes. Adapted from Lebreton et al., 2011.

6.4.3.2 INPP5B potential role in *L. monocytogenes* host cell colonization

INPP5B is an inositol 5-phosphatase that preferentially dephosphorylates the 5′ position of PI(4,5)P2 and PI(3,4,5)P3. This gene is known to be identical to the oculocerebrorenal syndrome of lowe gene (OCRL). The two proteins have primarily been studied in different contexts but appear to have at least partially overlapping functions and similar enzymatic properties (Zhang, Jefferson et al. 1995; Bohdanowicz, Balkin et al. 2012). Studies conducted by Kuhbacher et al. demonstrated diminishing the activity of OCRL leads to an increase in the internalization levels of *L. monocytogenes* in HeLa cells (Kuhbacher, Dambournet et al. 2012). They showed that OCRL regulates bacterial colonization of HeLa cells by modulating signaling cascades downstream of the Met receptor, which is a major entry point used by the bacteria. In the same study, they demonstrated that through its phosphatase enzymatic activity, OCRL restricts *L. monocytogenes* entry into host cell by modulating actin dynamics at bacterial internalization sites.

INPP5B may also play an important role in the regulation of *L. monocytogenes* at the phagosomal level. Upon internalization the bacteria pathogen are transiently entrapped in a tight phagosomal compartment before accessing their replicative niche. Work done by Bohdanowicz et al. demonstrated that INPP5B are important contributors to the signaling that is pivotal for phagocytosis (Bohdanowicz, Balkin et al. 2012). In the study, INPP5B was recruited to the phagosome by the adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing-2 (APPL1), which is encoded by the small GTPase Rab5 protein. Rab5 protein involvement in *L. moncytogenes* entry into host cells has also been demonstrated (Jiwanni, Wang et al. 2012).

6.4.3.3 LATS2 potential role in *L. monocytogenes* infection

Internalization of the bacterium *L. monocytogenes* into its replicative niche is mediated by two specific host proteins; the tyrosine kinase receptor MetA and the E-cadherin (E-cad) (Bonazzi, Lecuit et al. 2009). The latter protein is a transmembrane molecule that plays an important role in cell adhesion, forming adherens junctions to bind cells together within tissues. The bacteria exploit this host cellular process to gain access into the cell. It is therefore reasonable to believe that suppression in the expression of E-cad may play a vital role in bacterial colonization of the host cell. Recent studies have demonstrated the potential involvement of the tumor repressor gene LATS2 in the regulation of E-cad activation (Mazda, Nishi et al. 2011). The increase in bacteria host cell colonization observed in this study can be attributed to the E-cad inactivation following suppression of LATS2 transcriptional repressor by the LATS2 siRNA. Figure 6.27 shows the suggested E-cad response following suppression of the LATS2 transcriptional repressor gene.

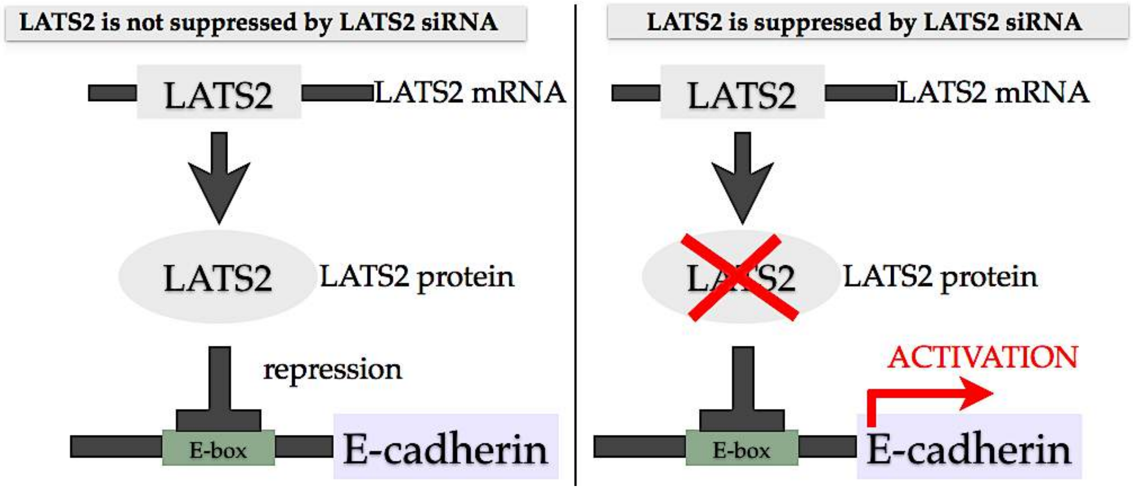


Figure 6.27: Possible mechanism of E-cadherin by LATS2 siRNA.

The transcription of E-cadherin is suppressed by LATS2 transcriptional repressor (left). When LATS2 is suppressed, the transcriptional repression of E-cadherin is alleviated. Adapted from Mazda et al., 2011 with modification (right).

CHAPTER 7

CONCLUDING REMARKS AND PERSPECTIVES

Although the bacterial pathogen *L. monocytogenes* has been used as a model organism to understand host-pathogen interactions for the past 15 years, researchers have not yet entirely elucidated the actual human host factors hijacked by the pathogen during infection. With the introduction of high-content array based screens, scientists are beginning to unveil some of the host adaptor proteins involved during infection. Here we described a high-density siRNA array platform and its subsequent implementation in search of human host factors involved during *L. monocytogenes* infection.

High throughput RNAi array-based phenotypic screening is increasingly becoming an important platform to screen for genes involved in biological pathways or affected during bacterial and viral infection. Researchers are currently using this high-density array platform to concurrently interrogate large number of genes involved in many cellular processes. Despite the impressive nature of the array technology, scientists are still faced with some major problems when performing high content screening (Echeverri, Beachy et al. 2006).

A major challenge faced when conducting a high-content RNAi-based screening is the off-target effect (Echeverri, Beachy et al. 2006; Sharma and Rao 2009; Mohr, Bakal et al. 2010), which if not properly controlled, may significantly contribute to false discovery of candidate proteins. The off-target effects experienced in array screening can be attributed to; [1] if the siRNA sequence used is complementary or nearly identical to a sequence in an unrelated mRNA, [2] if the 'seed region' of an siRNA binds with a weakly complementary sequence in the 3' untranslated region of the unrelated mRNA in which case, the siRNA can act like a microRNA and cause reduction of non-target proteins through mRNA degradation and lastly, [3] the use of saturated doses of siRNAs during screening may activate the antiviral type 1 interferon response in a sequence-independent manner (Echeverri, Beachy et al. 2006).

A second drawback seen in RNAi screens is the lack of overlap among screen hits when independent but related screens are performed. This effect has been demonstrated in studies carried out on mammalian cells infected with the human immuno-deficiency virus (König, Zhou et al. 2008; Zhou, Xu et al. 2008; Yeung, Houzet et al. 2009). Also, a similar study performed on two different Drosophila cell types, assaying for components of the JAK/STAT signaling pathway resulted in only 25% overlap in hits identified (Beller, Sztalryd et al. 2008; Guo, Walther et al. 2008).

Another challenge faced by researchers in the field of high-content screening is choosing the appropriate statistical methods to define the initial 'hit' list (Niederlein, Meyenhofer et al. 2009; Huang 2010). High-content screenings generally generate huge datasets that makes data handling very cumbersome and difficult to analyze. Also high throughput screeners are faced with the problem of choosing the appropriate arbitrary threshold for hit selection. As to the latter, different groups used different arbitrary threshold levels depending on the axis of apparent activity (Genovesio, Kwon et al. 2011).

To address these drawbacks, we adjusted our screening approach by optimizing the assay responses to the internal controls. The control values enabled us to define the baseline and signal values for the screen, which we reliably use to predict where the relevant targets should belong. The former was achieved by titrating the doses, incubation times and other variables used during validation screen assays. Also to minimize phenotypic variation in our screening approach we made use of a pool siRNA library that has been reported to greatly reduce the of-target effect (Tilesi, Fradiani et al. 2009). To increase the stringency in our hit selection, we used 13 measuring parameters, across fifteen screens yielding 614,250 experiments (including the internal controls). Data resulting from the screens were analyzed using a multi-dimensional analytical tool, the principal component analysis (PCA). The latter was used to create a correlation matrix of all variables by reducing the dimensionality of the dataset consisting of interrelated variables, while retaining as much of the variation present in the data as possible. This approach enabled us to generate different clusters of candidate genes depending on the different phenotypes expressed.

It was important to validate the screening platform to establish the high-density array was suitable for use in screening. To this end, a matrix of measurements was applied to determine print quality that were chosen based on the known outputs expected from RNAi array experiments, for example that spots should be printed and gene silencing be measurable. Criteria including the number of printed spots, dried spot intensity, shape and efficiency of RELA/p65 silencing were selected. Here, we demonstrated that printed nucleic acid spots size were regular with an average surface diameter of 351µm±0.5µm (Figure 4.1), gene silencing was measurable with a > 70 % silencing efficiency (Figure 4.3-4.5), and spot-to-spot cross contamination was minimal (Figure 4.2).

We further demonstrated the effectiveness of the array platform by screening for cellular components involved in the TNF-α to NF-κB signaling pathway. The NF-κB signaling pathway was chosen because it has been extensively characterized and documented. Here we used the high-density siRNA array comprising of all the kinases, phosphatases, ubiquitinylases and proteases to screen for genes involved in the regulation of TNF-α induced NF-κB/RELA nuclear translocation. To achieve this, a TNF-α induced NF-κB/RELA nuclear translocation assay was performed and the Z-factor (Z') for the assay calculated. A Z' >7 (Figure 5.5 D) indicated a suitable difference in signal and background values together with low variability making the assay suitable for use in screening for potential human host factors implicated in the NF-κB/RELA signaling pathway (Zhang, Chung et al. 1999). As documented in other studies (Ngo, Davis et al. 2006; Luo, Cheung et al. 2008; Sharma and Rao 2009), increasing the number of replicates in a screen helps to minimized false discovery in the screen. To achieve the latter standard, we performed a total of 32 replicate screens using the assay format described above. For data analysis, 12 criteria were incorporated each directed to measure the RELA nucleus to cytoplamic ratio intensity. For hit selection we focused only on genes whose silencing resulted in a decrease in nucleus to cytoplasmic ratio intensity. Here we identified 51 genes to be implicated in the TNF-α to NF-κB/RELA signaling pathway (Figure 5.9). Genes were interrogated for protein-protein interactions using the pathway studio and 26 of the genes had direct interaction with the NF-κB signaling pathway amongst which we identified two genes; CHUK (Conserved Helix-Loop-Helix Ubiquitous Kinase) and IKBKB (Inhibitor Of Kappa

Light Polypeptide Gene Enhancer In B-Cells, Kinase Beta), which are key regulators to the pathway (Figure 5.10).

Transposing the array format used in the TNF-α to NF-κB (RELA), we screened for human host factors involved in *L. monocytogenes* infection. To achieve this, we first conducted experiments to demonstrate that the bacteria were internalized in the cell rather than lying on the cell surface. The bacterial pathogen once internalized polymerizes the host actin cytoskeleton to generate an actin rocket tail. By staining the cell actin cytoskeleton we demonstrated that the bacteria were internalized as we could see the bacterial actin rocket tails by visual inspection of the fluorescently labeled images of the infected cells. This was further confirmed through a Z-scan series performed on the infected cells (Figure 6.3). For internalized bacterial quantification we used an off-the-shelf image analysis software from the Ultra image Xpress from Molecular Devices (Figure 6.4-6.6). The analysis platform was used to quantify for the optimal MOI for infection as well as the optimal infection time required for screening (Figure 6.9). Here we performed 15 replicate screens and data from the screens were normalized against the existing values of the non-targeting control. For hit identification, we employed 13 bacteria quantification parameters. Data were projected into the PCA and using a >2 STDV from the mean value of the control, we identified candidate genes whose silencing enhanced or hindered bacteria entry as well as those that affected the cell actin cytoskeleton.

To eradicate any errors that may have resulted from the array screening and to further confirm the stringency of the hits, we reproduced the screen in 96-well plates using a new set of pooled siRNAs corresponding to the primary hits. The screen was reproduced six times and extracted data were normalized against the existing values of the negative control (non-targeting siRNA). Using a >4 STDV cut-off mark from the control, we identified 15 candidate genes whose silencing greatly enhanced bacterial entry. Among the candidate genes obtained, we recognized a transcription factor, bromo adjacent homology domain containing 1 (BAHD1), whose interaction with the bacterial virulence factor LntA has been shown to mediate bacterial colonization of the host cell. The inositol polyphosphate-5-phosphatase (INPPB5) gene may achieve the latter by regulating bacterial internalization by modulating actin dynamics at bacterial internalization sites as shown with the oculocerebrorenal syndrome of lowe (OCRL) gene with which the INPP5B shares an identical sequence and overlapping enzyme activities (Bohdanowicz, Balkin et al. 2012; Kuhbacher, Dambournet et al. 2012). We also identified the large tumor suppressor kinase 2 (LATS2) that may play a potential role in bacterial colonization by modulating the expression of the host E-cad receptor (Mazda, Nishi et al. 2011).

Undoubtedly further validation steps are usually required to confirm primary candidate genes originating from the TNF-α to NF-κB (RELA) as well as for *L. monocytogenes* screens. The most commonly used approach to confirm primary candidate genes in siRNA pooled screens is to reproduce the results using the individual siRNA duplexes from the pooled siRNA. The idea behind this strategy is that if two or more different siRNA duplexes yield the same phenotypic expression for a particular gene, the chance that this is an off-target effect is considerably reduced and, therefore, the hit is confirmed. However, studies have shown that there is an estimated probability of > 95 % gene silencing when using siRNA pooled libraries (Echeverri and Perrimon 2006; Sharma and Rao 2009; Mohr, Bakal et al. 2010; Yang, Lacson et al. 2012).

Another approach that could be introduced in further validation of the prime candidates is using gene-specific quantitative methods such RT-PCR or microarray analysis (Yang, Lacson et al. 2012). These validation approaches will give a perspective on which RNAi reagents are over or under-represented in the experimental set(s) compared with the control set(s).

The list of 65 genes established during the screen likely does not exhaustively recapitulate all aspects involved in bacteria-host interactions during the pathogen's entire life cycle. It should however be noted that these candidate genes only provide us with an inventory list of hits that are or may be taking part in one or more stages of the pathogen intracellular life cycle. To obtain important information about a gene function as well as the molecular events that underlie a biological process or pathway, further studies will be required. This can be addressed through the use of classical genetic screens, which when properly performed will provide hint about the molecular events that underpin a particular biological process.

In conclusion, the list does however generate potential new candidates of cellular targets for antibacterial therapy development.

BIBLIOGRAPHY

AbdEl-Malek, A. M., S. F. H. Ali, et al. (2010). "Occurrence of *Listeria* species in meat, chicken products and human stools in Assiut city, Egypt with PCR use for rapid identification of *Listeria monocytogenes*." Vet World **3**: 353-359.

Agaisse, A., L. S. Burrack, et al. (2005). "Genome-Wide RNAi Screen for Host Factors Required for Intracellular Bacterial Infection." Science **309** 1248-1251.

Allerberger, F. (2003). "*Listeria*: growth, phenotypic differentiation and molecular microbiology." FEMS Immunol Med Microbiol **35**: 183-189.

Allerberger, F. and M. Wagner (2010). "Listeriosis: a resurgent foodborne infection." Clin Microbiol Infect **16**: 16-23.

Ashcroft, F. J., J. Y. Newberg, et al. (2011). "High content imaging-based assay to classify estrogen receptor-α ligands based on defined mechanistic outcomes." Gene **477**: 42-52.

Balestrino, D., M. A. Hamon, et al. (2010). "Single-cell Techniques Using Chromosomally Tagged Fluorescent Bacteria To Study *Listeria monocytogenes* Infection Processes." Appl Environ Microbiol **76**: 3625-3636.

Beauregard, K., K. Lee, et al. (1997). "pH-dependent perforation of macrophage phagosomes by listeriolysin O from *Listeria monocytogenes*." J Exp med **186**: 1159-1163.

Beller, M., C. Sztalryd, et al. (2008). "COPI complex is a regulator of lipid homeostasis." PLoS Biol **6**: e292.

Beltzner, C. C. and T. D. Pollard (2004). "Identification of functionally important residues of Arp2/3 complex by analysis of homology models from diverse species." Mol Biol **336**: 551-565.

Ben-Hur, A. and I. Guyon (2003). "Detecting stable clusters using principal component analysis. " Functional Genomics: Meth Prot: 159-182.

Bierne, H. and P. Cossart (2002). "InIB, a surface protein of *Listeria monocytogenes* that behaves as an invasin and a growth factor." Cell Sci **115**: 3357-3367.

Bierne, H. and P. Cossart (2007). "*Listeria monocytogenes* Surface Proteins: from Genome Predictions to Function." Microbiol Mol Biol Rev **71**: 377-397.

Bierne, H., C. Garandeau, et al. (2004). "Sortase B, a new class of sortase in *Listeria monocytogenes*." Bacteriol **186**: 1972-1982.

Bille, J., B. Catimel, et al. (1992). "API *Listeria*, a new and promising one-day system to identify *Listeria* isolates." Appl Environ Microbiol **58**: 1857-1860.

Blajecka, K., M. Marinov, et al. (2012). "Phosphoinositide 3-Kinase C2β Regulates RhoA and the Actin Cytoskeleton through an interaction with DbI." PLoS One **7**: e44945.

Blank, M. and Y. Shiloh (2007). "Programs for Cell Death: Apoptosis is Only One Way to Go." Cell Cycle **6**: 686-695.

Bohdanowicz, M., D. M. Balkin, et al. (2012). "Recruitment of OCRL and Inpp5B to phagosomes by Rab5 and APPL1 depletes phosphoinositides and attenuates Akt signaling." Mol Biol Cell **23**: 1176-1187.

Bonazzi, M., M. Lecuit, et al. (2009). "*Listeria monocytogenes* Internalin and E-cadherin: From Bench to Bedside." Cold Spring Harb Perspect Biol **1**: a003087.

Bonazzi, M., M. Lecuit, et al. (2009). "*Listeria monocytogenes* internalin and E-cadherin: from structure to pathogenesis." Cell Microbiol **11**: 693-702.

Bonazzi, M., L. Vasudevan, et al. (2011). "Clathrin phosphorylation is required for actin recruitment at sites of bacterial adhesion and internalization." J Cell Biol **195**: 525–536.

Bonizzi, G. and M. Karin (2004). "The two NF-kappaB activation pathways and their role in innate and adaptive immunity." Trends Immunol **25**: 280–288.

Bossche, J., B. Malissen, et al. (2012). "Regulation and function of the E-cadherin/catenin complex in cells of the monocyte-macrophages lineage and DCs." Blood **119**: 1623-1633.

Brass, A. L., D. M. Dykxhoorn, et al. (2008). "Identification of host proteins required for HIV infection through a functional genomic screen." Science **319**: 921-926.

Braun, L. and P. Cossart (2000). "Interactions between *Listeria monocytogenes* and host mammalian cells." Microbes Infect **2**: 803-811.

Braun, L., H. Ohayon, et al. (1998). "The InIB protein of *Listeria monocytogenes* is sufficient to promote entry into mammalian cells." Mol Microbiol **27**: 1077-1087.

Breed, R. S., E. G. D. Murray, et al. (1948). "Bergey's manual of determinative bacteriology, 6th ed." The William and Wilkins Co., Baltimore.

Bubert, A., I. Hein, et al. (1999). "Detection and differentiation of *Listeria* spp. by a single reaction based on multiplex PCR." Appl Environ Microbiol **65**: 4688-4692.

Camilli, A., L. G. Tilney, et al. (1993). "Dual roles of plcA in *Listeria monocytogenes* pathogenesis." Mol Microbiol **8**: 143-157.

Carlier, M. F. and D. Pantaloni (2007). "Control of actin assembly dynamics in cell motility." J Biol Chem **282**: 23005-23009.

Cartwright, E. J., K. A. Jackson, et al. (2013). "Listeriosis outbreaks and associated food vehicles, United States, 1998-2008." Emerg Infect Dis **19**: 1-9; quiz 184.

Cervantes, C. F., P. R. Markwick, et al. (2009). "Functional dynamics of the folded ankyrin repeats of I kappa B alpha revealed by nuclear magnetic resonance." Biochemistry **48**: 8023-8031.

Cheng, A., S. Magdaleno, et al. (2011). "Optimization of transfection conditions and analysis of siRNA potency using real-time PCR." Methods Mol Biol **764**: 199-213.

Chiang, H.-I., L. R. Berghman, et al. (2009). "Inhibition of NF-κB 1 (NF-κBp50) by RNA interference in chicken macrophage HD11 cell line challenged with *Salmonella enteritidis*." Genet Mol Biol **32**: 507–515.

Chong, R., R. Swiss, et al. (2009). "Regulatory mimicry in *Listeria monocytogenes* actin-based motility." Cell Host Microbe **6**: 268-278.

Connelly, M. A. and K. B. Marcu (1995). "CHUK, a new member of the helix-loop-helix and leucine zipper families of interacting proteins, contains a serine-threonine kinase catalytic domain." Cell Mol Biol Res **41**: 537–549.

Cossart, P. (2011). "Illuminating the landscape of host-pathogen interactions with the bacterium *Listeria monocytogenes*." Proc Natl Acad Sci U S A **108**: 19484-19491.

Cossart, P. and C. Archambaud (2009). "The bacterial pathogen *Listeria monocytogenes*: an emerging model in prokaryotic transcriptomics." J Biol **8**: 107.

Cossart, P. and M. Lecuit (1998). "Interactions of *Listeria monocytogenes* with mammalian cells during entry and actin-based movement: bacterial factors, cellular ligands and signaling." EMBO J **17**: 3797-3806.

Cossart, P. and P. J. Sansonetti (2004). "Bacterial invasion: the paradigms of enteroinvasive pathogens." Science **304**: 242-248.

Devling, T. W., C. D. Lindsay, et al. (2005). "Utility of siRNA against Keap1 as a strategy to stimulate a cancer chemopreventive phenotype." Proc Natl Acad Sci U S A **102**: 7280-7285A.

Dokainish, H., B. Gavicherla, et al. (2007). "The carboxyl-terminal SH3 domain of the mammalian adaptor CrkII promotes internalization of *Listeria monocytogenes* through activation of host phosphoinositide 3-kinase." Cell Microbiol **9**: 2497-2516.

Dow, L. E., P. K. Premsrirut, et al. (2012). "A pipeline for the generation of shRNA transgenic mice." Nat Protoc **7**: 374-393.

Dramsi, S. and P. Cossart (2002). "Listeriolysin O: a genuine cytolysin optimized for an intracellular parasite." Cell Biol **156** 943-946.

Echeverri, C. J., P. A. Beachy, et al. (2006). "Minimizing the risk of reporting false positives in large-scale RNAi screens." Nat Methods **3**: 777-779.

Echeverri, C. J. and N. Perrimon (2006). "High-throughput RNAi screening in cultured cells: a user's guide." Nat Rev Genet **7**: 373-384.

Firat-Karalar, E. N. and M. D. Welch (2011). "New mechanisms and functions of actin nucleation." Curr Opin Cell Biol **23**: 4-13.

Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**: 806-811.

Furge, K. A., Y. W. Zhang, et al. (2000). "Met receptor tyrosine kinase: enhanced signaling through adapter proteins." Oncogene **19**: 5582-5589.

Gaillard, J. L., P. Berche, et al. (1991). "Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci." Cell **65**: 1127-1141.

Gekara, N. O., N. Zietara, et al. (2010). "*Listeria monocytogenes* Induces T Cell Receptor Unresponsiveness through Pore-Forming Toxin Listeriolysin O." Infect Dis **202**: 1698-1707.

Genovesio, A., M. A. Giardini, et al. (2011). "Visual genome-wide RNAi screening to identify human host factors required for *Trypanosoma cruzi* infection." PLoS One **6**: e19733.

Genovesio, A., Y.-J. Kwon, et al. (2011). "Automated genome-wide visual profiling of cellular proteins involved in HIV infection." Biomol Screen **16**: 945-958.

Gewurz, B. E., F. Towfic, et al. (2012). "Genome-wide siRNA screen for mediators of NF-κB activation." Proc Natl Acad Sci U S A **109**: 2467-2472.

Glomski, I. J., A. L. Decatur, et al. (2003). "*Listeria monocytogenes* mutants that fail to compartmentalize listerolysin O activity are cytotoxic, avirulent, and unable to evade host extracellular defenses." Infect Immun **71**: 6754-6765.

Goff, S. P. (2008). "Knockdown screens to knockout HIV-1." Cell **135**: 417–420.

Goulet, V., C. Hedberg, et al. (2008). "Increasing incidence of listeriosis in France and other European countries." Emerg Infect Dis **14**: 734-740.

Goulet, V., C. Jacquet, et al. (2006). "Surveillance of human listeriosis in France, 2001-2003." Euro Surveill **11**: 79-81.

Graves, L. M., L. O. Helsel, et al. (2010). "*Listeria marthii* sp. nov., isolated from the natural environment, Finger Lakes National Forest." Int J Syst Evol Microbiol **60**: 1280-1288.

Gründling, A., M. D. Gonzalez, et al. (2003). "Requirement of the *Listeria monocytogenes* broad-range phospholipase PC-PLC during infection of human epithelial cells." Bacteriol **185**: 6295-6307.

Gunter, B., C. Brideau, et al. (2003). "Statistical and graphical methods for quality control determination of high-throughput screening data." J Biomol Screen **8**: 624-633.

Guo, Y., T. C. Walther, et al. (2008). "Functional genomic screen reveals genes involved in lipid-droplet formation and utilization." Nature **453**: 657–666.

Haas, J., K. Kusinski, et al. (2007). "A Ride with *Listeria monocytogenes*: A Trojan Horse." Eukaryon **3**: 1-8.

Haghighat, A. C. and S. Seveau (2010). "Quantification of host-microbe interactions by automated fluorescence microscopy." J Immunol Methods **352**: 186-191.

Hain, T., S. S. Chatterjee, et al. (2007). "Pathogenomics of *Listeria* spp." Int J Med Microbiol **297**: 541-557.

Hain, T., C. Steinweg, et al. (2006). "Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*." J Bacteriol **188**: 7405-7415.

Hamon, M., H. Bierne, et al. (2006). "*Listeria monocytogenes*: a multifaceted model." Nat Microbiol Rev **4**: 423-434.

Hamon, M. A., E. Batsche, et al. (2007). "Histone modifications induced by a family of bacterial toxins." Proc Natl Acad Sci USA **104**: 13467-13472.

Hamon, M. A., D. Ribet, et al. (2012). "Listeriolysin O: the swiss army knife of *Listeria*." Trends Microbiol **20**: 360-368.

Hampel, F. R., E. M. Ronchetti, et al. (1986). "Robust Statistics: The Approach Based on Influence Functions." John Wiley & Sons New York: ISBN: 9780471829218.

Hayden, M. S. and S. Ghosh (2008). "Shared Principles in NF-κB Signaling." Cell Rev **132**: 344-362.

Henmi, Y., K. Tanabe, et al. (2011). "Disruption of microtubule network rescues aberrant actin comets in dynamin2-depleted cells." PLoS One **6**: e28603.

Henry, R., L. Shaughnessy, et al. (2006). "Cytolysin-dependent delay of vacuole maturation in macrophages infected with *Listeria monocytogenes*." Cell Microbiol **8**: 107-119.

Hense, M., E. Domann, et al. (2001). "Eukaryotic expression plasmid transfer from the intracellular bacterium *Listeria monocytogenes* to host cells." Cell Microbiol **3**(9): 599-609.

Hoffmann, A., G. Natoli, et al. (2006). "Transcriptional regulation via the NFkB signaling module." Oncogene **25**: 6706-6716.

Hosseinpour, B., V. HajiHoseini, et al. (2012). "Female Gametophyte Development Identifies Novel Proteins and Relations." PLoS One **7**: e49931.

Huang, S. (2010). "Statistical issues in subpopulation analysis of high content imaging data." J Comput Biol **17**: 879-894.

Huang, S. M., M. K. Hancock, et al. (2009). "Negative regulators of insulin signaling revealed in a genome-wide functional screen." PLoS One **4**: e6871.

Hulphers, G. (1911). "Lefvernekos hos kanin orsakad af en ej forut beskrifven bakteria." Svenska Vet Tidskr **2**: 265–268.

Humphries, C. L., H. I. Balcer, et al. (2002). "Direct regulation of Arp2/3 complex activity and function by the actin binding protein coronin." J cell Biol **159**: 993-1004.

Ireton, K. (2013). "Molecular mechanisms of cell-to-cell spread of intracellular bacteria pathogens." Open Biol **3**: 130079.

Ivanek, R., Y. T. Grohn, et al. (2006). "*Listeria monocytogenes* in multiple habitats and host populations: review of available data for mathematical modeling." Foodborne Pathog Dis **3**: 319-336.

Jiwanni, S., Y. Wang, et al. (2012). "Identification of Components of the Host Type IA Phosphoinositide 3-Kinase Pathway That Promote Internalization of Listeria monocytogenes." Infect Immun **80**: 1252–1266.

Jolliffe, I. T. (1986). "Principal Component Analysis, Second Edition." Springer Verlag.

Jonquieres, R., H. Bierne, et al. (1999). "Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria." Mol Microbiol **34**: 902-914.

Jonquières, R., J. Pizarro-Cerda, et al. (2001). "Synergy between the N- and C-terminal domains of InlB for efficient invasion of non phagocytic cells by *Listeria monocytogenes*." Mol. Microbiol **42**: 955–965.

Karin, M., Y. Cao, et al. (2002). "NF-κB in cancer: from innocent bystander to major culprit." Nat Rev **2**: 301-310.

Kayal, S. and A. Charbit (2006). "Listeriolysin O: a key protein of *Listeria monocytogenes* with multiple functions." FEMS Microbiol Rev **30**: 514-529.

König, R., Y. Zhou, et al. (2008). "Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication." Cell **135**: 49–60.

Kota, K. P., B. Eaton, et al. (2013). "Integrating High-Content Imaging and Chemical Genetics to Probe Host Cellular Pathways Critical for *Yersinia Pestis* Infection." PLoS One **8**: e55167.

Krappmann, D. and C. Scheidereit (2005). "A pervasive role of ubiquitin conjugation in activation and termination of IkappaB kinase pathways." EMBO Rep **6**: 321-326.

Kuhbacher, A., D. Dambournet, et al. (2012). "Phosphatidylinositol 5-phosphatase oculocerebrorenal syndrome of Lowe protein (OCRL) controls actin dynamics during early steps of *Listeria monocytogenes* infection." J Biol Chem **287**: 13128-13136.

Lätzer, J., G. A. Papoian, et al. (2007). "Induced Fit, Folding, and Recognition of the NF-κB-Nuclear Localization Signals by IκBα and IκBβ." J Mol Biol **367**: 262–274.

Lebreton, A., P. Cossart, et al. (2012). "Bacteria tune interferon responses by playing with chromatin." Virulence **3**: 87-91.

Lebreton, A., G. Lakisic, et al. (2011). "A Bacterial Protein Targets the BAHD1 Chromatin Complex to Stimulate Type III Interferon Response." Science **331**: 1319-1321.

Lecuit, M. (2007). "Human listeriosis and animal models." Microbes Infect **9**: 1216-1225.

Lecuit, M., H. Ohayon, et al. (1997). "Internalin of *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to promote internalization." Infect Immun **65**: 5309–5319.

Lee, K.-C. and A. J. Liu (2008). "New Proposed Mechanism of Actin-Polymerization-Driven Motility." Biophysical **95**: 4529–4539.

Liu, D. (2006). "Identification, subtyping and virulence determination of *Listeria monocytogenes*, an important foodborne pathogen." J Med Microbiol **55**: 645-659.

Lodish, H., A. Berk, et al. (2000). "Molecular Cell Biology, 4th edition." New York: W. H. Freeman: ISBN-10: 10-7167-3136-7163.

Lu, W., G. Zhang, et al. (2010). "Tumor site-specific silencing of NF-kappaB p65 by targeted hollow gold nanosphere-mediated photothermal transfection." Cancer Res **70**: 3177-3188.

Luo, B., H. W. Cheung, et al. (2008). "Highly parallel identification of essential genes in cancer cells." Proc Natl Acad Sci U S A **105**: 20380-20385.

Machesky, L. M. (1997). "Cell motility: Complex dynamics at the leading edge." Curr Biol **7**: R164–R167.

Malo, N., J. A. Hanley, et al. (2006). "Statistical practice in high-throughput screening data analysis." Nat Biotechnol **24**: 167-175.

Marino, M., M. Banerjee, et al. (2002). "GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands." EMBO J **21**: 5623-5634.

Mazda, M., K. Nishi, et al. (2011). "E-cadherin is transcriptionally activated via suppression of ZEB1 transcriptional repressor by small RNA-mediated gene silencing." PLoS One **6**: e28688.

Mead, P. S., L. Slutsker, et al. (1999). "Food-related illness and death in the United States." Emerg Infect Dis **5**: 607-625.

Mengaud, J., H. Ohayon, et al. (1996). "E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells." Cell **84**: 923-932.

Mercurio, F., B. W. Murray, et al. (1999). "IκB kinase (IKK)-associated protein 1, a common component of the heterogeneous IKK complex." Mol Cell Biol **19**: 1526–1538.

Miyamoto, S. (2011). "Nuclear initiated NF-κB signaling: NEMO and ATM take center stage." Cell Res **21**: 116-130.

Mohr, S., C. Bakal, et al. (2010). "Genomic screening with RNAi: results and challenges." Annu Rev Biochem **79**: 37-64.

Mostowy, S. and P. Cossart (2011). "Septins as key regulators of actin based processes in bacterial infection." Biol Chem **392**: 831-835.

Mostowy, S., T. Nam Tham, et al. (2009). "Septins regulate bacterial entry into host cells." PLoS One **4**: e4196.

Moynagh, P. N. (2005). "The NFkB pathway." J Cell Sci **118**: 4389-4392.

Mullenders, J., A. W. Fabius, et al. (2009). "A large scale shRNA barcode screen identifies the circadian clock component ARNTL as putative regulator of the p53 tumor suppressor pathway." PLoS One **4**: e4798.

Murray, E. D. G., R. A. Webb, et al. (1926). "A disease of rabbits characterized by a large mononuclear leukocytosis, caused by a hitherto undescribed bacillus Bacterium monocytogenes (n. sp.)." Pathol Bacteriol **29**: 407–439.

Nadon, C. A., D. L. Woodward, et al. (2001). "Correlations between molecular subtyping and serotyping of *Listeria monocytogenes*." J Clin Microbiol **39**: 2704-2707.

Ngo, V. N., R. E. Davis, et al. (2006). "A loss-of-function RNA interference screen for molecular targets in cancer." Nature **441**: 106-110.

Niederlein, A., F. Meyenhofer, et al. (2009). "Image analysis in high-content screening." Comb Chem High Throughput Screen **9**: 899-907.

Niemann, H. H., V. Jäger, et al. (2007). "Structure of the Human Receptor Tyrosine Kinase Met in Complex with the *Listeria* Invasion Protein InlB." Cell **130**: 235-246.

Niforou, K. M., A. K. Anagnostopoulos, et al. (2008). "The proteome profile of the human osteosarcoma U2OS cell line." Cancer Genomics Proteomics **5**: 63-78.

Nikitin, A., S. Egorov, et al. (2003). "Pathway studio the analysis and navigation of molecular networks." Bioinformatics **19**: 2155–2157.

Nishikori, M. (2005). "Classical and Alternative NF-κB Activation Pathways and Their Roles in Lymphoid Malignancies." Clin Exp Hematopathol **45**: 15-24.

Nolan, G. P., S. Ghosh, et al. (1991). "DNA binding and I kappa B inhibition of the cloned p65 subunit of NF-kappa B, a rel-related polypeptide." Cell **64**: 961–969.

O'Riordan, M., C. H. Yi, et al. (2002). "Innate recognition of bacteria by a macrophage cytosolic surveillance pathway." Proc Natl Acad Sci U S A **99**(21): 13861-11386.

Oeckinghaus, A. and S. Ghosh (2009). "The NF-kappaB family of transcription factors and its regulation." Cold Spring Harb Perspect Biol **1**: a000034.

Paschen, A., K. E. Dittmar, et al. (2000). "Human dendritic cells infected by *Listeria monocytogenes*: induction of maturation, requirements for phagolysosomal escape and antigen presentation capacity." Eur J Immunol **30**(12): 3447-3456.

Patil, A. B., S. Nadiger, et al. (2007). "*Listeria monocytogenes* meningitis: an uncommon opportunistic infection in HIV/AIDS." Indian J Pathol Microbiol **50**: 671-673.

Pentecost, M., J. Kumaran, et al. (2010). "*Listeria monocytogenes* Internalin B Activates Junctional Endocytosis to Accelerate Intestinal Invasion." PLoS Pathog **6**: e1000900.

Pentecost, M., G. Otto, et al. (2006). "*Listeria monocytogenes* Invades the Epithelial Junctions at Sites of Cell Extrusion." PLoS Pathog **2**: e3.

Pistor, S., L. Gröbe, et al. (2000). "Mutations of arginine residues within the 146-KKRRK-150 motif of the ActA protein o*f Listeria monocytogenes* abolish intracellular motility by interfering with the recruitment of the Arp2/3 complex." Cell Sci **113**: 3277-3287.

Pizarro-Cerdá, J., M. Bonazzi, et al. (2010). "Clathrin-mediated endocytosis: What works for small, also works for big." Bioessays **32**: 496–504.

Pizarro-Cerdá, J. and P. Cossart (2009). "*Listeria monocytogenes* Membrane Trafficking and Lifestyle: The Exception or the Rule?" Cell Dev. Biol. **25**: 649-670.

Pizarro-Cerdá, J., S. Sousa, et al. (2004). "Exploitation of host cell cytoskeleton and signalling during *Listeria monocytogenes* entry into mammalian cells." Microbiol **327**: 523–531.

Portnoy, D. A., J. P. S., et al. (1988). "Role of hemolysin for the intracellular growth of *Listeria monocytogenes*." J Exp Med **167**(4): 1459-1471.

Poulsen, K. P. and C. J. Czuprynski (2013). "Pathogenesis of listeriosis during pregnancy." Anim Health Res Rev: 1-10.

Pust, S., H. Morrison, et al. (2005). "*Listeria monocytogenes* exploits ERM protein functions to efficiently spread from cell to cell." EMBO J **24**: 1287-1300.

Py, B. F., M. M. Lipinski, et al. (2007). "Autophagy limit *Listeria monocytogenes* intracellular growth in the early phase of primary infection." Autophagy **3**: 117-125.

Radtke, A. L., K. L. Anderson, et al. (2011). "*Listeria monocytogenes* exploits cystic fibrosis transmembrane conductance regulator (CFTR) to escape the phagosome." Proc Natl Acad Sci U S A **108** 1633–1638.

Ramaswamy, V., V. M. Cresence, et al. (2007). "*Listeria*-review of epidemiology and pathogenesis." J Microbiol Immunol Infect **40**: 4-13.

Regnier, C. H., H. Y. Song, et al. (1997). " Identification and characterization of an IκB kinase." Cell **90**: 373–383.

Rocourt, J. and C. Buchrieser (2007). "The Genus *Listeria* and *Listeria* monocytogenes Phylogenetic position, Taxonomy and Identification." *Listeria*, *Listeriosis* and Food Safety (3rd ed), FL: CRC Press, Boca Raton: p1-12 In ET Ryers and E H Marth (editors).

Rocourt, J. and H. H. Mollaret (1988). "Taxonomy of the genus *Listeria*." Infection **16**: 89-91.

Rossjohn, J., S. C. Fell, et al. (1997). "Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form " Cell **89**: 685-692.

Rothwarf, D. M., E. Zandi, et al. (1998). "IKKγ is an essential regulatory subunit of the IκB kinase complex." Nature **395**: 297–300.

Schmeck, B., W. Beermann, et al. (2006). "*Listeria monocytogenes* induced Rac1-dependent signal transduction in endothelial cells." Biochem Pharmacol **72**: 1367-1374.

Schmid, M. W., E. Y. W. Ng, et al. (2005). "Evolutionary history of the genus *Listeria* and its virulence genes." Syst Appl Microbiol **28**: 1-8.

Schnupf, P. and D. A. Portnoy (2007). "Listeriolysin O: a phagosome-specific Lysin." Microbes Infect **9**: 1176-1187.

Schnupf, P., J. Zhou, et al. (2007). "Listeriolysin O secreted by *Listeria monocytogenes* into the host cell cytosol is degraded by the N-end rule pathway." Infect Immun **75**: 5135-5147.

Schoumacher, M., R. D. Goldman, et al. (2010). "Actin, microtubules, and vimentin intermediate filaments cooperate for elongation of invadopodia." Cell Biol **189**: 541-556.

Schubert, W.-D., G. Göbel, et al. (2001). "Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain." J mol Biol **312**: 783–794.

Schuchat, A., C. Lizano, et al. (1991). "Outbreak of neonatal listeriosis associated with mineral oil." Pediatr Infect Dis J **10**: 183-189.

Schuerch, D. W., E. M. Wilson-Kubalek, et al. (2005). "Molecular basis of listeriolysin O pH dependence." Proc. Natl. Acad. Sci. U.S.A **102**: 12537-12542.

Schuppler, M. and M. J. Loessner (2010). "The Opportunistic Pathogen *Listeria monocytogenes*: Pathogenicity and Interaction with the Mucosal Immune System." Int J Inflam **2010**: 704321.

Serra, S. and R. Chetty (2008). "Revision 2: an immunohistochemical approach and evaluation of solid pseudopapillary tumour of the pancreas." Clin Pathol **61**: 1153-1159.

Seveau, S., J. Pizarro-Cerda, et al. (2007). "Molecular mechanisms exploited by *Listeria monocytogenes* during host cell invasion." Microbes Infect **9**: 1167-1175.

Sexton, J. Z., Q. He, et al. (2010). "High content screening for non-classical peroxisome proliferators." Int J High Throughput Screen **2010**: 127-140.

Sharma, S. and A. Rao (2009). "RNAi screening: tips and techniques." Nat Immunology **10**: 799-804.

Shatursky, O., A. P. Heuck, et al. (1999). "The Mechanism of Membrane Insertion for a Cholesterol-Dependent Cytolysin: A Novel Paradigm for Pore-Forming Toxins." Cell **99**: 293–299.

Shen, Y., M. Naujokas, et al. (2000). "InIB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase." cell **103**: 501-510.

Shieh, C. C., M. Coghlan, et al. (2000). "Potassium Channels: Molecular Defects, Diseases, and Therapeutic Opportunities." Pharmacological Rev **52**: 4557-4594.

Sims, D., A. M. Mendes-Pereira, et al. (2011). "High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing." Genome Biol **12**: R104.

Singh, R., A. Jamieson, et al. (2008). "GILT is a critical host factor for *Listeria monocytogenes* infection." Nat **455**: 1244-1247.

Sit, S. T. and E. Manser (2011). "Rho GTPases and their role in organizing the actin cytoskeleton." Cell Sci **124** 679-683.

Smith, G. A., H. Marquis, et al. (1995). "The two distinct phospholipases C of *Listeria monocytogenes* have overlapping roles in escape from a vacuole and cell-to-cell spread." Infect Immun **63**: 4231–4237.

Smith, G. A., J. A. Theriot, et al. (1996). "The tandem repeat domain in the *Listeria monocytogenes* ActA protein controls the rate of actin-based motility, the percentage of moving bacteria, and the localization of vasodilator-stimulated phosphoprotein and profilin." J Cell Biol **135**: 647–660.

Stavru, F., F. Bouillaud, et al. (2011). "*Listeria monocytogenes* transiently alters mitochondrial dynamics during infection." Proc Natl Acad Sci U S A **108**: 3612-3617.

Strezoska, Ž., A. Licon, et al. (2012). "Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens." PLoS One **7**: e42341.

Takaesu, G., R. M. Surabhi, et al. (2003). "TAK1 if critical for IκB kinase-mediated activation of the the NF-κB pathway." J Mol Biol **326**: 105-115.

Thi, E. P., U. Lambertz, et al. (2012). "Sleeping with the enemy: how intracellular pathogens cope with a macrophage lifestyle." PLoS Pathog **8**: e1002551.

Tian, F., T. Fan, et al. (2012). "A small interfering RNA targeting NF-κB p65 alone or combined with 5-FU inhibits growth of esophageal squamous cell carcinoma in nude mice." Pathol Res Pract **208**: 32-38.

Tilesi, F., P. Fradiani, et al. (2009). "Design and validation of siRNAs and shRNAs." Curr Opin Mol Ther **11**: 156–164.

Travier, L., S. Guadagnini, et al. (2013). "ActA promotes *Listeria monocytogenes* aggregation, intestinal colonization and carriage." PLoS Pathog **9**: e1003131.

Tweten, R. K. (2005). "Cholesterol-dependent cytolysin, a family of versatile pore-forming toxins." Infect Immun **73**: 6199-6209.

Vannberg, F. O., S. J. Chapman, et al. (2011). "Human genetic susceptibility to intracellular pathogens." Immunol Rev **240**: 105-116.

Vazquez-Boland, J. A., G. Dominguez-Bernal, et al. (2001). "Pathogenicity islands and virulence evolution in *Listeria*." Microbes Infect **3**: 571-584.

Vazquez-Boland, J. A., M. Kuhn, et al. (2001). "*Listeria* pathogenesis and molecular virulence determinants." Clin Microbiol Rev **14**: 584-640.

Warhurst, A. C., S. J. Hopkins, et al. (1998). "Interferon gamma induces differential upregulation of alpha and beta chemokine secretion in colonic epithelial cell lines." Gut **42**: 208-213.

Weiglein, I., W. Goebel, et al. (1997). "*Listeria monocytogenes* infection of HeLa cells results in listeriolysis O-mediated transient activation of the Raf-MEK-MAP kinase pathway." FEMS microbiol Letts **148**: 189-195.

Weih, F. and J. Caamaño (2003). "Regulation of secondary lymphoid organ development by the nuclear factor-κB signal transduction pathway." Immunol Rev **195**: 91–105.

Welch, M. D., J. Rosenblatt, et al. (1998). "Interaction of human Arp2/3 complex and the *Listeria monocytogenes* ActA protein in actin filament nucleation." Science **281**: 105-108.

Yamamoto, Y. and R. B. Gaynor (2004). "IkappaB kinases: key regulators of the NF-kappaB pathway." Trends Biochem Sci **29**: 72-79.

Yang, R., R. G. Lacson, et al. (2012). "A Genome-Wide siRNA Screen to Identify Modulators of Insulin Sensitivity and Gluconeogenesis." PLoS ONE **7**: e36384.

Yeung, K. Y. and W. L. Ruzzo (2000). "An empirical study on Principal Component Analysis for clustering gene expression data." Dept of Computer Science and Engineering, University of Washington: 1-26.

Yeung, M. L., L. Houzet, et al. (2009). "A genome-wide short hairpin RNA screening of Jurkat T-cells for human proteins contributing to productive HIV-1 replication." J Biol Chem **284**: 19463–19473.

Zhang, J. H., T. D. Chung, et al. (1999). "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays." J Biomol Screen **4**: 67-73.

Zhang, X., A. B. Jefferson, et al. (1995). "The protein deficient in Lowe syndrome is a phosphatidylinositol-4,5-bisphosphate 5-phosphatase." Proc Natl Acad Sci USA **92**: 4853-4856.

Zhou, H., M. Xu, et al. (2008). "Genome-scale RNAi screen for host factors required for HIV replication." Cell Host Microbe **4**: 495-504.

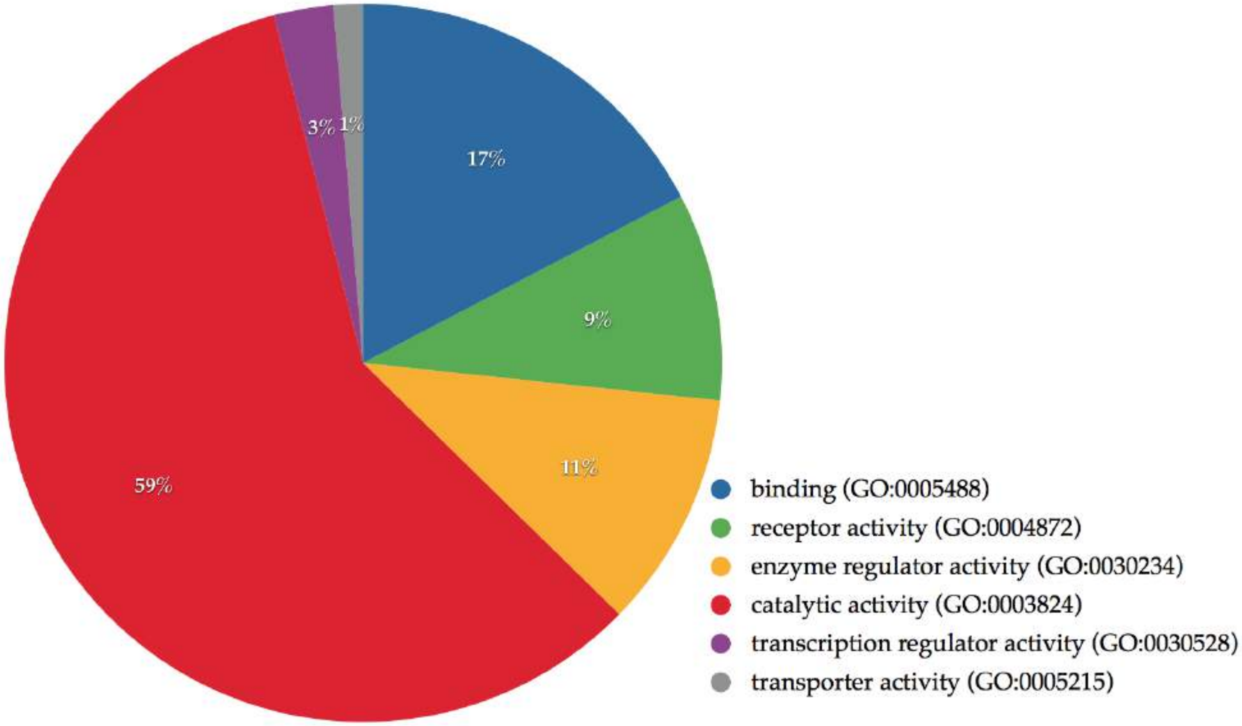
Ziauddin, J. and D. M. Sabatini (2001). "Microarrays of cells expressing defined cDNAs." Nature **411**: 107-110.

APPENDICES

Appendix table 1: p65 (NF-κB) nuclear import candidate genes

Gene Symbol	Gene Name	GO Molecular Function
ACVR2	Activin A receptor type 2A	Protein kinase activity
CDC14B	Cell division cycle 14B	Protein serine/threonine phosphatase activity
CDC25C	Cell division cycle 25C	Protein tyrosine phosphatase activity
CDKN2D	Cyclin-dependent kinase inhibitor 2D	Cyclin-dependent protein serine/threonine kinase inhibitor activity
CHUK	Conserved helix-loop-lelix ubiquitous kinase	Protein kinase activity
CIT	Citron (Rho-interacting, Serine/Threonine Kinase 21)	Protein serine/threonine kinase activity
CKM	Creatine Kinase, Muscle	Creatine kinase activity
CTDSP2	Carboxy-terminal domain RNA polymerase II polypeptide a small phosphatase	Protein binding
DUSP11	Dual specificity phosphatase 11 (RNA/RNP complex 1-interacting)	RNA binding
EPHB4	Ephrin type-B receptor 4	Ephrin receptor activity, Transmembrane receptor protein tyrosine kinase activity
EYA1	Eyes absent homolog 1 (Drosophila)	Protein tyrosine phosphatase activity
FGR	Feline gardner-rasheed sarcoma viral oncogene homolog	Non-membrane spanning protein tyrosine kinase activity
GAK	Cyclin G associated kinase	Protein serine/threonine kinase activity, ATP binding
GALK1	Galactokinase 1	Galactokinase activity
GNE	Glucosamine (UDP-N-Acetyl)-2-epimerase/N-acetylmannosamine kinase	ATP binding
GUCY2C	Guanylate cyclase 2C (heat stable enterotoxin receptor)	Protein binding, Guanylate cyclase activity
HK3	Hexokinase 3 (white cell)	Glucokinase activity
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase Beta	Protein kinase activity
ITPKB	Inositol-trisphosphate 3-kinase B	Protein binding
LATS2	Large tumor suppressor kinase 2	Protein binding activity, ATP binding
MAP2K1	Mitogen-activated protein kinase kinase 1	Protein kinase activity
MAPK11	Mitogen-activated protein kinase 11	Protein kinase activity, MAP kinase activity
MPP3	Membrane protein, palmitoylated 3 (mAGUK p55 subfamily member 3)	Guanylate kinase activity
MTMR14	Myotubularin related protein 14	Phosphatidylinositol-3-phosphatase activity
MYLK	Myosin light chain kinase	Myosin light chain kinase activity, Actin binding

NME6	NME/NM23 nucleoside diphosphate kinase 6	Nucleoside diphosphate kinase activity
NT5C	5', 3'-Nucleotidase, cytosolic	Nucleotidase activity
NT5C1A	5'-Nucleotidase, cytosolic IA	Nucleotide binding, 5'-nucleotidase activity
PFKFB1	6-Phosphofructo-2-kinase/Fructose-2,6-biphosphatase 1	6-phosphofructo-2-kinase activity
PGAM2	Phosphoglycerate mutase 2 (muscle)	Phosphoglycerate mutase activity
PIK3R4	Phosphoinositide-3-kinase, regulatory subunit 4	Protein kinase activity
PPM1K	Protein phosphatase, Mg2+/Mn2+ dependent, 1K	Catalytic activity
PPM1L	Protein phosphatase, Mg2+/Mn2+ dependent, 1L	Catalytic activity
PPM1M	Protein phosphatase, Mg2+/Mn2+ dependent, 1M	Catalytic activity
PPM2C	Pyruvate dehyrogenase phosphatase catalytic subunit 1	Catalytic activity
PPP1R8	Protein phosphatase 1, regulatory subunit 8	Protein serine/threonine phosphatase inhibitor activity
PPP2R1A	Protein phosphatase 2, regulatory subunit A, alpha	Antigen binding, Protein phosphatase type 2A regulator activity
PPP2R2A	Protein phosphatase 2, regulatory subunit B, alpha	Protein serine/threonine phosphatase activity
PPP2R4	Protein phosphatase 2A activator, regulatory subunit 4	Receptor binding, Protein tyrosine phosphatase activator activity
PPP2R5A	Protein phosphatase 2, regulatory subunit B', alpha	Protein binding
PPP4C	Protein phosphatase 4, catalytic subunit	NF-kappaB-inducing kinase activity
PRKAR1B	Protein kinase, CAMP-dependent, regulatory, type I, Beta	cAMP-dependent protein kinase regulator activity
PRKCH	Protein kinase C, Eta	Protein kinase activity
PTEN	Phosphatase and tensin homolog	Phosphatidylinositol-3-phosphatase activity
PTPDC1	Protein tyrosine phosphatase domain containing 1	Protein tyrosine phosphatase activity
PTPRJ	Protein tyrosine phosphatase, receptor type, J	Protein tyrosine phosphatase activity
RAGE	MAPK/MAK/MRK overlapping kinase	Protein serine/threonine kinase activity
RELA	V-Rel avian reticuloendotheliosis viral oncogene homolog A	DNA binding, Chromatin binding
RIOK3	RIO kinase 3	Catalytic activity
STK23	SRSF Protein Kinase 3	Protein kinase activity
ZAP70	Zeta-Chain (TCR) associated protein kinase 70kDa	Phosphotyrosine binding, Protein binding



Appendix figure 1: GO molecular funtion of p65 (NF-kB) nuclear import candidate genes.

Appendix table 2: Statistical quantification of candidate genes (< 2STDV): Cellular functions, biological processes and molecular functions

Gene Name	Cellular function/component(s)	Biological process /GO ID	Molecular function/GO ID
PPP6C	Catalytic subunit of protein phosphatase 6 (PP6). PP6 is a component of a signaling pathway regulating cell cycle progression in response to IL2 receptor stimulation. cytosol (GO:0005829)	G1/S transition of mitotic cell cycle GO:0000082) Protein dephosphorylation (GO:0006470)	Phosphoprotein phosphatase activity (GO:0004721) Protein serine/threonine phosphatase activity (GO:0004722) Protein binding (GO:0005515) Metal ion binding (GO:0046872)
EPHA4	Receptor tyrosine kinase, which binds membrane-bound ephrin family ligands residing on adjacent cells, leading to contact-dependent bidirectional signaling into neighboring cells. Cytoplasm (GO:0005737) Mitochondrial outer membrane (GO:0005741) Endoplasmic reticulum (GO:0005783) Golgi apparatus (GO:0005794) Integral to plasma membrane (GO:0005887)	Cell adhesion (GO:0007155) Axon guidance (GO:0007411) Adult walking behaviour (GO:0007628) Motor neuron axon guidance (GO:0008045) glial cell migration (GO:0008347)	Protein kinase activity (GO:0004672) GPI-linked ephrin receptor activity (GO:0005004) transmembrane-ephrin receptor activity (GO:0005005) Protein binding (GO:0005515) ATP binding (GO:0005524)
PAK7	Serine/threonine protein kinase that plays a role in a variety of different signaling pathways including Cytoskeleton regulation, cell migration, proliferation or cell survival. Nucleus (GO:0005634) Mitochondrion (GO:0005739)	Apoptotic process (GO:0006915) Cytoskeleton organization (GO:0007010) Signal transduction (GO:0007165) Cell proliferation (GO:0008283) Cell growth (GO:0016049)	Protein serine/threonine kinase activity (GO:0004674) ATP binding (GO:0005524)
CDKL2	This gene product is a member of a large family of CDC2-related serine/threonine protein kinases. It accumulates primarily in the cytoplasm, with lower levels in the nucleus. Nucleus (GO:0005634) Cytoplasm (GO:0005737) Centrosome (GO:0005813)	Protein phosphorylation (GO:0006468) Signal transduction (GO:0007165) Sex differentiation (GO:0007548)	Protein kinase activity (GO:0004672) cyclin-dependent protein kinase activity (GO:0004693) ATP binding (GO:0005524)
MARCH6	This gene encodes a member of a family of membrane-associated E3 ubiquitin ligases containing RING-CH-type zinc finger motifs. E3 ubiquitin-protein ligase that promotes ubiquitination of DIO2, leading to its degradation. Integral to endoplasmic reticulum membrane (GO:0030176)	Protein ubiquitination (GO:0016567) Protein K48-linked ubiquitination (GO:0070936)	Ubiquitin-protein ligase activity (GO:0004842 Zinc ion binding (GO:0008270) Enzyme binding (GO:0019899) ubiquitin conjugating enzyme binding (GO:0031624)
PTPN9	Protein-tyrosine phosphatase that could participate in the transfer of hydrophobic ligands or in functions of the Golgi apparatus.	Protein dephosphorylation (GO:0006470)	Protein tyrosine phosphatase activity (GO:0004725) Non-membrane spanning protein tyrosine phosphatase activity

	cytoplasmic part (GO:0044444)	peptidyl-tyrosine dephosphorylation (GO:0035335)	(GO:0004726) Protein binding (GO:0005515)
CDK17	May play a role in terminally differentiated neurons. Has a Ser/Thr-phosphorylating activity for histone H1 (By similarity).	Protein phosphorylation (GO:0006468)	Protein kinase activity (GO:0004672) Protein serine/threonine kinase activity (GO:0004674) Cyclin-dependent protein kinase activity (GO:0004693) ATP binding (GO:0005524
NLN	Hydrolyzes oligopeptides such as neurotensin, bradykinin and dynorphin A (By similarity). Mitochondrion (GO:0005739) Mitochondrial intermembrane space (GO:0005758)	Proteolysis (GO:0006508)	metalloendopeptidase activity (GO:0004222) Metal ion binding (GO:0046872)
MAPKAPK5	Tumor suppressor serine/threonine-protein kinase involved in mTORC1 signalling and post-transcriptional regulation. Nucleus (GO:0005634) NOT nucleolus (GO:0005730) Cytoplasm (GO:0005737)	MAPK cascade (GO:0000165) Activation of MAPK activity (GO:0000187) Regulation of translation (GO:0006417) Protein phosphorylation (GO:0006468) Response to stress (GO:0006950)	p53 binding (GO:0002039) Protein kinase activity (GO:0004672) Protein serine/threonine kinase activity (GO:0004674) MAP kinase kinase activity (GO:0004708) Protein binding (GO:0005515)
PON1	Hydrolyzes the toxic metabolites of a variety of organophosphorus insecticides. Extracellular region (GO:0005576) Extracellular space (GO:0005615) Microsome (GO:0005792) High-density lipoprotein particle (GO:0034364) Spherical high-density lipoprotein particle (GO:0034366)	Lipid metabolic process (GO:0006629) Cholesterol metabolic process (GO:0008203) Response to external stimulus (GO:0009605) Response to toxin (GO:0009636) Positive regulation of cholesterol efflux (GO:0010875)	Aryldialkylphosphatase activity (GO:0004063) Arylesterase activity (GO:0004064) Calcium ion binding (GO:0005509) Phospholipid binding (GO:0005543) Protein homodimerization activity (GO:0042803)
MAPK6	Atypical MAPK protein. Phosphorylates microtubule-associated protein 2 (MAP2) and MAPKAPK5. Nucleus (GO:0005634)	MAPK cascade (GO:0000165) Protein phosphorylation (GO:0006468)	Protein serine/threonine kinase activity (GO:0004674) MAP kinase activity (GO:0004707) Protein binding (GO:0005515)

	Cytoplasm (GO:0005737)	Cell cycle (GO:0007049) Signal transduction (GO:0007165)	ATP binding (GO:0005524) Protein kinase binding (GO:0019901)
PPP1R12A	Key regulator of protein phosphatase 1C (PPP1C). Mediates binding to myosin. As part of the PPP1C complex, involved in dephosphorylation of PLK1. Capable of inhibiting HIF1AN-dependent suppression of HIF1A activity. Kinetochore (GO:0000776) Centrosome (GO:0005813) Contractile fiber (GO:0043292) PTW/PP1 phosphatase complex (GO:0072357)	Protein dephosphorylation (GO:0006470) Mitosis (GO:0007067) Signal transduction (GO:0007165) Regulation of cell adhesion (GO:0030155) Regulation of myosin-light-chain-Phosphatase activity (GO:0035507)	Phosphoprotein phosphatase activity (GO:0004721) Enzyme inhibitor activity (GO:0004857) Signal transducer activity (GO:0004871) Protein binding (GO:0005515) Phosphatase regulator activity (GO:0019208
PPP2R2D	B regulatory subunit of protein phosphatase 2A (PP2A) that plays a key role in cell cycle by controlling mitosis entry and exit. Protein phosphatase type 2A complex (GO:0000159) Nucleoplasm (GO:0005654) Cytoplasm (GO:0005737) Cytosol (GO:0005829)	Mitosis (GO:0007067) Signal transduction (GO:0007165) Exit from mitosis (GO:0010458) Cell division (GO:0051301)	Phosphatidate phosphatase activity (GO:0008195) Phosphatidate phosphatase activity (GO:0008195)
ERBB4	Tyrosine-protein kinase that plays an essential role as cell surface receptor for neuregulins and EGF family members and regulates development of the heart, the central nervous system and the mammary gland, gene transcription, cell proliferation, differentiation, migration and apoptosis. Extracellular region (GO:0005576) Nucleus (GO:0005634) Nucleoplasm (GO:0005654) NOT nucleolus (GO:0005730) Cytoplasm (GO:0005737)	Neural crest cell migration (GO:0001755) Positive regulation of protein phosphorylation (GO:0001934) Transcription, DNA-dependent (GO:0006351) Apoptotic process (GO:0006915) Signal transduction (GO:0007165)	Protein tyrosine kinase activity (GO:0004713) Transmembrane receptor protein tyrosine kinase activity (GO:0004714) Receptor signaling protein tyrosine kinase activity (GO:0004716) Epidermal growth factor receptor binding (GO:0005154) Protein binding (GO:0005515)

Appendix Ttable 3: Statistical quantification of candidate genes (> 2STDV); Cellular functions, biological processes and molecular functions

	Cellular function	Biological process /GO ID	Molecular function/GO ID
NEK4	Seems to act exclusively upon threonine residues (By similarity). Nucleus (GO:0005634) NOT nucleolus (GO:0005730)	Protein phosphorylation (GO:0006468) Mitosis (GO:0007067) Cell division (GO:0051301)	Protein serine/threonine kinase activity (GO:0004674) ATP binding (GO:0005524) Metal ion binding (GO:0046872)
CAMKK2	Calcium/calmodulin-dependent protein kinase belonging to a proposed calcium-triggered signaling cascade involved in a number of cellular processes. Intracellular (GO:0005622) Cytoplasm (GO:0005737)	MAPK cascade (GO:0000165) Protein phosphorylation (GO:0006468) Calcium-mediated signaling (GO:0019722) Regulation of protein kinase activity (GO:0045859) Positive regulation of transcription, DNA-dependent (GO:0045893) Protein autophosphorylation (GO:0046777)	Protein serine/threonine kinase activity (GO:0004674) Calmodulin dependent protein kinase activity (GO:0004683) Protein tyrosine kinase activity (GO:0004713) Calcium ion binding (GO:0005509) Calmodulin binding GO:0005516)
CELF4	RNA-binding protein implicated in the regulation of pre-mRNA alternative splicing. Mediates exon inclusion and/or exclusion in pre-mRNA that are subject to tissue-specific and developmentally regulated alternative splicing. Nucleus (GO:0005634) NOT nucleolus (GO:0005730) Cytoplasm (GO:0005737)	Alternative mRNA splicing, via spliceosome (GO:0000380) Regulation of alternative mRNA splicing, via spliceosome (GO:0000381) mRNA splice site selection (GO:0006376) Germ cell development (GO:0007281) Embryo development (GO:0009790) Positive regulation of mRNA splicing, via spliceosome (GO:0048026) Negative regulation of excitatory postsynaptic membrane potential	Nucleotide binding (GO:0000166) Translation repressor activity, nucleic acid binding (GO:0000900) mRNA binding (GO:0003729) BRE binding (GO:0042835

		(GO:0090394)	
EPHB1	Receptor tyrosine kinase which binds promiscuously transmembrane ephrin-B family ligands residing on adjacent cells, leading to contact-dependent bidirectional signaling into neighboring cells. Cytoplasm (GO:0005737) Integral to plasma membrane (GO:0005887) Membrane (GO:0016020) Integral to membrane (GO:0016021) Axon (GO:0030424)	Cell-substrate adhesion (GO:0031589) Regulation of JNK cascade (GO:0046328) Protein autophosphorylation (GO:0046777) Ephrin receptor signaling pathway (GO:0048013) Cell chemotaxis (GO:0060326) Regulation of ERK1 and ERK2 cascade (GO:0070372) Angiogenesis (GO:0001525)	Protein tyrosine kinase activity (GO:0004713) Ephrin receptor activity (GO:0005003) Transmembrane-ephrin receptor activity (GO:0005005) Protein binding (GO:0005515) ATP binding (GO:0005524)
RFFL	Has E3 ubiquitin protein ligase activity. Regulates the levels of CASP8 and CASP10 by targeting them for proteasomal degradation. Cytoplasm (GO:0005737) Membrane (GO:0016020) Cytoplasmic membrane-bounded vesicle (GO:0016023)	Intracellular protein transport (GO:0006886) Apoptotic process (GO:0006886) Protein ubiquitination (GO:0016567)	Zinc ion binding (GO:0008270) Ligase activity (GO:0016874)
CDK12	Cyclin-dependent kinase which displays CTD kinase activity and is required for RNA splicing. Cyclin K-CDK12 complex (GO:0002944) Nucleus (GO:0005634) Nucleolus (GO:0005730) Nuclear speck (GO:0016607) Nuclear cyclin-dependent protein kinase holoenzyme complex (GO:0019908)	mRNA processing (GO:0006397) RNA splicing (GO:0008380) Regulation of MAP kinase activity (GO:0043405) Protein autophosphorylation (GO:0046777) Phosphorylation of RNA polymerase II C-terminal domain (GO:0070816)	Protein kinase activity (GO:0004672) Cyclin-dependent protein kinase activity (GO:0004693) Protein binding (GO:0005515) ATP binding (GO:0005524)) RNA polymerase II carboxy-terminal domain kinase activity (GO:0008353)
PTP4A1	Protein tyrosine phosphatase which stimulates progression from G1 into S phase during mitosis. Nucleus (GO:0005634)	Cell cycle (GO:0007049) Multicellular organismal development (GO:0007275)	Protein tyrosine phosphatase activity (GO:0004725) Phosphatase activity (GO:0016791)

	Cytoplasm (GO:0005737) Early endosome (GO:0005769) Endoplasmic reticulum (GO:0005783) Spindle (GO:0005819)	Positive regulation of cell migration (GO:0030335) Peptidyl-tyrosine dephosphorylation (GO:0035335)	
PRKCZ	Calcium- and diacylglycerol-independent serine/threonine-protein kinase that functions in phosphatidylinositol 3-kinase (PI3K) pathway and mitogen-activated protein (MAP) kinase cascade, and is involved in NF-kappa-B activation, mitogenic signaling, cell proliferation, cell polarity, inflammatory response and maintenance of long-term potentiation (LTP). Intracellular (GO:0005622) Membrane fraction (GO:0005624) Soluble fraction (GO:0005625) Nucleus (GO:0005634) Nuclear envelope (GO:0005635)	Microtubule cytoskeleton organization (GO:0000226) Positive regulation of cell-matrix adhesion (GO:0001954) Protein phosphorylation (GO:0006468) Inflammatory response (GO:0006954) Signal transduction (GO:0007165)	Protein kinase activity (GO:0004672) Protein serine/threonine Kinase activity (GO:0004674) Protein kinase C activity (GO:0004697) Protein binding (GO:0005515) ATP binding (GO:0005524) Insulin receptor substrate binding (GO:0043560)
KCNK13	Potassium channel displaying weak inward rectification in symmetrical K(+) solution (By similarity). Plasma membrane (GO:0005886) Integral to membrane (GO:0016021)	Synaptic transmission (GO:0007268) Potassium ion transmembrane transport (GO:0071805)	Voltage-gated ion channel activity (GO:0005244) Potassium channel activity (GO:0005267)
MOK	Able to phosphorylate several exogenous substrates and to undergo autophosphorylation (By similarity). Cytoplasm (GO:0005737)	Protein phosphorylation (GO:0006468) Signal transduction GO:0007165)	Protein kinase activity (GO:0004672) Protein serine/threonine kinase activity (GO:0004674) Cyclin-dependent protein Kinase activity (GO:0004693) ATP binding (GO:0005524)
ZNRF2	May play a role in the establishment and maintenance of neuronal transmission and plasticity via its ubiquitin ligase activity. Lysosomal membrane (GO:0005765) Plasma membrane (GO:0005886)	Protein ubiquitination (GO:0016567)	Zinc ion binding (GO:0008270) Ligase activity (GO:0016874)

	Endosome membrane (GO:0010008) Cell junction (GO:0030054) Presynaptic membrane (GO:0042734)		
TSSK6	Required for sperm production and function. Plays a role in DNA condensation during postmeiotic chromatin remodeling (By similarity).	Protein phosphorylation (GO:0006468) Multicellular organismal development (GO:0007275) Sperm chromatin condensation GO:0035092)	Magnesium ion binding (GO:0000287) Protein serine/threonine kinase activity (GO:0004674) Protein binding (GO:0005515) ATP binding (GO:0005524)
RNF24	May play a role in TRPCs intracellular trafficking. Golgi membrane (GO:0000139) Integral to membrane (GO:0016021)		Zinc ion binding (GO:0008270)
STK4	Stress-activated, pro-apoptotic kinase which, following caspase-cleavage, enters the nucleus and induces chromatin condensation followed by internucleosomal DNA fragmentation. Nucleus GO:0005634 Cytoplasm GO:0005737 Cytosol GO:0005829	Cell morphogenesis GO:0000902 Patterning of blood vessels GO:0001569 Neural tube formation GO:0001841 Endocardium development GO:0003157 Protein phosphorylation GO:0006468	Magnesium ion binding GO:0000287 Protein serine/threonine kinase activit GO:0004674 Protein binding GO:0005515 ATP binding GO:0005524 Transcription factor binding GO:0008134
SH3RF2	Inhibits PPP1CA phosphatase activity. May be a E3 ubiquitin-protein ligase (Potential). May play a role in cardiac function.	negative regulation of phosphatase activity (GO:0010923) protein ubiquitination (GO:0016567)	Protein phosphatase inhibitor activity (GO:0004864) Protein binding (GO:0005515) Protein phosphatase 1 binding GO:0008157) Zinc ion binding (GO:0008270) Ligase activity (GO:0016874) Phosphatase binding (GO:0019902)
C2	Component C2 which is part of the classical pathway of the complement system is cleaved by activated factor C1 into two fragments: C2b and C2a. C2a, a serine protease, then combines with complement factor 4b to generate the C3 or C5 convertase. Extracellular region GO:0005576 Extracellular space GO:0005615	Proteolysis GO:0006508 Complement activation GO:0006956 Complement activation, classical pathway GO:0006958 Regulation of complement activation	Serine-type endopeptidase activity GO:0004252

		GO:0030449 Innate immune response GO:0045087	
UCK2	Phosphorylates uridine and cytidine to uridine monophosphate and cytidine monophosphate. Cellular_component GO:0005575 Microsome GO:0005792 Cytosol GO:0005829 Intracellular membrane-bounded organelle GO:0043231	Pyrimidine nucleobase metabolic process GO:0006206 Feeding behavior GO:0007631 Nucleotide biosynthetic process GO:0009165 Pyrimidine nucleoside salvage GO:0043097 UMP salvage GO:0044206	Uridine kinase activity GO:0004849 Protein binding GO:0005515 ATP binding GO:0005524 Kinase activity GO:0016301 Phosphotransferase activity, alcohol group as acceptor GO:0016773 Nucleoside kinase activity GO:0019206
VRK1	Serine/threonine kinase involved in Golgi disassembly during the cell cycle: following phosphorylation by PLK3 during mitosis, required to induce Golgi fragmentation. Nucleus GO:0005634 NOT nucleolus GO:0005730 Cytoplasm GO:0005737 Golgi stack GO:0005795 Spindle GO:0005819	Protein phosphorylation GO:0006468 Mitosis GO:0007067 Protein autophosphorylation GO:0046777 Cell division GO:0051301 Golgi disassembly GO:0090166	Protein serine/threonine kinase activity GO:0004674 Protein binding GO:0005515 ATP binding GO:0005524 Identical protein binding GO:0042802

Appendix table 4: Statistical quantification of candidate genes (> 3STDV): Cellular functions, biological processes and molecular functions

Gene Name	Cellular function	Biological process /GO ID	Molecular function/GO ID
ST6GALNAC4	Involved in the biosynthesis of ganglioside GD1A from GM1B. Transfers CMP-NeuAc with an alpha-2,6-linkage to GalNAc residue on NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc of glycoproteins and glycolipids. Golgi membrane GO:0000139 Soluble fraction GO:0005625 Nucleus GO:0005634 Golgi apparatus GO:0005794 Integral to Golgi membrane GO:0030173	protein glycosylation GO:0006486 glycolipid metabolic process GO:0006664 O-glycan processing GO:0016266 Post-translational protein modification GO:0043687 Cellular protein metabolic process GO:0044267	sialyltransferase activity GO:0008373 (alpha-N-acetylneuraminyl-2,3-beta-galactosyl-1,3)-N-acetyl-galactosaminide 6-alpha-sialyltransferase activity GO:0047290

ARIH2	E3 ubiquitin-protein ligase mediating 'Lys-48'-and 'Lys-63'-linked polyubiquitination and subsequent proteasomal degradation of modified proteins. May play a role in myelopoiesis. Nucleus GO:0005634 Cytoplasm GO:0005737	Protein polyubiquitination GO:0000209 ubiquitin-dependent protein catabolic process GO:0006511 Multicellular organismal development GO:0007275 Protein ubiquitination involved in ubiquitin-dependent protein catabolic process GO:0042787 Developmental cell growth GO:0048588	Nucleic acid binding GO:0003676 Ubiquitin-protein ligase activity GO:0004842 Protein binding GO:0005515 Zinc ion binding GO:0008270
ZACN	Zinc-activated ligand-gated ion channel. Membrane fraction GO:0005624 Plasma membrane GO:0005886 Integral to membrane GO:0016021 Postsynaptic membrane GO:0045211	Response to zinc ion GO:0010043	Receptor activity GO:0004872 Extracellular ligand-gated ion channel activity GO:0005230 Zinc ion binding GO:0008270
AIFM3	Induces apoptosis through a caspase dependent pathway. Reduces mitochondrial membrane potential. NOT nucleus GO:0005634 Cytoplasm GO:0005737 Mitochondrion GO:0005739 Mitochondrial inner membrane GO:0005743 Endoplasmic reticulum GO:0005783	Induction of apoptosis GO:0006917 Activation of cysteine-type endopeptidase activity involved in apoptotic process by cytochrome c GO:0008635 Electron transport chain GO:0022900 Cell redox homeostasis GO:0045454 Mitochondrial depolarization GO:0051882	Cysteine-type endopeptidase activator activity involved in apoptotic process GO:0008656 Oxidoreductase activity GO:0016491 Metal ion binding GO:0046872 Flavin adenine dinucleotide binding GO:0050660 2 iron, 2 sulfur cluster binding GO:0051537
PKIB	Extremely potent competitive inhibitor of cAMP-dependent protein kinase activity, this protein interacts with the catalytic subunit of the enzyme after the cAMP-induced dissociation of its regulatory chains (By similarity)		cAMP-dependent protein kinase inhibitor activity GO:0004862
SRPK3	Serine/arginine-rich protein-specific kinase which specifically phosphorylates its substrates at serine residues located in regions rich in arginine/serine dipeptides, known as	protein phosphorylation GO:0006468 skeletal muscle tissue development	protein serine/threonine kinase activity GO:0004674 ATP binding GO:0005524

	RS domains. cellular_component GO:0005575	GO:0007519 cell differentiation GO:0030154 muscle tissue development GO:0060537	
KLK7	May catalyze the degradation of intercellular cohesive structures in the cornified layer of the skin in the continuous shedding of cells from the skin surface. Specific for amino acid residues with aromatic side chains in the P1 position. Extracellular region GO:0005576 Epidermal lamellar body GO:0097209	Proteolysis GO:0006508 epidermis development GO:0008544	serine-type endopeptidase activity GO:0004252 serine-type peptidase activity GO:0008236
ZAP70	Tyrosine kinase that plays an essential role in regulation of the adaptive immune response. Regulates motility, adhesion and cytokine expression of mature T-cells, as well as thymocyte development. immunological synapse GO:0001772 Intracellular GO:0005622 Cytoplasm GO:0005737 Cytosol GO:0005829 plasma membrane GO:0005886	adaptive immune response GO:0002250 protein phosphorylation GO:0006468 immune response GO:0006955 intracellular protein kinase cascade GO:0007243 peptidyl-tyrosine phosphorylation GO:0018108	phosphotyrosine binding GO:0001784 protein kinase activity GO:0004672 protein tyrosine kinase activity GO:0004713 non-membrane spanning protein tyrosine kinase activity GO:0004715 protein binding GO:0005515
RNF122	May induce necrosis and apoptosis. May play a role in cell viability. endoplasmic reticulum GO:0005783 Golgi apparatus GO:0005794 integral to membrane GO:0016021		zinc ion binding GO:0008270
ME1	soluble fraction GO:0005625 Mitochondrion GO:0005739 Cytosol GO:0005829	carbohydrate metabolic processGO:0005975 malate metabolic processGO:0006108 NADP biosynthetic processGO:0006741 response to hormone stimulusGO:0009725 response to carbohydrate stimulusGO:0009743	malic enzyme activity GO:0004470 malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity GO:0004473 electron carrier activity GO:0009055 malate dehydrogenase (oxaloacetate-decarboxylating) activity GO:0016619 manganese ion binding GO:0030145

MAPK9	Serine/threonine-protein kinase involved in various processes such as cell proliferation, differentiation, migration, transformation and programmed cell death. Nucleoplasm GO:0005654 Mitochondrion GO:0005739 Cytosol GO:0005829 cytosolic part GO:0044445	release of cytochrome c from mitochondria GO:0001836 positive regulation of protein phosphorylation GO:0001934 toll-like receptor signaling pathway GO:0002224 MyD88-dependent toll-like receptor signaling pathway GO:0002755 MyD88-independent toll-like receptor signaling pathway GO:0002756 stress-activated MAPK cascade GO:0051403	protein kinase activity GO:0004672 JUN kinase activity GO:0004705 MAP kinase activity GO:0004707 protein binding GO:0005515 ATP binding GO:0005524
KPNA6	Functions in nuclear protein import as an adapter protein for nuclear receptor KPNB1. Binds specifically and directly to substrates containing either a simple or bipartite NLS motif. Nucleus GO:0005634 Cytoplasm GO:0005737	protein import into nucleus GO:0006606 NLS-bearing substrate import into nucleus GO:0006607	protein binding GO:0005515 protein transporter activity GO:0008565
HIPK1	Serine/threonine-protein kinase involved in transcription regulation and TNF-mediated cellular apoptosis. Plays a role as a corepressor for homeodomain transcription factors. Nucleus GO:0005634 Cytoplasm GO:0005737 PML body GO:0016605 Nuclear speck GO:0016607	Eye development GO:0001654 Transcription, DNA-dependent GO:0006351 Regulation of transcription, DNA-dependent GO:0006355 Apoptotic process GO:0006915 Smoothened signaling pathwayGO:0007224	DNA binding GO:0003677 Protein serine/threonine kinase activity GO:0004674 Protein binding GO:0005515 ATP binding GO:0005524
AZU1	This is a neutrophil granule-derived antibacterial and monocyte- and fibroblast-specific chemotactic glycoprotein. Binds heparin. Extracellular region GO:0005576 Azurophil granule GO:0042582	Microglial cell activation GO:0001774 Proteolysis GO:0006508 Inflammatory response GO:0006954 Protein kinase C-activating G-protein	Serine-type endopeptidase activity GO:0004252 Heparin binding GO:0008201 Toxin binding GO:0015643

		coupled receptor signaling pathway GO:0007205 Glial cell migration GO:0008347	
RFPL3			Zinc ion binding GO:0008270
DUSP1	Dual specificity phosphatase that dephosphorylates MAP kinase MAPK1/ERK2 on both 'Thr-183' and 'Tyr-185', regulating its activity during the meiotic cell cycle (By similarity). Soluble fraction GO:0005625 Nucleus GO:0005634 Nucleoplasm GO:0005654	Inactivation of MAPK activity GO:0000188 Endoderm formation GO:0001706 Protein dephosphorylation GO:0006470 Response to oxidative stress GO:0006979 Cell cycle GO:0007049	Non-membrane spanning protein tyrosine phosphatase activity GO:0004726 Protein binding GO:0005515 Protein tyrosine/threonine phosphatase activity GO:0008330 MAP kinase tyrosine/serine/threonine phosphatase activity GO:0017017
WDSUB1	Ubiquitin ligase complex GO:0000151		Ubiquitin-protein ligase activity GO:0004842
ADCY3	Mediates odorant detection (possibly) via modulation of intracellular cAMP concentration Intracellular GO:0005622 Membrane fraction GO:0005624 Nucleus GO:0005634 Cytoplasm GO:0005737 Plasma membrane GO:0005886	Energy reserve metabolic process GO:0006112 cAMP biosynthetic process GO:0006171 Water transport GO:0006833 Female meiosis GO:0007143 Signal transduction GO:0007165	Nucleotide binding GO:0000166 Adenylate cyclase activity GO:0004016 Protein binding GO:0005515 Calmodulin binding GO:0005516 ATP binding GO:0005524

Appendix table 4: Statistical quantification of candidate genes (> 4STDV): Cellular functions, biological processes and molecular functions

Gene Name	Cellular function	Biological process /GO ID	Molecular function/GO ID
UQCRC1	This is a component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is part of th mitochondrial respiratory chain. This protein may mediate formation of the complex between cytochromes c and c1. Mitochondrion GO:0005739 Mitochondrial inner membrane GO:0005743 Mitochondrial respiratory chain GO:0005746 Mitochondrial respiratory chain complex III GO:0005750	Oxidative phosphorylation GO:0006119 Mitochondrial electron transport, ubiquinol to cytochrome c GO:0006122 Proteolysis GO:0006508 Aerobic respiration GO:0009060 Response to activity GO:0014823 Respiratory electron transport chain GO:0022904 Response to alkaloid GO:0043279 Small molecule metabolic process GO:0044281 Oxidation-reduction process GO:0055114	Metalloendopeptidase activity GO:0004222 Protein binding GO:0005515 Ubiquinol-cytochrome-c reductase activity GO:0008121 Zinc ion binding GO:0008270 Protein complex binding GO:0032403 Metal ion binding GO:0046872
GUCY2F	Probably plays a specific functional role in the rods and/or cones of photoreceptors. It may be the enzyme involved in the resynthesis of cGMP required for recovery of the dark state after phototransduction. Membrane fraction GO:0005624 Nuclear outer membrane GO:0005640 Integral to plasma membrane GO:0005887	cGMP biosynthetic process GO:0006182 Receptor guanylyl cyclase signaling pathway GO:0007168 Visual perception GO:0007601 Intracellular signal transduction GO:0035556	Guanylate cyclase activity GO:0004383 Protein kinase activity GO:0004672 Receptor activity GO:0004872 Protein binding GO:0005515 ATP binding GO:0005524 GTP binding GO:0005525 Protein homodimerization activity GO:0042803 Protein heterodimerization activityGO:0046982
BAHD1	Heterochromatin protein that acts as a transcription repressor and has the ability to promote the formation of large heterochromatic domains. Chromatin silencing complex GO:0005677	Transcription, DNA-dependentGO:0006351 Heterochromatin	DNA binding GO:0003677 Chromatin binding GO:0003682 Protein binding GO:0005515

	Chromosome GO:0005694	assemblyGO:0031507 Negative regulation of transcription, DNA-dependentGO:0045892	
INPP5B	Hydrolyzes phosphatidylinositol 4,5-bisphosphate (PtIns(4,5)P2) and the signaling molecule phosphatidylinositol 1,4,5-trisphosphate (PtIns(1,4,5)P3), and thereby modulates cellular signaling events. Cytoplasm GO:0005737 Endoplasmic reticulum-Golgi intermediate compartment GO:0005793 Golgi apparatus GO:0005794 Cytosol GO:0005829 Microtubule cytoskeleton GO:0015630	In utero embryonic development GO:0001701 Small GTPase mediated signal transduction GO:0007264 Spermatogenesis GO:0007283 Sperm motility GO:0030317 Phosphatidylinositol phosphorylation GO:0046854	Phosphatidylinositol-4,5-bisphosphate 5-phosphatase activity GO:0004439 Protein binding GO:0005515 Metal ion binding GO:0046872 Inositol-1, 4, 5-trisphosphate 5-phosphatase activity GO:0052658
LATS2	Negative regulator of YAP1 in the Hippo signaling pathway that plays a pivotal role in organ size control and tumor suppression by restricting proliferation and promoting apoptosis. The core of this pathway is composed of a kinase cascade wherein STK3/MST2 and STK4/MST1, Spindle pole GO:0000922 Nucleus GO:0005634 Microtubule organizing center GO:0005815 Cytosol GO:0005829	G1/S transition of mitotic cell cycle GO:0000082 Protein phosphorylation GO:0006468 Mitosis GO:0007067 Intracellular protein kinase cascade GO:0007243 Hormone-mediated signaling pathway GO:0009755	Protein serine/threonine kinase activity GO:0004674 Protein binding GO:0005515 ATP binding GO:0005524 Metal ion binding GO:0046872
CYP4A22	Catalyzes the omega- and (omega-1)-hydroxylation of various fatty acids such as laurate and palmitate. Shows no activity towards arachidonic acid and prostaglandin A1. Endoplasmic reticulum membrane GO:0005789 Microsome GO:0005792		Monooxygenase activity GO:0004497 Iron ion binding GO:0005506 Electron carrier activity GO:0009055 Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen GO:0016705 Alkane 1-monooxygenase activity GO:0018685
CNOT7	Ubiquitous transcription factor required for a diverse set of processes. It is a component of the CCR4 complex involved in the control of gene expression (By similarity). Cytoplasmic mRNA processing body GO:0000932	Nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay GO:0000288	Nucleic acid binding GO:0003676 Sequence-specific DNA binding transcription factor activity GO:0003700

	Nucleus GO:0005634 Cytosol GO:0005829 CCR4-NOT complex GO:0030014	Nuclear-transcribed mRNA poly(A) tail shortening GO:0000289 Deadenylation-dependent decapping of nuclear-transcribed mRNAGO:0000290 Carbohydrate metabolic process GO:0005975 Transcription, DNA-dependent GO:0006351	Poly (A)-specific ribonuclease activity GO:0004535 Signal transducer activity GO:0004871 Protein binding GO:0005515
TLK1	Rapidly and transiently inhibited by phosphorylation following the generation of DNA double-stranded breaks during S-phase. This is cell cycle checkpoint and ATM-pathway dependent and appears to regulate processes involved in chromatin assembly. Nucleus GO:0005634	Regulation of chromatin assembly or disassembly GO:0001672 Protein phosphorylation GO:0006468 Intracellular protein transport GO:0006886 Response to DNA damage stimulus GO:0006974 Cell cycle GO:0007049	Protein serine/threonine kinase activity GO:0004674 ATP binding GO:0005524 kinase activity GO:0016301
ACOT6	Cytosol GO:0005829		carboxylesterase activity GO:0004091
NPR2	Receptor for the C-type natriuretic peptide NPPC/CNP hormone. Has guanylate cyclase activity upon binding of its ligand. May play a role in the regulation of skeletal growth. Intracellular GO:0005622 Membrane fraction GO:0005624 Plasma membrane GO:0005886 Integral to membrane GO:0016021	Ossification GO:0001503 cGMP biosynthetic process GO:0006182 Signal transduction GO:0007165 Cell surface receptor signaling pathway GO:0007166 Receptor guanylyl cyclase signaling pathway GO:0007168	Guanylate cyclase activity GO:0004383 Protein kinase activity GO:0004672 Receptor activity GO:0004872 Transmembrane signaling receptor activity GO:0004888 ATP binding GO:0005524
ADAMTS1	Cleaves aggrecan, a cartilage proteoglycan, and may be involved in its turnover (By similarity). Has angiogenic inhibitor activity. Active metalloprotease, which may be associated with various inflammatory processes as well as development of cancer	Ovulation from ovarian follicle GO:0001542 Kidney development GO:0001822	Metalloendopeptidase activity GO:0004222 Protein binding GO:0005515 Heparin binding GO:0008201

	cachexia. May play a critical role in follicular rupture. Basement membrane GO:0005604 Extracellular matrix GO:0031012 Cytoplasmic vesicle GO:0031410	Proteolysis GO:0006508 Integrin-mediated signaling pathway GO:0007229 Negative regulation of cell proliferation GO:0008285	Metallopeptidase activity GO:0008237 Zinc ion binding GO:0008270
EPHB2	Receptor tyrosine kinase which binds promiscuously transmembrane ephrin-B family ligands residing on adjacent cells, leading to contact-dependent bidirectional signaling into neighboring cells. Integral to plasma membrane GO:0005887 Integral to membrane GO:0016021 Axon GO:0030424 Dendrite GO:0030425 Neuronal cell body GO:0043025	Cell morphogenesis GO:0000902 Angiogenesis GO:0001525 Urogenital system development GO:0001655 Protein phosphorylation GO:0006468 Transmembrane receptor protein tyrosine kinase signaling pathway GO:0007169	Protein tyrosine kinase activity GO:0004713 Receptor activity GO:0004872 Ephrin receptor activity GO:0005003 Transmembrane-ephrin receptor activity GO:0005005 Receptor binding GO:0005102
ECEL1	May contribute to the degradation of peptide hormones and be involved in the inactivation of neuronal Peptides. Intracellular GO:0005622 Integral to plasma membrane GO:0005887	Proteolysis GO:0006508 Neuropeptide signaling pathway GO:0007218 Intracellular signal transduction GO:0035556	Metalloendopeptidase activity GO:0004222 Metallopeptidase activity GO:0008237 Metal ion binding GO:0046872
ADAM30	May be involved in spermatogenesis and fertilization. Integral to membrane GO:0016021	Proteolysis GO:0006508	Metalloendopeptidase activity GO:0004222 Metallopeptidase activity GO:0008237 Zinc ion binding GO:0008270
NEK6	Protein kinase, which plays an important role in mitotic cell cycle progression. Required for chromosome segregation at metaphase-anaphase transition, robust mitotic spindle formation and cytokinesis. Spindle pole GO:0000922 Nucleus GO:0005634 Cytoplasm GO:0005737 Microtubule organizing center GO:0005815	Cytokinesis GO:0000910 Protein phosphorylation GO:0006468 Apoptotic process GO:0006915 Chromosome segregation GO:0007059 Mitosis GO:0007067	Magnesium ion binding GO:0000287 Protein kinase activity GO:0004672 Protein serine/threonine kinase activity GO:0004674 Signal transducer activity GO:0004871 Protein binding GO:0005515

	Microtubule GO:0005874