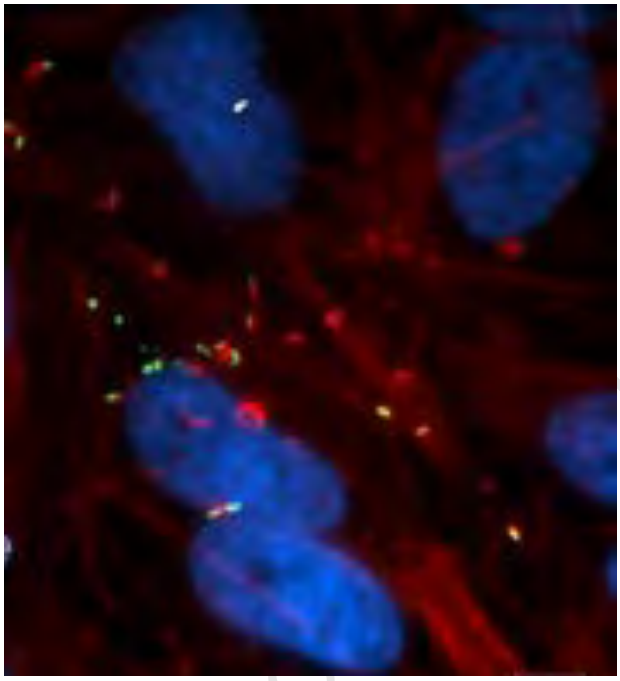
LncRNA discovery in the *Listeria monocytogenes* infection model.



Loretta Q. Magagula

MGGLOR001

A dissertation submitted to the University of Cape Town in fulfillment of the requirements for the degree of

MSc (Med) Immunology

Supervisors: Dr. Musa Mhlanga[1] & Prof. Frank Brombacher[2]

[1]Gene Expression and Biophysics Lab
Synthetic Biology Department
Biosciences
C.S.I.R.

[2]Department of Clinical Laboratory Sciences
Division of Immunology
Faculty of Health Sciences
University of Cape Town

February 2015

PLAGARISM DECLARATION

I, **Loretta Q. Magagula** hereby declare that the work on which this dissertation/thesis is based is 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: **14 February 2015**

ACKNOWLWEDGEMENTS

Firstly, I would like to thank my supervisor Musa Mhlanga, PhD for his mentorship, guidance, the countless challenges and opportunities he has awarded me over the past few years. A special thank you to Prof. Frank Brombacher for sharing his knowledge, expertise and advice during the tenure of this work.

A tremendous thank you to my past and present colleagues at Mhlanga Lab for their support, friendship and especially for staying awake and engaged enough to provide their constrictive criticisms at our infamously long lab meetings ;). A special thank you goes to Dr. Youtaro Shibayama, Dr. Stephanie Fanucchi, Dr. Samantha Brachievy and Dr. Janine Scholefield for their mentorship, training, guidance, assistance, kindness and friendships. I am forever indebted. Thank you to Dr. Anca Savulescu for her listening ear, sound advice and editorial services!

A special thank you to Joanna Cruz and Dr. Luis Pedro Coelho for their contribution to this work and bioinformatic know-how. Also, thank you Dr. Emiliano Dalla for being my FANTOM5 liaison. Thank you to Dr. Reto Guler and Dr. Suraj Pahir for their continued engagement and advice throughout this project.

To Allan Gray and everyone at the Allan Gray Orbis Foundation; staff and fellows thank you for an amazing fellowship journey that has and will continue to fuel my fire for science and our nation.

My undying gratitude goes out to the friends who have been there for me and supported me despite never understanding what I was complaining about. My rocks, my treasure, my rod and my champ, I couldn't have survived without any of you.

The most special thank you goes to my parents for the unconditional love and sacrifice that has led me to a field I am madly in love with. To my baby sisters Rhea and Kiana, thank you for the joy you bring to my life.

Finally, thank you Loretta for never giving up.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ABBREVAIATIONS

ANRIL	Antisense noncoding RNA in the INK4 locus
Δhly	Listeriolysin deficient
ΔLLO	Listeriolysin deficient
A-value	Relative expression value
cDNA	Complementary DNA
CHIP	Chromatin immune-precipitation
Ct	Cycle threshold
CV	Co-efficient of variation
DegSeq	Differential expression analysis package
DeSeq	Differential expression analysis package based on the negative binomial distribution
EMD-lncRNA	lncRNA within emerin 3'UTR
eRNA	Enchancer RNA
EST	Expressed sequence tags
FDR	False discovery rate
GO	Gene ontology
H. pylori	*Helibacteur pylori*
HDAC	Histone deacytelase
IFN-γ	Interferron gamma
IL	Interleukin
IL1β-eRNA	Interleukin 1 beta eRNA
IpaH9.8	*Shigella flexineri* E3 ubiquitin-protein ligase
iPSC	Induce pluripotent stem cell
ISG	IFN-γ-stimulated gene
KEGG	Kyto encyclopedia of genes and genomes
Lethe	Pseudogene Rps15a-ps4
linc-Cox2	LincRNA proximal to Cox2
Linc-NDRG1+	LincRNA near NDRG1, positive strand
lincRNA	Long intergenic noncoding RNA
Listeria	*Listeria* monocytogenes
Lm-MUT	*Listeria* monocytogenes EDGe *Δhly* strain
Lm-WT	*Listeria* monocytogenes EDGe wild type strain
lnc-IL7R	LncRNA transcribed from 3"UTR of ILR gene
lncRNA	Long noncoding RNA
LntA	*Listeria* nuclear targeted proein
LpqH	*M. tb* lipoprotein
LPS	Lipopolysaccharide
M-value	Log2 fold change ratio of gene expression
M. tb	*Microbacterium tuberculosis*
Matal1	Metastasis-associated lung adenocarcinoma transcript 1
MDM	Monocyte-derived macrophage
MHC2TA	MHCII transactivator
miR	MicroRNA
miRNA	MicroRNA
ncRNA	Noncoding RNA
*Neat*1	Nuclear enriched autosomal transcript 1
NeST	Nettoie Salmonella pas Theiler's translated to cleanup Salmonella not Theiler's
NF-κβ	Nuclear factor kappa beta
NGS	Next-generation sequencing
NMD	Nonsense mediated decay
ORF	Open reading frame
Osp	Outer *Shigella* protein
PACER	P50-associated COX-2 xtragenic RNA
PANDA	P21 associated ncRNA DNA damage activated

piwiRNA	P-element induced wimpy testis RNA
Pol II	RNA polymerase II
PTEN-P1	Phosphatase and tensin homolog pseudogene 1
RNA-*Seq*	RNA-Sequencing
RNAi	RNA interference
RPKM	reads per kilobase of transcript per million mapped reads
RT-PCR	Reverse transcription polymerase chain reaction
S. flexineri	*Shigella flexineri*
SAGE	Serial analysis of gene expression analysis
siRNA	Small interfering RNA
SOAP2	Single oligonucleotide
svm-predict	Support vector machine predicted
T3SS	Type III secretion system
THRIL	TNFα and hnRNPL related immunoregulatory lincRNA
TLR	Toll-like receptor
TNF-α	Tumor necrosis factor alpha
TSS	Transcriptional start site
TUCPS	Transcript of unknown coding potential
UTR	Untranslated region
XIST	X-inactive specific transcript

Abstract

A growing body of evidence indicates that long noncoding RNAs (lncRNAs), the most abundant noncoding RNA (ncRNA) species of the pervasively transcribed mammalain genome, have functional roles in the gene regulation of an array of cellular processes. These observations have since discredited the long standing central dogma formulated by Franscis Crick in 1958, which states that genetic output is entirely conducted by protein. Recent studies collectively indicate that lncRNAs play important functional roles in the transcriptional regulation of a wide array of cellular processes. In the last year alone, a handful of studies have identified lncRNAs linc-Cox2, Lethe, PACER and THRIL as central players in host cell innate immune response against microbial infection. These discoveries and the vast numbers of uncharacterized lncRNAs identified by high-throughput next-generation transcriptome sequencing technologies, set a precedence for further investigation and characterization of lncRNAs in infection biology. Importantly, lncRNAs may serve as important diagnostic markers of infection as well as therapeutic targets. These aspect of lncRNAs field although extensively being explored in cancer research, have been neglected in infection biology, particularly in microbial infection.

In this study, next-generation technologies were used to identify subtle vairations in transcriptional activity, with particular emphasis to lncRNA differential expression, and uncover their physiological relevance during *Listeria monotocytogenes* infection. To this end, an RNA-*Seq* dataset of *Listeria*-infected HeLa cells was subjected to several variations of data analysis lncRNA discovery pipelines. Potential lncRNA functioning was hypothesized using the Rinn & Chang “guilt by association” approach in which lncRNA functioning was hypothesized based on the known functions of tightly co-expressed protein coding mRNAs. “Guilty” lncRNAs were then knocked down in the HeLa cells using transcription activator-like nucleases (TALENs) to validate their candidacy as infection-regulating lncRNAs. Preliminary investigations conducted in this study have revealed potential *Listeria* infection-regulating lncRNA candidates. Furthermore, we explored the use of the physiologically relevant cellular model of iPSC-MDMs to validate identified lncRNA candidates. This work provides a framework for lncRNA discovery from RNA-*Seq* data by iterative and intergrative analysis.

Chapter 1: Introduction

1.1 Host-pathogen interactions: pathogens alter host gene expression

During their long coexistence with their eukaryotic hosts, pathogens have evolved several mechanisms to exploit host processes in order to establish successful infection within the host. In a similar manner, the mammalian host has developed complex intra- and inter-cellular systems to eradicate pathogenic infection. At the interface of this evolutionary conflict is an intricate network of connections and interactions between the pathogen and its host spanning an array of cellular processes. Fundamentally these interactions result in alterations in the functioning host transcriptional programmes during infection which directly and indirectly affect pathogen virulence, immune subversion, survival, multiplication and host cell survival.

Upon microbial infection host cells rapidly respond by augmenting the transcriptional activity of innate immune genes in an attempt to mount an appropriate response. Inherently, this involves host factors including chromatin modification complexes and transcriptional factors. Thus, these factors are attractive evolutionary targets for the reprogramming of host innate immune transcriptional responses by bacterial pathogens. Immune subversion by the alteration of histone modifications is an important mechanism used by multiple pathogenic bacteria during infection affecting a host signaling cascades and ultimately the host transcriptional profiles via an array of secreted factors and proteins (Bhavsar *et al*., 2007).

1.1.2. Bacterial pathogens alter host chromatin

One of the best agonists of the inflammatory response is the major gram-negative microbial cell wall component lipopolysaccharide (LPS). In addition to activating TLR4 receptors, LPS is able to transpose the nucleosome toward the promoter of immuno-responsive gene IL12, where it induces the acetylation of H4K14 and phosphorylation of H3S10 in TLR4-dependent manner (Weinmann *et al*., 1999, Weinmann *et al*., 2001). These LPS-dependent histone modifications facilitate chromatin accessibility for transcription factors such as NF-κβ at the IL2 promoter

and subsequent transcription of IL2 (Saccani *et al*., 2002; Arbibe *et al.*, 2007) **(Figure 1.1)**.

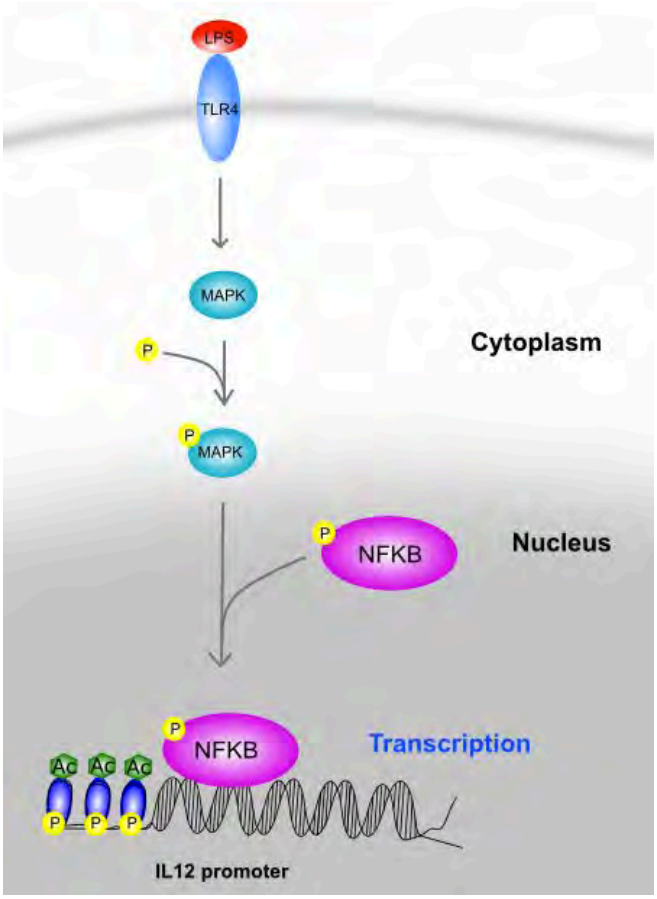


Figure 1.1: **LPS activates host IL 12 expresion.** The binding of LPS to TLR4 actiavtes MAPK phosphorylation. Phosphyrlated MAPK translocates ino the nucleus where it activates increased NF-κβ occupancy and subsequent transcriptional activation at the IL 12 promoter.

Bacterial pathogens use a cohort of nuclear-targeted injected and/or secreted factors/effectors to exploit normal cellular functions. For example, the gram-negative bacterium *Shigella flexineri* harbours a large virulence plasmid that encodes a type III secretion system (T3SS) whose secreted factors are responsible for *S.* flexineri pathogenesis. Among these effectors, T3SS-injected *S. flexineri* effector OspF, transverses the nucleus where it prevents the MAPK-dependent phosphorylation of H3S10 (**Figure 1.2**) (Saccani *et al*., 2002, Li *et al*., 2007, Zurawski *et al*., 2009). In addition, the LXCXE site on OspF associates with the C terminus of chromatin remodeling protein Retinoblastoma (Rb), which recruits histone deacetylases (HDACs), methlytransferases and other transcriptional repressive factors to the IL 8 promoter (**Figure 1.2**) (Macaluso *et al*., 2006, Zurawski *et al*., 2009).

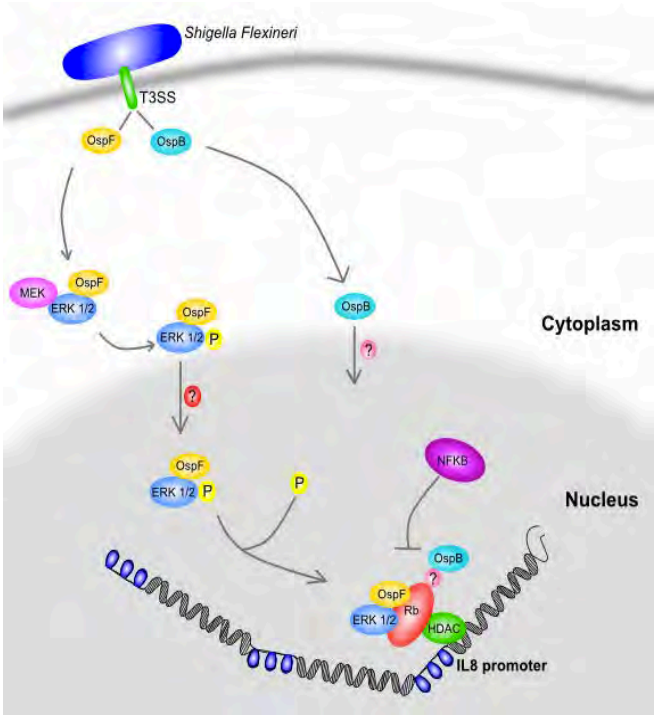


Figure 1.2: **Model of *Shigella flexineri* effector functioning OspF and OspB in epithelial cells**. OspF and OspB are secreted via the *Shigella* T3SS and enter the nucleus by unknown mechanisms. In the case of OspB, the N-terminus is required for nuclear import. OspF associates directly with ERK1/2 and Rb. OspB associates directly with Rb via an unknown binding site or unknown intermediary protein (signified by ?). OspF and OspB in turn use Rb to recruit factors such as histone deacetylases (HDAC), which down-regulate IL-8 production (and other genes) by remodeling the chromatin architecture. Adapted from **Zurawski *et al*., 2009.**

OspB, an additional T3SS-injected nuclear-localized effector directly interacts with Rb correlating with decreased immuno-responsive IL8 secretion (**Figure 1.2**) (Zurawski *et al.*, 2008). These renovations in the nuclear architecture encompassing the IL8 promoter decrease the accessibility of key immune activators such as NF-κβ to this locus (Saccani *et al*., 2002, Arbibe *et al*., 2007). A similarly targeted T3SS effector, IpaH9.8 binds the U2AF host splicing factor interfering with its ability to bind and splice IL8 transcripts (Cuellar-Mata *et al*., 2002, Okuda *et al.*, 2005).

Helibacter pylori, the gram negative bacterium that leads to chronic gastritis and peptic ulcers, has also been shown to be correlated with a number of histone modifications within the host nucleus. In gastric cells*, H. pylori* infection led to a dose- and time-dependent phosphorylation of H3S10 and decreased H3K23 acetylation identified by CHIP (Chromatin immuno-precipitation) assays (Ding *et al*., 2010). These *H. pylori*-dependent histone modifications are correlated with the upregulation of the host inflammatory-induced gene c-Jun (Ding *et al*., 2010). These modifications are specifically dependent on the expression of *H. pylori* virulence factor cag pathogenicity island (cagPAI).

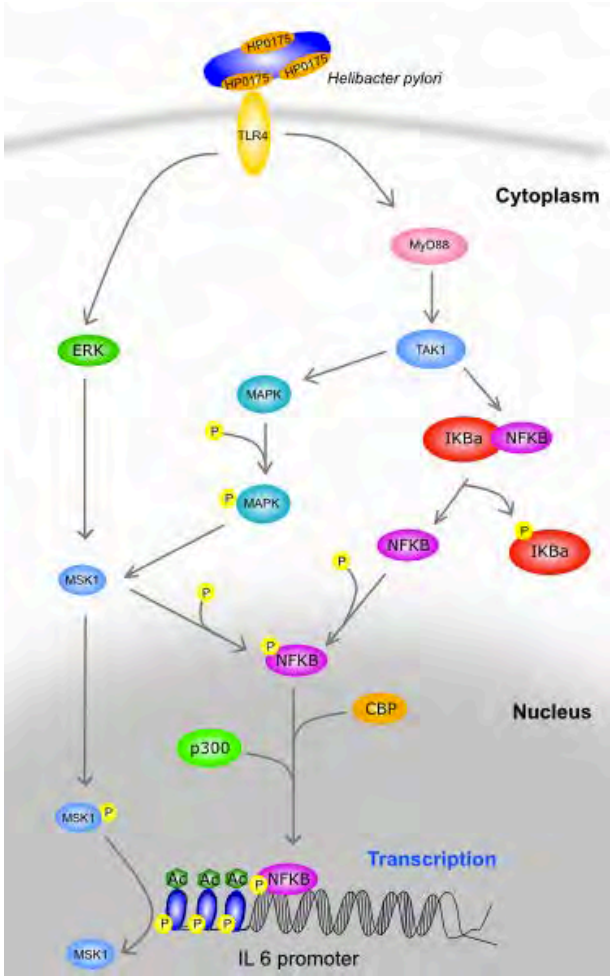


Figure 1.3: ***Helibacter phylori* HP0175 activates IL 6 transcription**. HP0175 activates the MYD88, and MAPK ERK signaling cascades in a TLR4-dependent manner. Activation of MYD88 leads to the phosphorylation of MAPK and the dissociation of inhibitory Iκβα from NF-κβ allowing NF-κβ to be phosphorylated and translocate into the nucleus, where it recruits transcriptional co-activators p300 and CBP to the IL 6 promoter. Activation of the MAPK and ERK pathways results in the phosphorylation of MAPK, which in turn phosphorylates MSK1 and NF-κβ. Phosphylated MSK1 in turn phosphorylates H3S10 at the IL 6 promoter. Adapted from **Pathak *et al.* 2006.**

HP0175, a *H. pylori* secreted peptidyl prolyl cis-, trans-isomerase is responsible for the H3S10 phosphorylation-mediated increase IL 6 expression in mouse macrophages (Pathak *et al*., 2006). HP0175 mediates overexpression of IL-6 in a TLR4-dependant manner resulting in the induction of NF-κβ, ERK and MAPK (mitogen-activated protein kinase) signaling cascades. The induction of the ERK and MAPK pathways leads to the activation of nuclear mitogen- and stress-activated protein kinase (MSK1), which phosphorylates H3S10 and p65 at Ser^{276} (Pathak *et al*., 2006) **(Figure 1.3)**. As mentioned earlier, H3S10 phosphorylation is required for NF-kB-dependent transcription. The phosphorylation of the NF-κβ p65 subunit enables the p50-p65 complex to interact with transcriptional co-activators p300 and CREBB binding protien (CBP) (Zhong *et al*., 2002) **(Figure 1.3)**. Pathak and colleagues (2006), also showed the association of phosphorylated p65 at the IL6 promoter using CHIP in macrophages, which correlated with the overexpression and release of IL 6 from/in macrophages induced by HP0175.

The 19 kDa cell wall protein of *Mycobacterium tuberculosis*, LpqH, interacts with TLR2 and inhibits IFN-γ-induced expression of class II transactivator (CIITA)-regulated major histocompatibility complex II (MHCII) genes and genes involved in antigen processing and presentation in macrophages (Pennini *et al*., 2006). IFN-γ induction of MHCII genes results in the activation of the JAK signaling cascade, which leads to the phosphorylation of transcriptional factor STAT1, allowing STAT1 to translocate into the nucleus where it mediates the transcription of several genes such as the CIITA gee, *MHC2TA.* (**Figure 1.4**). Thus, LpqH inhibits CIITA expression by TLR2-induced MAPK-dependent signaling that results in the deacetylation of H3 and H4 at the *MHC2TA* promoter (**Figure 1.4**) (Pennini *et al*., 2006).

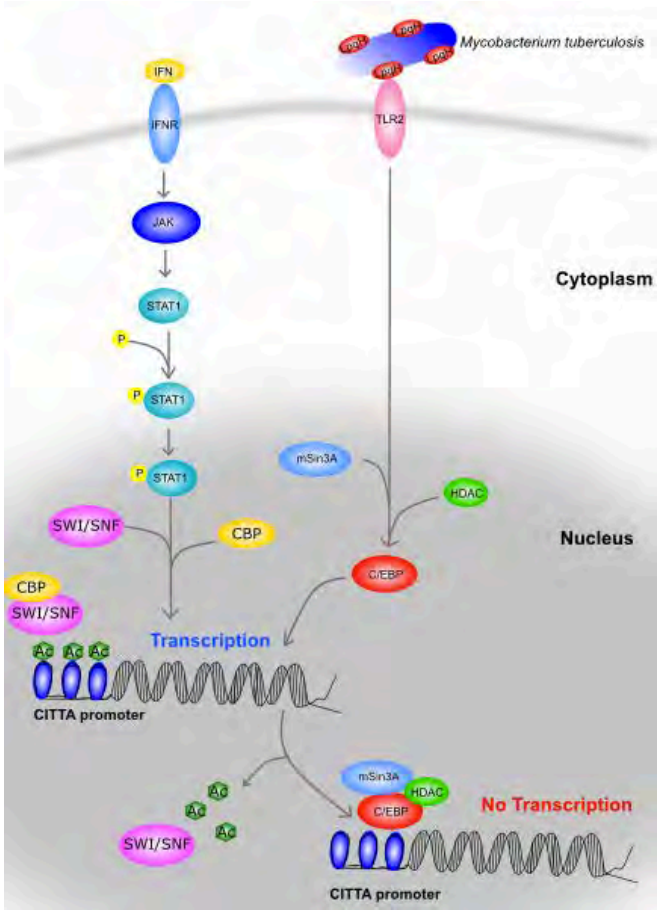


Figure 1.4: **Schematic representation of *Mycobacterium tuberculosis* LpqH-induced inhibition of CIITA expression.** *M.* tb/LpqH inhibits STAT1/JAT-induced histone acetylation and SWI/SNF and CBP transcriptional activator binding to CIITA . LpqH leads to the recruitment of transcriptional co-repressor C/EBP and HDACs to the pIV CIITA promoter which leads to the displacement of SWI/SNF and CBP at the CITTA promoter. *M.*tb/LpqH inhibits leads to the recruitment of HDACs/mSin3A repressive chromatin remodeling complex at the same promoter inhibiting CITTA expression. **Hamon and Cossart, 2008.**

Furthermore, incubation of macrophages with LpqH led to the loss of chromatin remodeling Brg1 protein, a major component of the SWI/SNF remodeling complex, and CBP, a transcriptional co-activator and histone acetyl transferase, IFN-γ-induced binding at the CIITA pIV promoter (**Figure 1.4**) (Pennini *et al.,* 2006). LpqH therefore mediates inhibition of IFN-γ-induced CIITA expression at the level of histone modifications as well as protein-specific binding at the pIV promoter.

Additionally, LpqH induced the binding of C/EBP transcriptional repressor, to the CIITA promoter (Pennini *et al.*, 2007).

The inhibition of CIITA expression in macrophages leads to the decreased CIITA-induced expression of MHCII class proteins such as HLA-DR. With regards to HLA-DR expression, this effect is compounded by the binding of the HDACs/mSin3A co-repressor complex binding at the HLA-DR promoter (Wang *et al.*, 2005). LpqH appears to mediate transcriptional repression of several genes by remodeling chromatin and altering histone modifications in order to repress an IFN-γ-induced immune response.

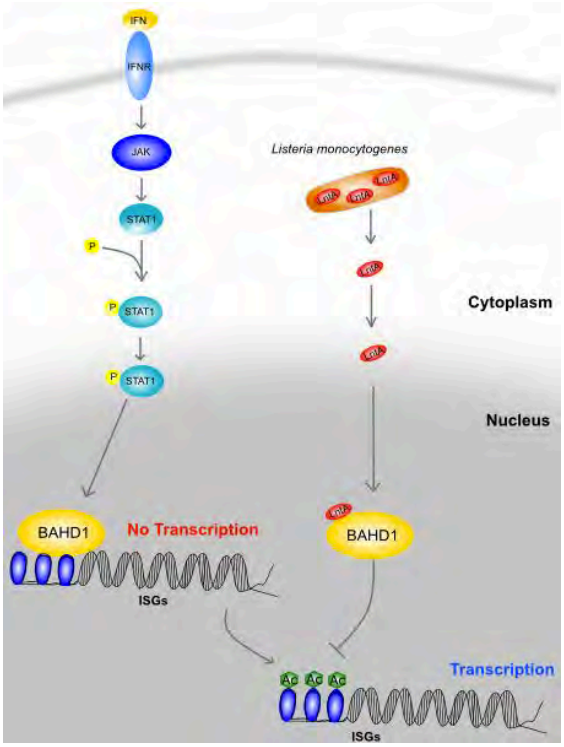


Figure 1.5: ***Listeria monocytogenes* protein LntA relieves BAHD1-mediated heterochromatin to stimulate IFN type II response.** IFN- type III stimulated genes (ISG) are subject to BAHD1-mediated repression. *Listeria* secreted and nuclear targetted protein LntA relieves this repression by direct BAHD1 binding. LntA-BAHD1 displaces BAHD1 silencing complex (containing TRIM28 and HP1) from ISGs, promoting elevated H3K9 occupancy and ISG transcription. **Lebreton *et al*., 2011**.

The pathogenic strain of food-borne intracellular bacterium *Listeria monocytogenes*, expresses nuclear-targeted LntA secreted protein. The LntA protein is expressed by the virulence gene *lmo0438/lntA* gene which is absent in nonpathogenic strains of the bacterium. This 205aa basic protein with a N-terminal signal peptide accumulates within the nuclei of fibroblasts during late infection. Lebreton *et al.* (2011) revealed a strong interaction between LntA and host silencing factor, BAHD1 resulting in the displacement of the BAHD1 heterochromatin promoting complex containing HP1 and TRIM28 from IFN-γ-stimulated genes (ISGs) during late infection (**Figure 1.5**). Thus, LntA specifically relieves BAHD1-mediated

transcriptional repression leading to the enrichment of H3K9Ac and subsequent elevation of transcription at ISGs such as Ccl5 (Lebreton *et al.* 2011).

Collectively, these observations demonstrate a remarkable precision in which microbial factors can function in concert to remodel the nuclear architecture and subsequently alter the transcription of specific immune response genes. It is important to note that in addition to the bacterial proteins described here, there is a growing number of bacterial proteins targeted to the nucleus to alter host transcription. Further study of nuclear-targeted microbial factors and their functions may help develop strategies and identify therapeutic targets in order to combat infection.

1.2. Long noncoding RNAs

The current revised dogma of biology establishes an importance on the transcribed portion of the genome that does not translate into functional protein. These noncoding RNAs (ncRNAs) have challenged the predicted flow of genetic information in eukaryotes. In addition, several regulatory ncRNA classes have been established in a plethora of integral cellular functional processes. Many of the known functional mechanisms however, are primarily of short regulatory transcripts including microRNAs (miRNAs) and small interfering RNAs (siRNAs). Recently, deep sequencing technologies have reproducibly demonstrated the existence of long noncoding RNAs (lncRNAs) that are estimated to represent 70-90% the transcribed mammalian genome (Mercer *et al*., 2009*,* Derrien *et al.* 2012, Hangauer *et al.,* 2013).

LncRNAs are currently classified as ncRNAs greater than 200 bp in length. This size inclusion is primarily employed to distinguish lncRNAs from smaller ncRNAs such as miRNAs, siRNAs and piwi-RNAs. Similar to translated mRNAs, lncRNAs can be capped, polyadenylated and a distingiushed by the lacking an obvious capacity to encode protein (Bertone *et al.*, 2004, Kapranov *et al.,* 2010). Unlike their mRNA counterparts, lncRNAs have poor sequence conservation and are expressed at several orders of magnitude lower than mRNAs, with the exceptions of a few abundant nuclear body forming lncRNAs (Pang *et al*., 2006, Clemson *et al*., 2009, Tripathi *et al*., 2010, Andersson *et al*., 2014).

Several classes of lncRNAs have been established based on the characteristics of their genomic loci (**Figure 1.6**). Long intergenic/intervening RNAs (lincRNAs) are transcribed from the noncoding portions of the genome between protein coding genes (Guttman *et al.* 2009, 2010). These lncRNAs are typically named according to their closest mRNA transcriptional start site (TSS). This class of lncRNAs includes a class of lincRNAs transcribed from enhancer regions termed enhaner RNAs (eRNAs) (Kim *et al*., 2010). Antisense lncRNAs are transcribed from the opposite strand of known protein coding genes (Derrien *et al*., 2012). These lncRNAs overlap at least in part with annotated spliced or intronless sense mRNAs. Intronic lncRNAs overlap with the introns of protein-coding genes either in the sense or antisense direction (Louro *et al.* 2008). Bidirectional lncRNAs are transcribed “head-to-head” or “tail-to-tail” with protein-coding genes within <1kb (Engström *et al.,* 2006).

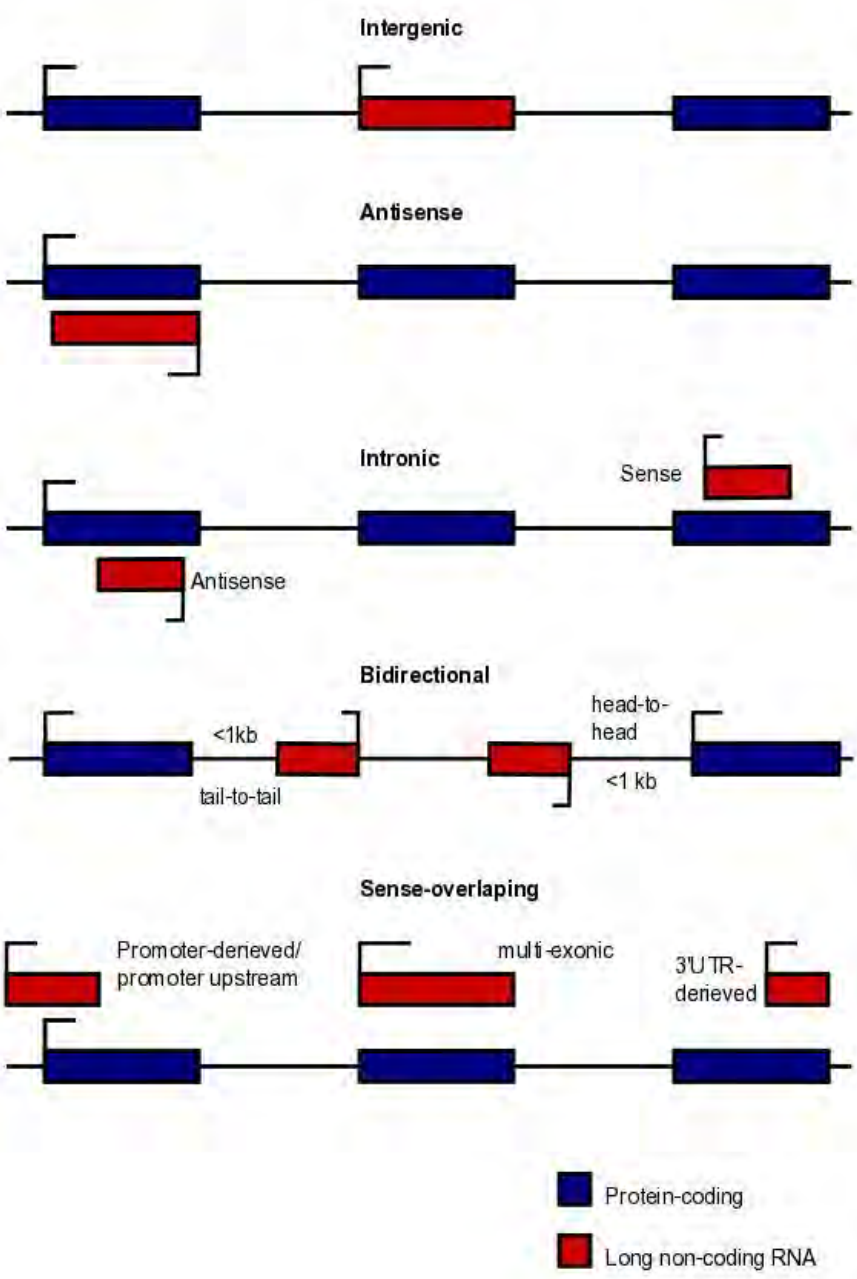


Figure 1.6: **Genomic characteristics of long non-coding RNAs**

Bidirectional lncRNAs include promoter-derived or -associated lncRNAs that overlap promixal protein coding promoters either in the sense or antisense direction within 1 kb from protein-coding genes suggesting these lncRNAs and mRNAs are transcribed by the same transcriptional machinery (Hung *et al.* 2011). Sense-overlapping lncRNAs are considered transcript variants of annotated genes yet are not translated either due to a lack of substantial ORFs, retained introns, or are degradation by nonsense mediated decay (NMD) (Derrien *et al.,* 2012). Within this

class there are: promoter-derieved/upstream lncRNAs, multi-exonic sense lncRNAs as well as 3'UTR-derieved lncRNAs. (**Figure 1.6**)

In addition, lncRNAs can be characterized according to histone signatures occupying their genomic loci. In 2009, Guttman and collegues identified a distinctive consensus chromatin signature in four different mouse cell types for identifying lncRNAs. The signature, similar to that of actively transcribed genes, referred to as the "K4-K36 domain", consists of a short region with H3K4me3 occupancy (corresponding to the promoter) followed by a longer H3K36me3 occupied region (corresponding to the transcribed region) (Guttman *et al.* 2009, Khalil *et al*., 2009). Other epigeentic signatures used to identify lncRNAs include the high H3K4me1-K3K4me3 ratio, which has been used to identify eRNAs (De Santa *et al*., 2010, Marques *et al*., 2013).

1.2.1 Molecular mechanisms of gene regulation by lncRNAs

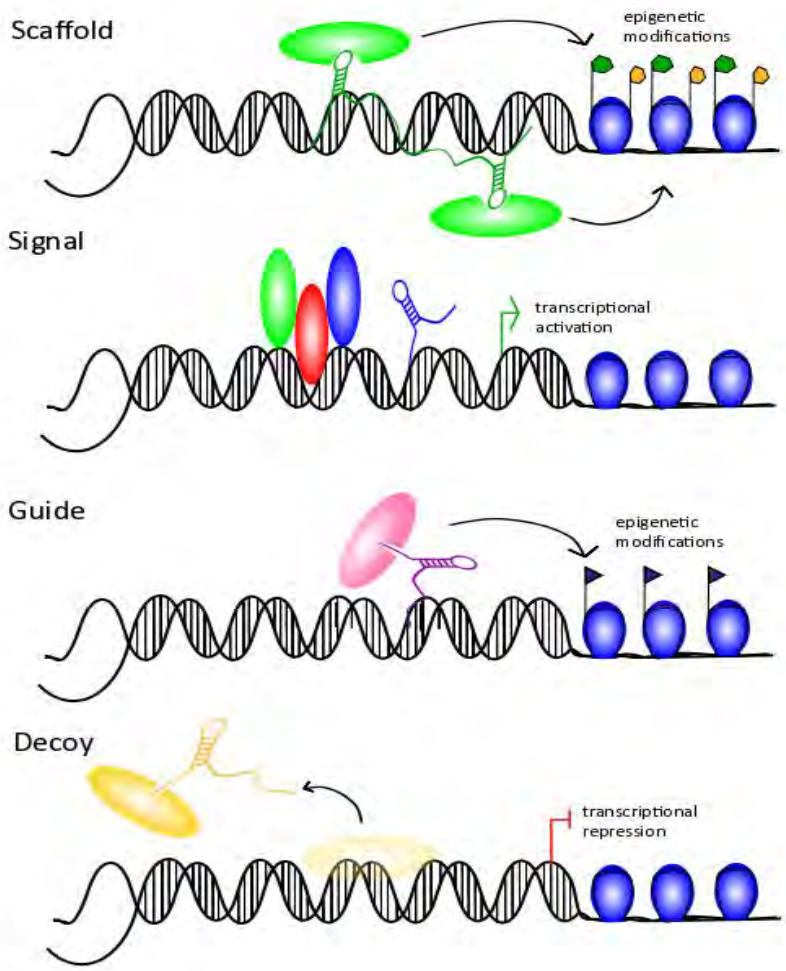


Figure 1.7: **Molecular mechanisms of lncRNAs**. Adapted from **Wang and Chang (2011)**

1.2.1.1 Molecular signal lncRNAs

The expression and functioning of lncRNAs is tightly spatio-temporally controlled allowing for specific lncRNAs to serve as molecular "cues" or signals for different cellular contexts (Wang and Chang, 2011) (**Figure 1.7**). This mode of cellular functioning is best illustrated by lncRNAs transcribed within the HOX clusters, with parallel expression and spatial positioning within the cluster and the anterior-posterior anatomic axis of the mammalian body (Wang *et al*., 2009). Hox cluster-transcribed lncRNAs HOTAIR (Hox antisense intergenic RNA, HoxC cluster) and HOTTIP (Hox A transcript at the distal tip, HoxA cluster) are expressed in distal and posterior cell types whereas Frigidair (HoxC) is expressed in anterior cells (Rinn *et al*., 2007). Similarly, p53-dependent lncRNAs linc-p21 and PANDA (p21 associated ncRNA DNA damage activated) control the switch between apoptosis and cell survival upon p53 activation (Huarte *et al.* 2010, Hung *et al.* 2011). These discoveries demonstrate tightly controlled cell-, context-, spatio- and temporal regulation of lncRNA expression and functioning.

1.2.1.2 Molecular sponge lncRNAs

Several lncRNAs regulate gene expression simply by titrating transcriptional proteins away from specific cytoplasmic or nuclear sites and/or proteins (Wang and Chang, 2011). Effectively these lncRNAs function as molecular "sinks" or decoys for transcriptional protein repressors and translational miRNA repressors (**Figure 1.7**). The signal lncRNA PANDA has been shown to titrate transcription factor NF-YA away from pro-apoptotic genes preventing their transcription following the activation of p53 tumor suppressor (Hung *et al.,* 2011). The 3'UTR of pseudogene *PTEN-P1* (Phosphatase and tensin homolog pseudogene 1) has been shown to bind and "sink" miRNAs targeted against *PTEN* mRNA (Poliseno *et al*., 2010). Thus *PTEN-P1* competes with endogenous preventing RNAi-mediated decay of *PTEN* transcripts effectively, albeit indirect, increasing functional transcriptional output of the *PTEN* gene. Similarly, lncRNA-ATB competitively binds the miR200 family resulting in the up-regulation of miR200 targets ZEB1 & 2 (Yuan *et al.* 2014). Additionally, the nuclear abundant lncRNA *Malat1* (metastasis-associated lung adenocarcinoma transcript 1) specifically sequesters SR splicing factors to nuclear speckles and decreases alternative mRNA splicing (Tripathi *et al*., 2010).

1.2.1.3 Guide lncRNAs

One of the most well represented lincRNAs, *XIST* (X-inactive specific transcript) functions by physically 'coating' one of the two X chromosomes in female mammals. *XIST* sequesters the Polycomb repressive complex 2 (PRC2) and subsequently H3K27me3-marked histones onto target chromosome producing a transcription-intolerant local environment devoid of RNA polymerase II and basal transcription factors (Plath *et al*., 2003; Lee *et al,* 2010). The imprinted IGF2R intronic lncRNA *Air* also functions as a "guide" lncRNA for H3K9 methylation to facilitate gene- and paternal-allele -targeted repression (Nagano *et al*., 2008). In addition, the extensively studied p53-activated antisense lncRNA, *linc-p21*, has been reproducibly shown to recruit heterogeneous nuclear ribonuclear protein (hnRNP) K to p53-regulated gene promoters mediating p53-dependent transcriptional repression (Huarte *et al.*, 2010). Collectively, these studies demonstrate the capacity of lncRNAs to regulate transcriptional output by guiding chromatin-modifying complexes to specific gene loci (**Figure 1.7**).

1.2.1.4 Molecular scaffold lncRNAs

LncRNAs can also alter the transcriptional microenvironment by acting as molecular scaffolds for protein-DNA and protein-protein interactions (**Figure 1.7**). For example, the 5' domain of HOTAIR (Hox antisense intergenic RNA) binds PRC2, whereas the 3' domain binds the LSD1/CoREST/REST complex coordinating chromatin targetting and subsequent H3K27 methylation and H3K4 demethylation at the HOXC gene cluster (Rinn *et al*., 2007, Tsai *et al.* 2010). The p15/CDKN2B (cyclin-dependent kinase inhibitor 2B) antisense lncRNA *ANRIL* (Antisense noncoding RNA in the INK4 locus) functions as a molecular scaffold for the interaction between PRC1 and PRC2 complexes, leading to epigenetic silencing of the INK4a locus (Yap *et al.* 2010, Kotake *et al.*, 2011). The abundant sub-nuclear lincRNA *Neat1* (nuclear enriched autosomal transcript 1) is a scaffold for nuclear paraspeckles and splicing factors p54 and PSP1 (Paraspeckle protein 1) (Clemson *et al*., 2010). LncRNA-ATB (lncRNA (lincRNA) activated by TGF-β) binds IL 11 mRNA transcripts reinforcing mRNA stability and modulating IL 11 gene expression (Yuan *et al.* 2014). These studies uncover a class of lncRNAs that exert their molecular function by binding specific molecules in order to provide a framework for consequent cellular processes.

1.3. LncRNAs in innate immunity and pathogenic infection

Table 1: LncRNAs in the innate immune response

lncRNA	Genomic localization	Molecular mechanism	Function	Ref.
THRIL	*Intergenic,*	Guide, Scaffold	Forms an RNP complex with hnRNPL that binds to TNFα promoter to regulating basal and stimulated transcription of TNFα.	Li *et al*., 2014
NeST	*Intergenic, Antisense*	Guide, Scaffold	Binds and recruits WDR5 to the IFN -γ promoter, increasing expression in TH1 CD4+ T cells, CD8+ T cells, and natural killer cells.	Gomez *et al*., 2013
***linc*-Cox2**	*Intergenic, Antisense*	Scaffold	Pam3CSK4-induced scaffold for hnRNP-A/B and hnRNP-A2/B1 to promote transcriptional repression.	Carpenter *et al.* 2013, Heward and Lindsay 2014
PACER	*Intergenic, Antisense*	Sponge	Occludes repressor complex (p50/p50) from Cox2 promoter to facilitate Cox2 expression.	Krawczyk and Emerson, 2014
Neat1	*Intergenic*	Sponge	Sequesters the SFPQ repressor away from the IL8 promoter to nuclear paraspeckles in order to facilitate IL8 transcription.	Hirose *et al.,* 2014
IL1β-eRNA	*Intergenic, enhancer-derived*	Unknown	Activates LPS-induced expression and release of IL1 and CXCL8 in an NF-Kβ-dependent manner.	Illot *et al.,* 2013
IL1β-RBT46	*Bidirectional*	Unknown	Mediates LPS-induced expression and release of the proinflammatory mediators, IL1 β and CXCL8.	Illot *et al.,* 2013
lnc-IL7R	*Intragenic, 3'UTR-derived*	Unknown	Inhibits the LPS-induced inflammatory response by increasing H3K27 trimethylation at proximal promoters.	Cui *et al.,* 2014
Lethe	*Pseudogene*	Sponge	Binds and sequesters p65 (RelA) away from NF-κβ responsive genes preventing their transcription.	Rapicavoli *et al*., 2013

Although the advert of next generation sequencing technologies such as RNA-Seq has allowed the identification of thousands of lncRNAs only a few of them have been characterized from a functional standpoint. However, a prevailing theme unveiled by lncRNA functional studies is transcriptional regulation typically mediated by the ability of lncRNAs reconfigure the transcriptional landscape of their target genes.

The plethora of lncRNA functions makes these transcripts potential targets for bacterial pathogens and/or proteins in order to elicit pathogen-imposed cellular responses. Conversely, lncRNAs may function as part of the host response against infection. Several transcriptome-wide studies have revealed definitive differences in the ncRNA expression profiles of infected and non-infected host cells revealing that pathogens or pathogen-derived molecules elicit changes on the expression of the noncoding portions of the genome (Mortazavi *et al*., 2008, Henn *et al.* 2012, Lisnic *et al.* 2013, Ilott *et al.* 2014). These findings set precedence for further investigation and functional characterization of lncRNAs in infection biology. Although several studies have sequenced whole transcriptomes of bacterial pathogens as well as dual host and pathogen sequencing, only a few lncRNAs have been shown to be involved in infection (**Table 1.1**).

1.3.1 THRIL "guides" and functions as a molecular scaffold for hnRNPL at the TNF-α promoter

A ubiquitously expressed 2-2.5 kb lincRNA termed THRIL (TNFα and hnRNPL related immunoregulatory lincRNA) was identified as an innate immune responsive lncRNA by microarray analysis of Pam3CSK4 stimulated THP1 monocytes (Li *et al.* 2014). Pam3CSK4 is a synthetic lipopeptide and potent activator of proinflammatory transcription factor NF-κβ that mimics the acetylated amino acids on bacterial lipopolysacchrides. Transcribed from the reverse strand of the gene Bri3bp (BRI3 binding protein), with ~450 bp overlapping with the Bri3bp mRNA 3′ UTR, THRIL has been shown to bind and guide hnRNPL to the proinflammatory cytokine, TNF-α, promoter stimulating TNFα transcription (**Figure 1.8**). hnRNPs are multifunctional ribonuclear RNA binding proteins that also act as mediators of lncRNA-induced transcriptional repression (Huarte *et al*., 2010, Hasegawa *et al*., 2010, Carpenter *et al.*, 2013).

Expression analysis and other functional studies including RNAi and RT-PCR identified THRIL as regulator of basal and Pam3CSK4-mediated TNF-α expression (**Figure 1.8**). shRNA (small hairpin RNA) knockdown of THRIL reduced the expression of 444 genes in non-stimulated cells and blocked Pam3CSK4-mediated differential expression of 317 of the 618 including multiple inflammatory genes such as IL6, CXCL8, CXCL10, CCL1, and CSF1.

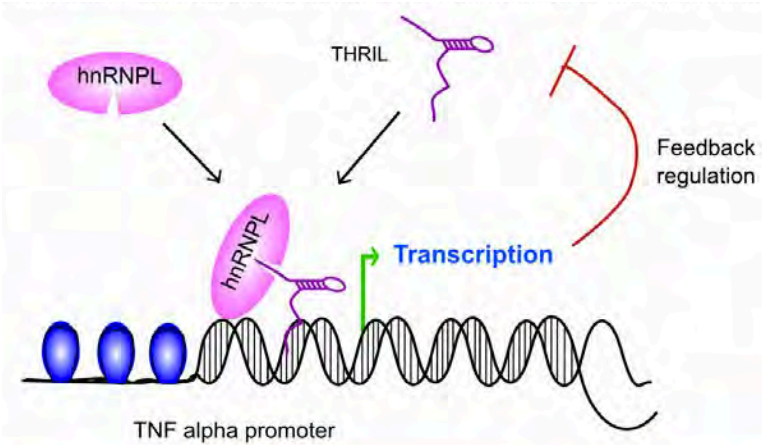


Figure 1.8: **THRIL "guides" hnRNPL to TNF-α promoter.** THRIL and hnRNPL form an RNP complex that regulates basal and stimulated transcription of TNFα by binding to the TNFα promoter. In infected cells, high levels of TNFα secretion initiate a negative feedback loop in which THRIL and subsequently TNFα expression is down-regulated. **Li et al., 2014.**

These findings suggest TNF-α activated-THRIL functions as a scaffold lncRNA binding hnRNPL and recruiting it to the TNF-α promoter to drive TNF-α expression. Importantly, this study proposed a functional mechanism, which reinforced the responsiveness of lncRNAs to microbial pathogen-derived molecules and their ability to mount an inflammatory response to combat infection by modulating gene expression.

1.3.2 NeST "guides" WDR5 to IFN-γ to increase transcription

In 2013, Gomez *et al.* characterized a pathogen-regulated, guide, enhancer-like lncRNA NeST (Nettoie Salmonella pas Theiler's translated to cleanup Salmonella not Theiler's), formally known as Tmevpg1. NeST is transcribed ~45kb downstream and antisense from the IFN-γ locus in both mice and humans. Originally, NeST was initially identified as a susceptibility candidate gene in T cells that are unable to clear

Theiler's virus pincovirus (Vigneau *et al.* 2001). The expression of at least 0.15 NeST copies per cell in transgenic mice promotes virus persistence.

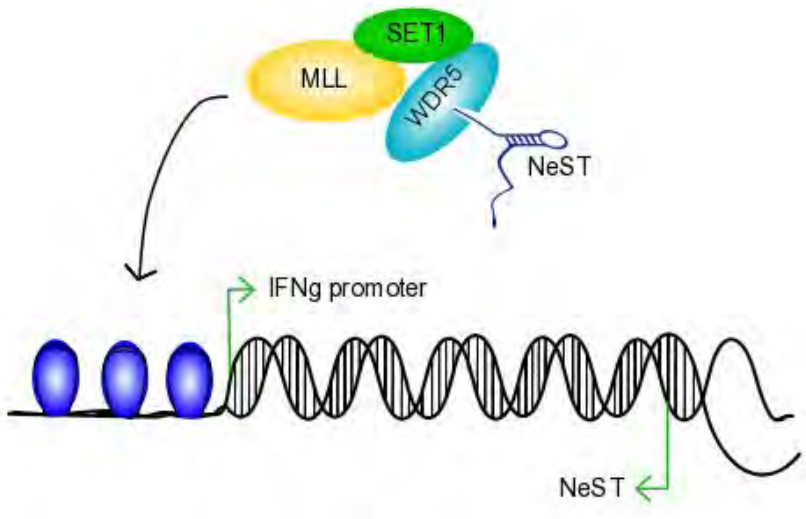


Figure 1.9: **NeST up-regulates IFN-γ.** IFN-γ downstream and antisense lncRNA NeST that binds and recruits WDR5 to the IFN-γ promoter leading to H3K4 methylation and IFN-γ expression in in TH1 CD4+ T cells, CD8+ T cells, and natural killer cells in mice. **Gomez *et al*., 2013.**

Intriguingly the persistence of *Salmonella* infection in NeST expressing cells is greatly reduced. NeST displays recurring molecular lncRNA functioning mechanisms by binding protein, altering epigenetics and enhancing transcription in pathogen-invaded cells. This is achieved by the binding of NeST lncRNA to WDR5 subunit of the H3K4 methyltransferase complex MLL/SET and this binding event regulates the methylation state of H3K4 and occupancy at the IFN-γ locus (**Figure 1.9**, Gomez *et al.* 2013). This transcription-favorable epigenetic environment leads an inflammatory response that efficiently clears *Salmonella* infection by increasing IFN-γ expression. Thus NeST is a uniquely functioning lncRNA involved in the susceptibility of viral and bacterial infection in mouse models.

1.3.3 hnRNP-linc-Cox2 scaffold represses innate immune system gene expression

RNA-*Seq* of LPS-stimulated bone marrow derived macrophages (BMDMs) and RT-PCR of Pam3CSK4 stimulated bone marrow derived dendritic cells (BMDCs) revealed an increased induction in immune response genes and lincRNA-Cox2 (Guttman *et al*., 2009, Carpenter *et al.* 2013). Identified using the epigenetic K4-K36

domain chromatin signature, *linc-Cox2* is located ~51 kb downstream of the Cox2 gene (Pgts) (Guttman *et al*., 2009). Similarly to LPS- and Pam3CSK4-stimulated primary cells, the transcription of Cox2 and *linc-Cox2* in *Listeria monocytogenes*-infected BMDMs and splenocytes was elevated (Carpenter *et al.* 2013). Cox2 and *linc-Cox2* up-regulation is induced by TLR ligands in a MyD88- and NF-κβ-dependent manner. RNA-*Seq* of *linc-Cox2*-silenced BMDM cell revealed that *linc-Cox2* induced the expression of 787 genes, the majority of which were immune system genes including chemokine's and IFN-γ-stimulated genes (ISGs).

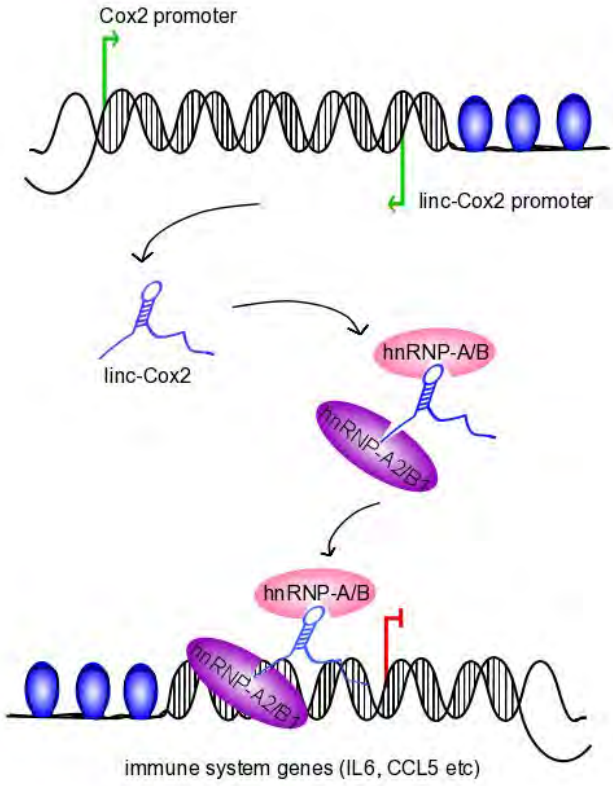


Figure 1.10: **hnRNP- linc-Cox2 scaffold represses innate immune genes.** Downstream and antisense to Cox2, lincRNA-Cox2 binds *hnRNP-A/B and hnRNP-A2/B1* to repress basal and pathogen-induced expression of innate immune system genes including IL6 and CcL5. **(Heward and Lindsay 2014).**

Linc-Cox2 binds both nuclear and cytoplasmic populations of hnRNP-A/B and hnRNP-A2/B1 to repress immune system genes including IL6 and Ccl5 in bacteria-infected primary mouse cells (**Figure 1.10**, Carpenter *et al.* 2013). These findings suggest like other hnRNP-binding lncRNAs such as THRIL, *linc-Cox2* may function as a bacterial-induced scaffold for hnRNP-A/B and hnRNP-A2/B1 to promote the transcription of targeted innate immune system genes.

1.2.2.4 PACER: p50 molecular sponge regulates Cox2 expression

PACER (p50-associated COX-2 extragenic RNA) is an intergenic lncRNA located 1100 bp upstream and antisense of the Cox2 TSS (**Figure 1.11**, Krawczyk and Emerson, 2014). The expression of this lncRNA is required for LPS-induced Cox2 expression in phorbal microstate acetate (PMA) differentiated U937 macrophages. This action is mediated by PACER's ability to bind and sequester the repressive p50 homodimer of NF-κβ from Cox2 promoter association. This occlusion results in decreased p50 occupancy at the Cox2 promoter, providing a permissive environment for the binding of transcription factors and chromatin modifying complexes (Krawczyk and Emerson, 2014).

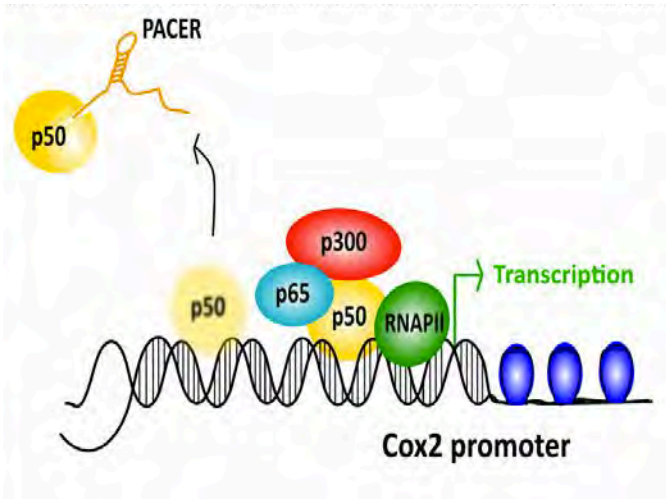


Figure 1.11: **PACER regulates Cox-2 expression.** PACER lncRNA physically interacts and restricts excess p50 from binding the promoter to facilitate the exchange of repressive p50/50 homodimers for p50/p65 heterodimers, thus allowing recruitment of p300, induction of histone hyper acetylation/chromatin remodeling, and consequent assembly of RNAP II complexes competent for transcription activation. **(Krawczyk and Emerson, 2014).**

Specifically, the expression of PACER resulted in significant increases in the acetylation state of H3 and H4 upstream of the Cox2 locus. The histone-modifying action of PACER is mediated by the recruitment of the p300 histone acetyltransferase to the Cox2 locus in LPS-stimulated macrophages (**Figure 1.11**). In addition, the presence of PACER at the Cox2 locus correlated with the assembly of RNA Pol II initiation complexes and subsequent transcription at the Cox2 locus (**Figure 1.11**). These results present a model of lncRNA functioning where PACER functions as a p50 molecular sponge decreasing p50 occupancy at the Cox2 promoter to facilitate the recruitment of activating p50/p65 heterodimers, p300 and RNA Pol II complexes for the transcriptional activation of Cox2 in bacteria infected macrophages (Krawczyk and Emerson, 2014).

1.4. Next-generation sequencing decodes the transcriptome

The term “transcriptome” describes the complete repertoire of transcripts expressed in a specific cell type or tissue at a certain developmental stage and/or under a specific physiological condition. Inherently, it is highly dynamic with varied transcript quantities and combinations in different cellular contexts. Thus, elucidating the transcriptome using global gene expression analysis has emerged as a powerful tool to investigate cell behavior under different physiological conditions and classifying disease states.

1.4.1 Quantifying the transcriptome

1.4.1.1 Hybridization-based microarrays

Early gene expression analysis tools have included Northern blotting, reverse-transcription PCR (RT-PCR), expressed sequence tags (EST) and serial analysis of gene expression analysis (SAGE). However, these tools have failed to provide rapid, high-throughput, unbiased and a complete picture of the transcriptome. The advent of hybridization-based DNA microarrays over a decade ago has bridged this gap providing a rapid, high-throughput method to quantify the differential expression of annotated genes in different cellular contexts in parallel (Schena *et al.*, 1995). DNA microarrays involve the hybridization of pre-determined immobilized DNA oligonucleotide probes to reverse transcribed target cDNA. The cDNA target is fluorescently labeled prior to probe hybridization on the microarray. A successful hybridization event between the labeled cDNA target and the immobilized probe results an increase of fluorescence intensity over a background level. These levels of hybridization, measured by fluorescence intensity, are then converted into quantitative expression measurements (Schena *et al.*, 1995).

Although microarrays are both highthroughput and inexpensive, they are prone to several limitations. Such hybridization-based transcriptomic approaches are biased toward known and predicted genes, relying on pre-existing knowledge of genome sequence with the exception of whole-genome high-density tiling microarrays that utiliize regularly spaced or ‘tiled’ probes complementary to the genome at regular intervals to assay transcription without prior consultation of existing gene annotation (Selinger *et al*., 2000, Yamada *et al.*, 2003, Schadt *et al*., 2004, Stolc *et al.*, 2004). In

general, hybridization-based transcript quantification techniques are subject to cross-hybridization artifacts which lead to high background levels, limited dynamic range and require complicated normalization methods for comparison analysis between experiments (Okonieswski & Miller 2006, Royce & Gerstein, 2007). Collectively, these limitations have inhibited microarrays from providing comprehensive transcriptome datasets.

1.4.1.2 Sequencing-based technolgies: RNA-Seq

Sequencing-based quantification and characterization of the transcriptome effectively circumvants the limitations of the microarrays listed above. Unlike microarray technologies, high-throughput next-generation sequencing (NGS) technologies allow for genome-wide analysis of transcription at single-nucleotide resolution facilitating the identification of novel transcripts, alternatively spliced transcripts and post-transcriptional RNA editting events. Furthermore, NGS technologies do not require prior knowledge of genome sequence and surmount background noise associated with fluorescence quantification used by microarray analysis by using digital gene expression measurements, which provides quantitative transcript information.

Offering single-nucleotide resolution, high-throughput digital quantification data at a genome-wide scale and 'infinite' dynamic range dependent on sequencing depth, massively parallel cDNA NGS sequencing (RNA-*Seq)* has become the most widespread NGS technology used for measuring transcript levels. RNA-*Seq* has been shown to be a highly accurate expression level technology verified by quantitative PCR (RT-PCR) (Nagalakshmi *et al.* 2008). Furthermore, RNA-*Seq* results have shown high reproducibility with technical and biological replicates (Cloonan *et al*., 2008, Nagalakshmi *et al.* 2008, Mortazavi *et al*., 2008). Collectively, these advantages in addition to the lower cost as compared to microarrays, allow the RNA-*Seq* method to be an appealing technique for capturing a snap-shot of the entire reptoire of transcripts within a particular cell context.

RNA-*Seq* involves the sequencing of short cDNA fragments with adapters ligated to one or both ends converted from total or fractionated RNA which can then be mapped to reconstruct an entire transcriptome (**Figure 1.12**) (Wang *et al.,* 2009, Doerr, 2012). RNA-*Seq* library preparation involves the depletion of contaminating DNA, and optionally rRNA (ribosomal RNA), followed by RNA fragmentation,

typically performed using RNAse III enzymatic digestion prior to reverse transcription into cDNA to faciliate sequencing. The quality of fragmented RNA is typically analyzed using a Agilent Bioanalyzer or a urea gel (Schroeder *et al*., 2006).

Short cDNA sequences are then ligated to sequence adaptors, which will bind complementary DNA templates or primers during sequencing. At this point, PCR strategies may be used to amplify starting material (Tang *et al*., 2010). Each cDNA molecule is then sequenced from the adaptor ends either by single- (3' or 5' end) or pair-ended (3' and 5' end) sequencing using cDNA-DNA hybridzation and polymerization to obtain short read sequences (**Figure 1.12**).

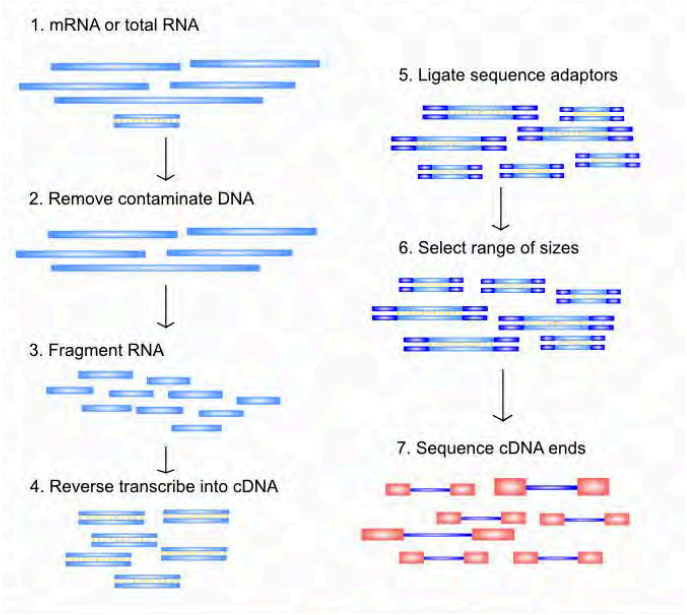


Figure 1.12: **RNA-*Seq* cDNA library preparation.** Total or fractionated (i.e. poly (A)+) RNA extracted is subjected to DNA contaminant removal. rRNA can also be removed. Resultant RNA is fragmented into 200-250 bp and reverse transcribed into cDNA. Double stranded cDNA is then ligated with 3' and 5' sequence adapters. Each molecule with or without amplification is then sequenced from cDNA ends to obtain short read sequences. **Martin & Wang, 2011**.

Several RNA-*Seq* platforms have been developed, including Illumina, Roche 454, Helicos BioSciences and Life Technologies/APG's SOLiD 3 (Metzker *et al*., 2010). Although each of these platforms uses slightly different protocols and detection methods, they each rely on the principle of cDNA-DNA hybridization on a solid support. Helicos Bio-Sciences sequencing uses a one-color cyclic reversible termination readout of cDNA hybridization onto short DNA templates (Braslavsky *et al*., 2003, Harris *et al*., 2008), whereas Illumina sequencing uses a four-color cyclic reversible termination readout corresponding to the four different nucleotides (Bentley *et al.*, 2008). Life/APG's SOLiD 3 platform uses a ligation-based four-color two-base encoded probe sequencing method (Valouev *et al.*, 2008). Roche/454's Titanium uses a pyrosequencing-based method detecting pyrophosphate release on

nucleotide incorporation (Margulies *et al*., 2005) as compared to dye-conjugated nucleotide incorporation detections of the other platforms. In all these platforms, single nucleotide addition events during cDNA-DNA hybridization are measured digitally.

Typically, seqeuncing reads lengths vary between 30-400 bp, dependent on the method of sequencing used (Wang *et al.*, 2009, Metzker, 2010). Notably, the number of reads per transcript is proportional to the relative abundance of the transcript multiplied by transcript length. Therefore, if expressed at equivalent levels, longer transcripts will be sequenced more frequently than shorter transcripts (Trapnell *et al*., 2012). To mitigate this bias, read counts are normalized by quantifying transcript abudance by the fragments or reads per kilobase of transcript per million mapped fragments/reads (FPKM and RPKM respectively) (Mortazavi *et al*., 2008, Trapnell *et al*., 2010).

1.4.1.2.1 RNA-Seq data analysis

RNA-*Seq* data analysis remains computationally intensive and challenging given that the number of reads per sample is in the order of tens of millions, particularly when reads are longer to increase sequencing accuracy. It is relatively easier to quantify known and abundant transcripts. However, common applications for RNA-Seq are the identification and quantification of novel transcripts including long noncoding RNAs (lncRNAs), alternatively spliced transcripts, novel isoforms and polymorphisms.

Notably, with sequencing technologies such as RNA-*Seq,* replicates are crucial to estimate variability within the same experimental condition between treatments. Some analysis packages require technical replicates to eliminate technical artifacts, however little variation has been observed between technical replicates (Bullard *et al*., 2010). However, biological replicates are particularly imperative for reducing false positives. Increasing RNA-*Seq* biological replicates regardless of sequencing depth has been shown to increase the statistical power and accuracy (Liu *et al.* 2013).

1.4.1.2.2 Differential expression analysis & genome mapping

A popular preliminary differential expression analysis RNA-*Seq* pipeline is the Tuxedo suite, which combines Tophat and Cufflinks analysis tools (Trapnell *et al*., 2012). In this pipeline, raw reads are mapped using Tophat and assembled with RPKM or FPKM values using Cufflinks, then combined with reference gene annotations using Cuffmerge and differential expression patterns determined by Cuffdiff and plotted by CummeRbund (Trapnell *et al.,* 2012). Other algorithms including Bowtie (Langmead *et al.*, 2009) RSEM (Li & Dewey 2011) and ERANGE (Mortazavi *et al.,* 2008) have been used to map RNA-*Seq* data. Similarly, several software packages have been designed to date to specifically detect differential gene expression including edgeR (Robinson *et al*., 2009) DESeq (Griffith *et al.,* 2010) and DEGseq (Wang *et al.,* 2010). However, to date no consensus has been reached as to the best algorithm or pipeline to use for optimal RNA-*Seq* data analysis. Although these algorithms and pipelines are adequate for preliminary RNA-*Seq* data expression annotation and quantification, no consensus algorithms and pipelines for identifying genes of interest and/or network analysis of annotated data have been developed.

1.4.1.2.3 Predicting lncRNAs from RNA-Seq data

To date, several validated and characterized mammalian lncRNAs have been identified from RNA-*Seq* data including linc-Cox2 (Carpenter *et al.*, 2013), Lethe (Rapicavoli *et al*., 2013), IL1β-eRNA and IL1β-RBT46 (Illot *et al*., 2014). Although different computational analysis pipelines identified each of these lncRNAs, common features in lncRNA identification from RNA-*Seq* data involved key features described in John Rinn and Howard Chang's (2012) functional discovery pipeline. These include: differential expression analysis, coding potential determination, chromatin signatures, "guilt by association", functional assays to determine lncRNA functionalities and mechanisms (**Figure 1.13**).

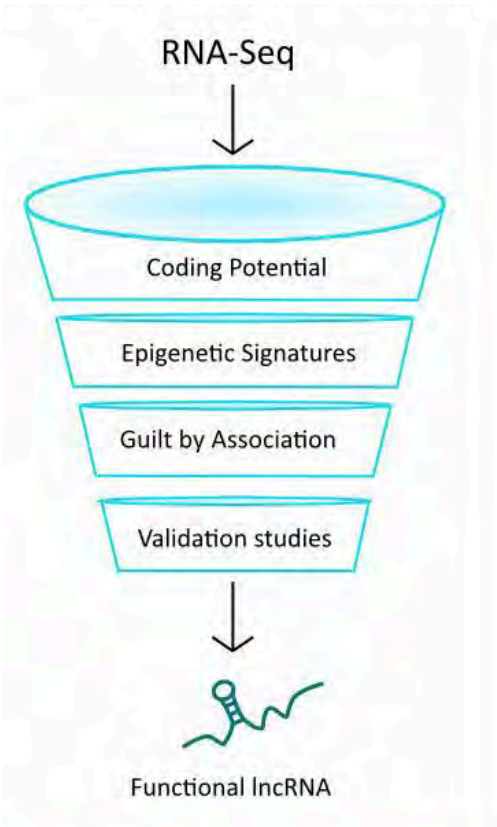


Figure 1.13: **Functional lncRNA discovery pipeline.** A combination of bioinformatic tools and functional assays identifies lncRNA functionalities and mechanisms within a specific biological context.. Adapted from **Rinn & Chang (2012**)

In current literature, no consensus pipeline for lncRNA discovery has been identified. However, studies aimed at identifying lncRNAs, have used an integrative approach, which combined RNA-*Seq* with Chip-*Seq* (chromatin immuno-precipitation sequencing) data (Rapicavoli *et al.,* 2013 and Illot *et al.*, 2014) or RT-PCR (quantitative PCR) and RNAi (RNA interference) (Carpenter *et al.,* 2013). Although RNA-*Seq* data analysis can adequately identify noncoding transcripts and may infer functional relationships by guilt by association, it fails to provide any other useful information about lncRNA functionality, localization, interacting partners and cellular targets. Thus, a complete lncRNA discovery pipeline from RNA-*Seq* data, as described in Rinn and Chang's pipeline, requires functional assays for lncRNA validation and characterization.

1.5 *Listeria monocytogenes*: a model bacterial pathogen

1.5.1 Listeriosis

Listeria monocytogenes is a gram-positive, intra-cytosolic opportunistic food-borne pathogen. *Listeria* enters its mammalian hosts through nonphagocytic epithelial gut cells by induced phagocytosis. Following initial infection, the bacterium can spread into the lymph nodes, enter the blood stream where it can infect a variety of cell types including macrophages, hepatocytes, splenocytes and neurons (**Figure 1.12**). Listeriosis is facilitated by the bacteria's ability to cross host barriers: intestinal, blood-brain and placental. This results in severe systemic illness including meningitis, meningo-encephalities in immune compromised individuals, as well as materno-fetal infection, fetal and neonatal pathologies (Farber and Peterkin, 1991, Schlech, 2000).

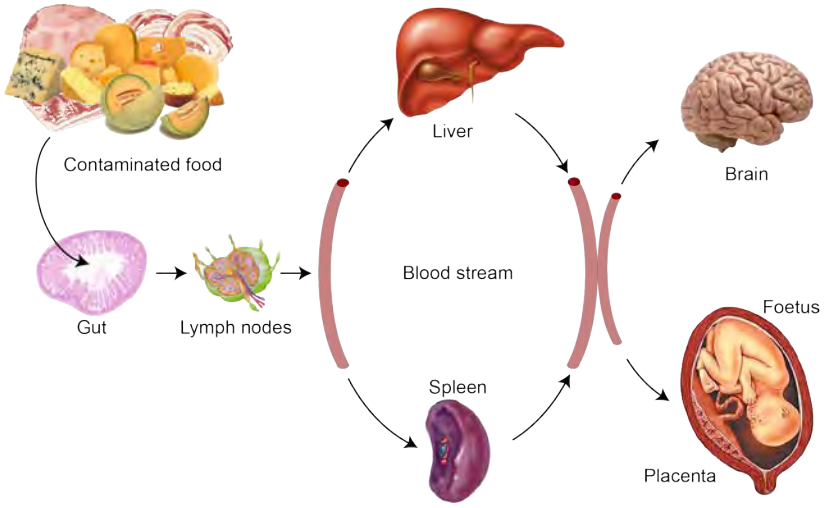


Figure 1.14: **The progression of listeriosis.** *Listeria monocytogenes* enters the mammalian host through epithelial cells in the gut progressing into the blood stream via the lymphatic system. Once in the blood stream, *Listeria* can infect and replicate in hepatocytes, splenocytes and cross the blood-brain and placental barriers infecting leading to severe illness such as splenomegaly, septicemia, meningitis, abortive pregnancy, fetal and neonatal pathologies in immune-compromised individuals. **Gouin *et al.* 2005.**

1.5.2 The lifecycle of *Listeria monocytogenes*

The lifecycle of this pathogen in various cell types consists of four main stages: (1) bacterial entry, (2) phagosomal escape, (3) actin-based motility, and (4) cell – to – cell spread (**Figure 1.15**).

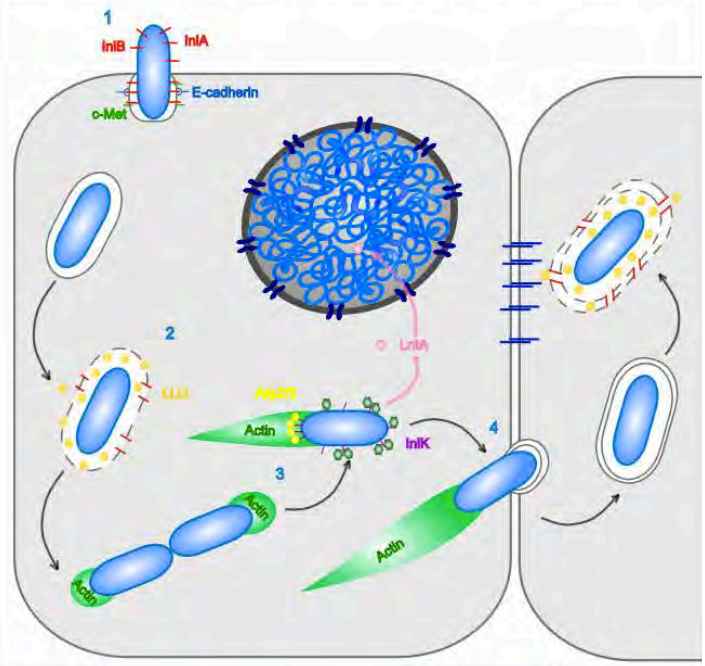


Figure 1.15: **The intra-cellular life-cyle of *Liseria monocytogenes.*** *Listeria* enters host cells using host-pathogen receptor-ligand binding interactions of E-cadherin-InlA and InlB-cMet in non-phagocytic cells. These interactions promote bacterial internalization forming a bacterial-containing phagosome. *Listeria* then escapes the phagosome by secreting pore-forming lysin LLO which disintergrates the phagosomal membrane allowing the bacterium to enter the host cytoplasm. Once in the cytoplasm, the bacteria replicate and use a variety host mimicking bacterial effectors including Arp2/3 and InlK to hijack the host cellular actin polymerization machinery, which permits intra- and inter-cellular bacterial motility. **Cossart and Toledo-Arana, 2008.**

1.5.3 Bacterial entry into host cells

Listeria internalization results from the tight encircling of the host plasma membrane around an extracellular bacterium, which is mediated by the interaction of bacterial surface ligands with host cellular receptors (Cossart and Toledo-Arana, 2008). This mechanism of entry, termed the “zipper mechanism”, results in the formation of intracellular vacuole-encased bacterium.

Two independent ligand-receptor binding and internalization pathways for *Listeria* entry into host cells have been observed. The first involves the binding of bacterial internalin A (InlA) to host cell receptor E-cadherin (**Figure 1.15**). This interaction

involves the N-terminius of the E-cadherin ecto-domain and the InlA N terminal leucine-rich repeats (Schubert *et al.*, 2002). The second ligand-receptor interaction for bacterial entry occurs with the internalin B (InlB), which interacts with host cell Met, a tyrosine kinase, and hepatocyte growth receptor (HGF) receptors (Shen *et al.*, 2000). InlB, as InlA, uses a leucine-dense repeat domain to interact with the first immunoglobulin-like domain and the transmembrane Met domain. InlB-Met binding facilitates bacterial entry via p13-kinase activation (Cossart, 2001).

1.5.4 Phagosomal escape

A key feature in *Listeria* pathogenesis is the escape of phagosome-engulfed bacteria into the cell cytosol. *Listeria* is typically entrapped within the phagosomal vacuole for approximately 30 min following bacterial entry (Cossart and Toledo-Arana, 2008). During this time, *Listeria* sets into play several strategies to resist and escape early intra-phagosomal killing. Phagosomal escape is largely mediated by bacteria-secreted listeriolysin O (LLO), a cholesterol-dependent pore-forming cytolysin, encoded by the *hly* gene, which ruptures the phagosomal membrane forming pores thereby facilitating bacterial escape into the cytosol (**Figure 1.15**) (Schnupf and Portnoy, 2007). LLO pore formation begins with the low-pH dependent oligomerization of cholesterol-bound monomers into a prepore complex proceeded by insertion into the vacuole lipid bilayer (Tweten, 2005; Schuerch et al., 2005).

Although LLO is the major factor required for bacterial escape, several other bacterial factors are required for efficient LLO-dependent phagosomal escape. In cells, in which LLO-dependent pore-formation is inefficient, pore forming vacuole lysis is performed by bacterial phospholipases C (PLC) or in synchrony with LLO (Smith *et al.*, 1995, Marquis *et al.*, 1995). Other factors involved in phagosomal escape include cytoplasmic *Listeria* superoxide dismutase (MnSOD), which is thought to inhibit phagolysosomal degradation (Archambaud *et al.,* 2006). PgdA, a *Listeria*-encoded enzyme, prevents lysozyme-mediated killing of *Listeria* within the phagosomal vacuole by mediating the deacetylation of the N-acetylglucosamine residues of the bacterial peptidoglycan thus conferring increased sensitivity to lysozyme *in vitro* (Boneca *et al.*, 2007). Recently, a chloride channel, cystic fibrosis transmembrane conductance regulator (CFTR), was shown to increase intra-vacuole chloride concentrations in murine macrophages thus potentiating LLO pore formation and vacuole lysis (Radtke *et al*., 2011).

1.5.5 Actin-based motility

Once in the host cytoplasm, *Listeria* undergoes cellular division and growth before host actin filaments begin to form dense coating on the surface of the bacteria referred to as an “actin cloud”. Subsequently, due to the non-polarized ActA secretion and septal-originating cell wall growth, this “actin cloud” as well as ActA localization becomes polarized to one end of the bacterium. This results in the formation of a polarized comet tail consisting of a dense array of cross-linked actin filaments with their barbed ends at the actin-bacterium interface (**Figure 1.16**).

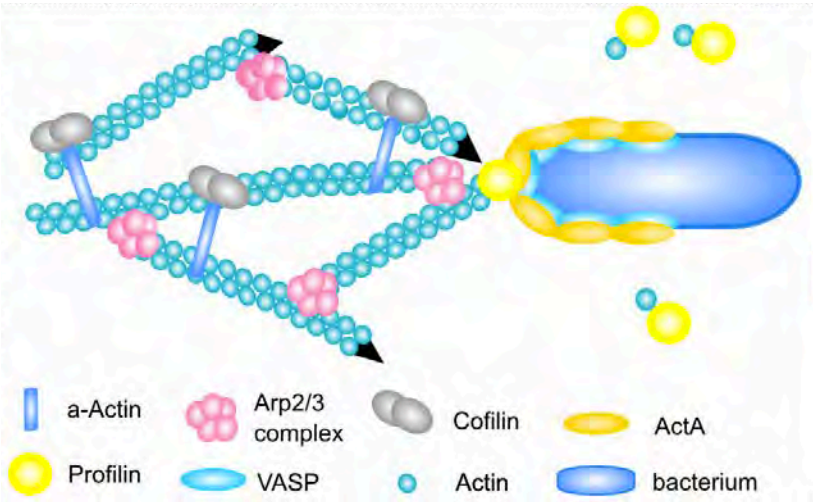


Figure 1.16: **Molecular components required for actin-based motility of *Listeria monocytogenes*.** **a.** Interactions between host-cell proteins and ActA at the bacterial surface.The ActA N-terminal domain activates actin filament nucleation through Arp2/3. The central proline-rich domain binds VASP which binds profilin and F-actin to promote filament elongation. **b.** Host proteins in the Listeria comet tail. α-actinin crosslinks and stabilizes filaments, and ADF/cofilin disassembles old filaments increasing the G-actin monomer pool. **Cameron *et al.*, 2000**

The *Listeria* protein, ActA binds host factors via a variety of motifs facilitating host protein interactions to regulate actin dynamics. Specifically, ActA mimics host actin filament nucleation promoting factors, zyxin–vinculin and WASP-Wave (Wiskott-Aldrich syndrome family) proteins and constitutively activates actin polymerization proteins Ena-VASP (Enabled/vasodilator-stimulated phosphoprotein) and Arp2/3 (Actin related protein 2 and 3) (**Figure 1.16**). (Skoble *et al*., 2000). The ActA N-terminus consists of three domains including a VCA (verprolin-cofilin-acidic) domain, which mimics the Arp2/3- and actin binding activity of host cell actin nucleation-promoting factors (Welch *et al*., 1998, Takenawa *et al*., 2007). A second ActA N-terminal domain, the cofilin homology domain (CHD), similar to that in WASP proteins, also binds the host Arp2/3 complex and is essential for actin nucleation (Bi

and Zigmond, 1999, Skoble *et al.*, 2000). ActA-Arp2/3 binding activates Arp2/3–mediated actin nucleation through its CHD (Skoble *et al*., 2000). The third ActA N-terminal domain, a central proline-rich repeat region, binds VASP, which in turn binds F-actin and profilin, an actin monomer exchange factor (**Figure 1.16**). The combination of these components at the bacterial surface facilitates actin filament/comet tail elongation. Perpetuated actin nucleation and polymerization at the actin-bacteria interface results in a high-speed (up to 30nm/s) propulsive force pushing the bacterium both inter- and intra-cellularly (Dabiri *et al*., 1990, Robbins *et al*., 1999).

1.5.6 Cell-to-cell spread

Listeria cell-to-cell spread can be divided into main three stages: (1) actin-based motility and filopodia formation; (2) bacterial uptake and formation of the double-membrane spreading vacuole and (3) membrane dissolution (Alberti-Segui *et al*., 2007). However, the detailed molecular mechanisms of cell-to-cell spread remain elusive with studies on this matter limited to macrophage cell types.

The actin polymerization of *Listeria* comet tails is essential for the formation and release of bacteria-containing membrane-encased protrusions extending from infected cells into neighboring cells without exposure to the extracellular matrix (**Figure 1.15**) (Tilney and Portnoy, 1989). The formation of these filopodia appears to be dependent on ezrin-radixin-moesin (ERM) family proteins ezrin and CD44, which serve as intermediates between the actin cytoskeleton and transmembrane proteins (Pust *et al*., 2005). The propulsive force generated by actin polymerization at the distal end of the bacterium, allows the bacterium to penetrate the neighbouring cell membrane (Tilney and Portnoy, 1989, Robbins *et al*., 1999).

Listeria has been shown to require exofacial phosphatidylserine (PS)-presenting primary infected cell protrusions binding to the TIM4 receptor to induce efferocytosis, to effectively penetrate and enter the neighboring cell (Czuczman *et al*., 2014). This leads to bacterial uptake resulting in the formation of a double membrane secondary vacuole within the secondary host cell (**Figure 1.15**). An investigation by Alberti-Segui and colleagues suggested early endosome fusion with the double membrane vacuole activates PC-PLC-mediated dissolution of the inner vacuole membrane and subsequent late endosome and vacuole fusion activates LLO-mediated outer vacuole membrane disintegration. Although this model has not

yet been validated, collective studies point towards a synchronous PC-PLC- and LLO-mediated secondary vacuole escape mechanism in macrophages and epithelial cells (Marquis and Hager *et al*., 2000, Gedde *et al*., 2000, Alberti-Segui *et al*., 2006, Czuczman *et al*., 2014). Cell-to-cell spread of Listeria infection requires more investigation to determine the molecular mechanisms governing the steps of secondary infection.

Chapter 2: Aims and Objectives

2.1 Aims

This study aims to identify and investigate *Listeria*-mediated alterations in the HeLa cell transcriptome, specifically the lncRNA-ome, and nuclear architecture in an attempt to identify novel lncRNAs regulating *Listeria* infection.

The aims of this study are:

- To validate RNA-*Seq* data of *Listeria*-infected HeLa cells;
- To identify potential lncRNAs from RNA-Seq data of *Listeria*-infected HeLa cells;
- To validate the functional relevance of these identified lncRNAs by knockdown experiments in *Listeria*-infected HeLa cells;
- To determine potential modes of action of lncRNA candidates.

2.2 Objectives

At the time of the inception of this study, no lncRNAs functioning during microbial infection had been characterized. However, in the past year alone a handful of studies have identified lncRNAs linc-Cox2 (Carpenter *et al*., 2013), Lethe (Rapicavoli *et al.,* 2013), PACER (Krawczyk *et al.*, 2014), THRIL (Li *et al.*, 2014) and NeST (Gomez *et al.* 2014) as central players in host cell innate immune response against microbial infection. These initial discoveries validate the main aim of this study, which is to identify novel *Listeria* infection-regulating lncRNAs in epithelial HeLa cells.

We deep sequenced the transcriptome of *Listeria*-infected HeLa cells at specific time points marking the different hallmarks of the bacterial intracellular lifecycle. The filtration, mining and analysis of this data to identify novel infection-regulating lncRNAs using a variety of platforms and pipelines made up a bulk of this work. This allowed for the formulation of hypotheses on lncRNA functioning during *Listeria* infection. To test the hypotheses formulated, we used TALEN-mediated knockdown studies of lncRNA candidates to determine their *Listeria* infection regulating capabilities.

Chapter 3: Methodologies

3.1 RNA-*Seq* of *Listeria*-infected HeLa cells

(RNA Samples extracted by Dr. Y. Shibayama, Dr. Suraj Parihar & Dr. Reto Guler and RNA-Sequencing performed by BGIAmericas Technologies)

Hela cells were grown in T75 flasks for RNA extraction and on coverslips in 12 well plates for microscopy and CFU (colony forming unit) analysis of the infection time course. Cell in T75 flasks and 12 well plates were infected with non-fluorescent *Listeria* EDGe at an moi of 50 in 10ml and 0.4 ml infection media respectively. *Listeria* strains used for the infection were non-fluorescent wild type EDGe strain and the Δ*LLO* knockout strains grown to an optical density 600nm=0.6 equating to $2X10^6$ CFU/ml. Cells were infected for time points: 20, 60, 120 and 240 min. Prior to infection, bacteria was washed X2 in PBS (Gibco, Life Technologies) X1 in DMEM (Gibco, Life Technologies) and X1 in DMEM+10% FBS (Biochrom, Merck Millipore) media.

Infections were conducted with a pre-extraction wash at 60 min poi to remove extracellular bacterium. This wash step consisted of X2 PBS washes followed by X1 DMEM-10% FBS media wash. For time points longer than 60 min (i.e. 120 and 240 min), cells were re-incubated until the end of the set infection time points. Once the infection time point was reached, cells underwent a second extraction-wash which included X2 PBS washes prior to RNA extraction.

3.1.1 CFU analysis

Following the end of the relevant time poi, cells in 12 well plates were washed in X2 PBS, followed by a 10 min incubation in 1ml 0.1% Triton-X-200 (Sigma Aldrich) at room temperature. Cells were further lysed by vigourous pipetting then collected and serial diluted to the following dilutions: 10^{-1}, 10^{-2}, 10^{-3}, and 10^{-4} in PBS. 5 ul of each of the dilutions was then plated on to trypic soy agar (TSA, Sigma Aldrich) plates. The CFUs were counted and calculate 24 hrs later.

Bacterial uptake was calculated as follows:

$$\frac{Infection\ CFU}{No\ of\ infected\ cells} \times 100\ = \%\ bacterial\ uptake\ \ (1)$$

where the

$$CFU/ml\ \times\ Infection\ volume = Infection\ CFU\ (2)$$

The rate of replication in infected cells was calucated as time taken for bacterial doubling events:

$$\frac{CFU\ at\ 4\ hr\ poi}{CFU\ at\ 1\ hr\ poi} = no.of\ doubling\ events\ in\ 3\ hrs\ (3)$$

$$\frac{no\ of\ doubling\ events\ in\ 3\ hrs}{180\ min} = \min for\ a\ single\ \ replication\ event\ (4)$$

3.1.2 RNA Extractions

Following the extraction washes, 1.5 ml of PBS was added to cells then cells were scraped off the bottom of the flask and total RNA was extracted using the Ambion mirVana kit as per manufacturers specifications. Eluted RNA was stored at -80°C until futher use. 10% (10 ul) of the eluted RNA, frozen separately, was used to determine concentration and integrity. RNA concentration, 260/280 and 260/230 ratios were measured on the Nanodrop ND-1000 spectrophotometer (Nanodrop technologies) using 2 ul of the eluted RNA. RNA intergrity was determined by running 8ul eluted RNA on a 12% urea gel with a 2ul reference RiboRuler Low Range RNA (Thermo Scientific) ladder.

3.1.3 Staining and microscopy analysis

Listeria-infected HeLa cells grown on 12 well coverslips were fixed at the relevant time poi in 4% formaldehyde in PBS for 10 min at room temperature, washed twice PBS and stored in PBS overnight. Cells were permeabilized in 0.2% Triton X-100 (Sigma Aldrich) in PBS for 10 min at 37°C and washed twice in PBS. To visualize *Listeria* bacteria, coverslips were then washed twice in PBS, blocked with 1% BSA in PBS at room temperature for 30 min prior to antibody staining. Following BSA blocking step, cells were incubated with 1 ng rabbit polyclonal anti-*Listeria* antibody (Abcam) for 1 hr, followed by two PBS washes then incubated with 1ng donkey anti-

rat 488 antibody (Abcam) for 30 min. Stained cells were then washed twice PBS before being incubated with 10 nM phalloidin 565 (Attotec, Germany) for 10 min, then washed twice PBS again and finally stained with 1ng DAPI (Life Technologies) for 5 min. Coverslips were then mounted in VectaShield (Vector Labs) mounting medium and imaged at 60X magnification using a widefield Nikon microscope. Images were processed using ImageJ (NIH, USA).

3.1.4 BGIAmericas RNA-Seq

(Performed by BGIAmericas)

RNA samples sent to BGIAmericas were subjected to BGIAmericas' experimental pipeline (**Figure 1.12**). Oligo (dT) beads were used to isolate and enrich poly(A) mRNA followed by the addition of a fragmentation buffer to fragment mRNA. First-strand cDNA synthesis was performed using random hexamer primers followed by second-strand cDNA synthesis. Fragmented RNA was purified with QiaQuick PCR extraction kit and resolved with EB for end reparation and poly(A) addition. Sequencing adaptors were then added to short cDNA fragments which were then amplified by PCR and sequenced using the Illumina HiSeq™ 2000 (**Figure 1.12**).

BGIAmericas' single-end sequencing generated 123-130 raw million reads per sample. Reads with adaptors, reads with unknown nucleotides larger than 5% and reads with low quality (more than half of the bases' qualities are less than 5) bases were removed to obtain clean reads. Clean reads were then mapped to NCBI Build 36.1 reference human genome using SOAP2 (Short Oligonucleotide Analysis Package 2) (Li et al., 2009) resulting in 9-12 million mapped reads per sample which tabulated to 74-85% of the raw reads (**Figure 3.1**).

Mapped reads were then assessed on the basis of mismatches where reads with more than 5 nucleotide mismatches were filtered out. Gene expression annotation and differential expression analysis was performed using an algorithm developed by BGIAmericas where differential expression was determined using statistical functions including the p-value, FDR (false discovery rate) and RPKM ratios per condition or in this case per time point post infection. This was the final subset of data used for subsequent analysis in this project although BGIAmericas also performed GO (gene onotology), KEGG (Kyto encyclopedia of genes and genomes) pathway, SNP (single nucleotide polymorphisms), alternative splicing analysis.

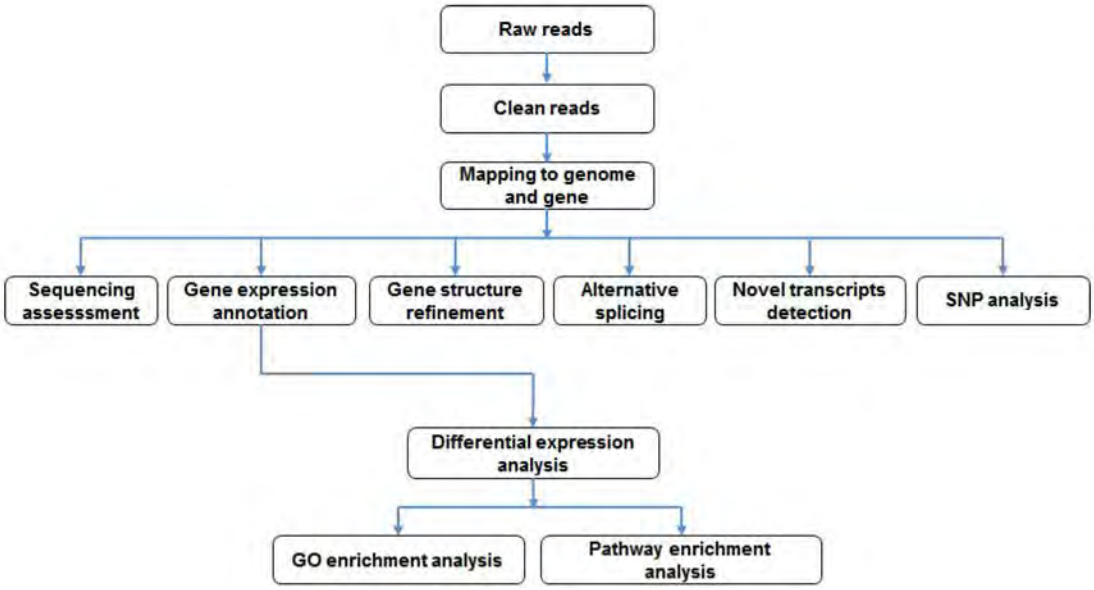


Figure 3.1**: BGIAmericas pipeline of bioinformatic analysis.**

In addition to this subset, the novel transcripts data predicted by BGIAmericas was also used for subsequent analysis. Novel transcripts were predicted as gene models found in intergenic regions (200 bp away from upstream or downstream genes) which were between 10-16 000 transcripts per sample or time point post infection. Notably, some novel transcripts within the ncRNA dataset were support vector machine predicted. However, it is unclear what algorithm BGIAmericas and what parameters were used for these predictions.

3.2 Bioinformatics: LncRNA discovery

3.2.1 Strategy 1: Differential expression of mRNA during Listeria infection

3.2.1.1 mRNA heat map

Although a limited number of lncRNAs have been fully characterized to date, many have been shown to regulate the expression of proximal genes. Thus, the first lncRNA discovery strategy employed from the BGIAmericas data received involved the determining significantly differentially expressed genes during *Listeria* infection. Simply put, the annotated and poly-A tailed mRNA component of the transcriptome data was filtered based on calculated fold changes and relative expressions. The fold change ratio used describes the fold change log2 ratio of gene expression in infected verses non-infected data (5). It has been refered to as the *M-value* as it has been widely used in mRNA expression microarry analysis (Allison et al., 2006).

$$M = \log_2 \frac{RPKM_{infected}}{RPKM_{uninfected}} \quad (5)$$

The *A-value,* which describes the relative expression values was also used to filter the mRNA component of the transcriptome data (6).

$$A = \frac{1}{2}\log_2\left(RPKM_{uninfected} \times RPKM_{infected}\right) \quad (6)$$

Genes with *M-values* ≥0.5 and ≤-0.5 and *A-value* ≥7.8 at any of the four time points or two infection conditions were then compiled into a heat map. In an attempt to identify lncRNAs that were potentially be regulating the expression of the genes determined by the heat map, the 100kb genomic regions surrounding heat map-identified genes were examined for proximal lncRNAs in the RNA-*Seq* ncRNA data set. Particularly lncRNAs with correlative expression patterns to proximal genes. Thus, CHIP-*Seq* peaks/traces within 100kb of the lncRNA loci in the UCSC hg19 ENCODE genome browser (https://genome.ucsc.edu) were explored. The traces surveyed included the characterized lncRNA 'K4-K36 domain' (Khalil et al., 2009) H3K4me3, H3K36me3, active transcription mark Pol II Std and enhancer associated-marks H3K27ac and p300.

3.2.2 Strategy 2: Differential expression of LncRNAs during Listeria infection

To begin lncRNA discovery from the RNA-*Seq* data, M- and A-values for the ncRNA dataset were calculated to determine differential expression as well as transcript abundance. These values were then plotted into MA (M value vs. A value) plots which are often used in microarray studies to determine the data distribution as well as detect technical artifacts. Within the ncRNA dataset obtained from BGIAmericas ncRNA isoforms were labeled as different transcripts. Thus, prior to differential expression filtration, each isoform belonging to a specific ncRNA was represented by the same ncRNA ID. Following which the M-values of the ncRNAs across the different time points and conditions were plotted to ascertain a particular infection or time point condition of interest containing the largest number of differentially expressed ncRNAs. The MA plots were also used to determine the infection condition of interest, which was used for subsequent differential expression analysis.

3.2.2.1 Microarray-based filtration of lncRNAs

Currently, no standardized lncRNA data filtration/threshold parameters exist in literature. This is largely due to the the low abundance of lncRNAs even in clonal cell populations with many lncRNAs expressed as low as <1 copy per cell. Thus, initially the lncRNA data subset was filtered based on standard microarray data filtration parameters *M-value* ≥0.5 and ≤-0.5 and *A-value* ≥10 at each time point and infection condition. These strigent parameters were bias towards highly abundant transcripts with large changes in fold expression. As in previous strategies, the genomic loci of these filtered lncRNAs were investigated on the UCSC genome browser for characteristic histone and chromatin modification signals.

3.2.2.2 The hunt for enhancer RNAs

3.2.2.2.1 Signature chromatin modification marks

Next-generation sequencing data has shown that the expression of lncRNAs and their target gene transcripts is highly correlative. However, the number of detected lncRNAs is typically over 20-fold lower than detected mRNA transcripts (Andersson et al., 2014). Thus, subsequent ncRNA filteration strategies in this study excluded

relative expression values (*A-value)* and only used fold-change expression values (*M-value*s).

Recent studies have identified several lncRNA-associated chromatin modification signatures linked to different classes of lncRNAs. Guttman *et al.* (2009) identified a chromatin modification signature associated with thousands of lncRNAs in various cell types identical to that of actively transcribed genes. This signature termed the 'K4-K36 domain', consists of a promoter-corresponding short H3K4me3 region followed by a longer H3K36me3 region corresponding to the transcribed gene body. In addition to the K4-K36 domain, recent data has suggested enhancer associated lncRNAs (eRNAs) genomic loci are typically associated with enhancer chromatin marks including p300, H3K27Ac. eRNAS are also thought to have a unique chromatin signature with H3K4me1 occupancy and H3K4me3 exclusion (De Santa et all., 2010, Marques et al., 2013).

In the hunt for eRNAs, significantly up-regulated (M value>0.5, sheet 1) and downregulated (M value<-0.5, sheet 2) were identifed and their H3K4me3-H3K4me1 ratios were investigated on ENCODE. Notably, svm-predict (support vector machine predicted) transcripts were excluded from analysis.

3.2.2.2.2 Gene ontology

In addition to expression analysis and hallmark chromatin signatures of eRNAs, the gene ontology GOSTATS data of the mRNA data subset provided by Joana Cruz was used to identify potential eRNAs in the study. This strategy involved identifying a selection of gene ontologies involved in the immune response including *response* to bacterial origin, innate immune response. Following the identification of the gene ontologies differentially expressed genes within these ontology classes were analyzed.

The genomic loci of the identified genes were then analysed in the UCSC genome browser an attempt to identify characteristic lncRNA and eRNA traces within 100kb of genes within the GOSTATS data. The genomic loci of these regions were then used to search for ncRNAs in the RNA-*Seq* data.

The gene ontologies studied involved immune response ontologies:

response to stimulus, response to stress, cellular response to stimulus, response to external stimulus, regulation of gene expression, response to bacterial origin, response to bacterium, inflammatory response, response to cytokine stimulus, toll-like receptor 3 signaling pathway, toll-like receptor 2 signaling pathway, tumor necrosis factor superfamily cytokine production, regulation of inflammatory response, cellular response to extracellular stimulus, toll-like receptor 4 signaling pathway, tolerance induction to lipopolysaccharide, negative regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, toll-like receptor 5 signaling pathway, negative regulation of toll-like receptor 5 signaling pathway, positive regulation of immune system process, negative regulation of antigen processing and presentation, positive regulation of chemokine mediated signaling pathway, cellular response external stimulus, lipopolysaccharide-mediated signaling pathway, innate immune response-activating signal transduction, activation of innate immune response, positive regulation of toll-like receptor 2 signaling pathway, negative regulation of toll-like receptor 4 signaling pathway, immune system process, positive regulation of innate immune response, regulation of immune system process, regulation of interleukin-6 production, interleukin-6 production, regulation of cytokine production, regulation of B cell activation, cellular response to lipopolysaccharide, cytokine production, immune response-activating signal transduction, cytokine biosynthetic process, immune response-regulating signaling pathway, positive regulation of cytokine-mediated signaling pathway, negative regulation of interleukin-6 production, regulation of tumor necrosis factor biosynthetic process.

3.2.2 Stratgey 3: LncRNA barcode

In an attempt to identify a differential expression lncRNA barcode for early *Listeria* infection in this cell line, the ncRNA *Lm*-WT 60 min poi data was filtered under strigent parameters. NcRNAs were filtered under the following parameters: -0.5<M value>0.5, >200 bp, >100 rpkm/million cells. In addition to the exclusion of svm-predicted transcripts, intragenic ncRNAs that span more than 2 exons and interagenic ncRNAs with similar M values to their parental transcripts were excluded from this data set. These filteration exclusions were not applied to antisense ncRNAs. NcRNAs whose parental transcript expression values could not be found in the data were also excluded as their differential expression cannot be compared to their parental transcripts. This filteration strategy would yield significantly and

differentially expressed ncRNAs whose expression patterns can be easily compared to their parental transcripts in the case of intragenic ncRNAs. In addition, the likelihood of misclassifying splice variants as lncRNAs is reduced. Thus, the ncRNAs identified by this strategy are amicable to wet lab validation experiments such as RT-PCR analysis.

3.3 Wet lab experiments

3.3.1 Molecular Biology

3.3.1.1 Transcription activator-like endonucleases (TALENs)

All TALENs used in this project were designed and constructed according to the Zhang protocol (Sanjana et al., 2012) with minor modifications.

Amplification and normalization of monomer library with ligation adaptors for 18-mer TALE DNA-binding domain construction. PCRs were performed with forward and reverse primer mixes for each of the 18 positions mixed to a final concentration of 10 µM for each primer. PCRs* were conducted using the following parameters and final concentrations: 50 pg/µl monomer template plasmid, 1 mM dNTPs (Thermo Scientific), 1X Herculase II PCR Buffer (Agilent Technologies), 200 nM primer mixes, Herculase II Fusion polymerase (Agilent Technologies), and appropriate volumes of dddH_2O Monomer amplifications were verified by gel electrophoresis** (1µl) displaying ~150bp products with the exception of monomers positioned at the ends of each hexamer (monomers 1,6,7,12,13,18) which were ~170 bp. Monomer bands were excised from gel and purified using Zymogen gel extraction kit according to manufacturers specifications and resuspended in 0.5M Tris (Sigma Aldrich). Monomer concentrations were normalized gel electrophoresis** images (2 µl per monomer, imaged for 40 ms using a Gbox Chemi HR16 system where product bands were quantified by ImageJ (NIH, USA) gel quantification plugin and with a Qubit 2.0 fluorometer (Molecular Probes) according to manufacturers specifications. Purified monomers were diluted to 15 ng/µl with the exception of monomers 1,6,7,12,13,18 which were adjusted to 18 ng/µl and stored at -20 °C and used for all subsequent TALE assemblies.

Construction of custom 20-bp-targeting TALES. Six appropriate monomers (1 µl of library each) were pooled into hexamers of designed TALEs (3 hexamers per TALE) and used for GOLDEN Gate reactions to assemble each hexamer. Golden Gate reactions were done on a BioRad T100 Thermal Cylcer; digestion 37°C, 5 min and ligation 20°C, 5 min for a total of 15 cycles. Golden Gate reactions were carried out using final concentrations 0.375 U/µl BsmBI (Fermentas), 1X Fast digest buffer (Fermentas), 1 mM DTT, 75 U/µl T7 ligase (Fermentas), 1 mM ATP (Fermentas) per

hexamer assembly. ~700 bp ligation products were verified by gel electrophoresis**. Noncircular (incomplete) ligation products were selectively degraded using 0.66 U/µl PlasmidSafe (Epicentre) exonuclease in 1X PlasmidSafe reaction buffer (Epicentre) and 1 mM ATP per hexamer for 30 min at 37°C. PlasmidSafe exonuclease was inactivated at 70°C. PlasmidSafe-treated hexamers were amplified with hexamer forward and reverse primers (Hex-F and Hex-R) using Herculase II Fusion DNA polymerase. Entire volume of PCR products were verified by gel electrophoresis and ~650 bp bands were excised and purified using the Zymogen gel extraction kit according to manufacturers specifications. Hexamer concentrations were determined using the Qubit 2.0 fluorometer according to manufacturers specifications and adjusted to 20 ng/µl in TE buffer (pH 8.0). For hexamer assembly into appropriate TALEN backbone vector a second Golden Gate reaction was carried out using 10 ng/µl TALE backbone vector, 1.5 U/µl BsaI-HF (New England Biolabs), 1X NEBuffer 4 (New England Biolabs), 1X BSA (New England Biolabs), 1 mM ATP (New England Biolabs) and 75 U/µl and 2 ng/µl of each of the three appropriate hexamers. Golden gate reactions were carried out using the following cycling conditions: cycles 1-20 (digest: 37°C, 5 min; ligate 20°C, 5 min) and enzymes were inactivated at 80°C for 20 min using a BioRad CFX96 Thermocycler. Ligation products of ~ 1.8 kbp were verified by gel electrophoresis** on a 1% agarose gel.

Verification of correct TALE repeat assembly. Ligation products (5 µl) were transformed into 50 µl of ice-cold competent Stbl3 *E. coli* (Life Technologies) cells incubated on ice for 5 min, incubated at 42°C for 45 s, incubated on ice for 5 min followed by the addition of 500 µl of LB (Lysogeny broth, Life Technologies) medium containing 100 ug/ml ampicillin (Sigma Aldrich) and incubated at 37 °C for 1 hr on a shaking (250 rpm) and plated on LB containing 100 ug/ml ampicillin plates and incubated at 37°C overnight. Colonies (8-20) were picked using a sterile 20 µl pipette tip and streaked on a new gridded LB-amp plate to save colony then tip was swirled in 100 µl of $dddH_2O$ to dissolve colony for subsequent PCRs. Gridded plates were incubated at 37 °C overnight. Colony PCRs* were performed using 0.02 U/µl DreamTaq (Fermentas), 100 mM each of TALE-Seq-F1 and TALE-Seq-R1 primers. Colony PCR products (1 µl) were checked on 1% agarose (Sigma Aldrich) gel using gel electrophoresis** for single band of 2175bp and product concentrations were determined using ImageJ (NIH, USA) gel quantification plugin. 100 ng of each TALE was also digested using AfeI (Fermentas) at 37°C for 10 min and AfeI digestion products were verified by gel electrophoresis** on a 1% agarose gel to produce 165,

2 118, 2 803 and 3 236 bp bands for TALENs. Clones with the correct band sizes were inoculated from gridded plate into 3 ml LB with 100 ug/ml ampicillin at 37 °C in shaking incubator (250 rpm) overnight. Plasmid DNA from overnight cultures were isolated using Machenry Nagel Xtra Maxi kit according to manufacturers specifications and resuspended in TE buffer (pH 8.0). Plasmid concentrations were verified using the Nanodrop ND-1000 spectrophotometer (Nanodrop technologies) as well as by gel electrophoresis and ImageJ (NIH, USA) gel quantification. Plasmid concentrations were adjusted to 100 ng/µl in TE buffer and stored at -20°C until transfections were performed.

**** Polymerase Chain Reaction***

All PCR reactions were performed under the following conditions, unless stated otherwise, using a BioRad T100 Thermal Cylcer:

Cycle number	Denature	Anneal	Extend
1	95 °C, 2 min		
2-36	95 °C, 20s	60 °C, 20 s	72 °C, 30 s
37			72 °C, 3 min

Final concentrations for PCRs were 200 nM for each primer and 1 mM for each dNTP.

*****Gel electrophoresis***

All gel electrophoresis experiments were performed under the following conditions, unless stated otherwise,: Samples were diluted in 2-5 µl of Loading dye (Fermentas) and run on a 2% (w/v) agarose (Sigma Aldrich, Germany) in 1X TBE (Sigma Aldrich, Germany) with 10 ng/µl ethidium bromide (BioRad) and run in cold 1X TBE (Sigma Aldrich, Germany) using the BioRad PowerPac universal power supply.

TALEN transfections. TALEN transfections were performed on HeLa cells growing on 6 well plate or coverslips in 12 well plate 24 hr after seeding. Transfections were carried out using 1 ug of total plasmid (500 ng left TALEN and 500 ng right TALEN) using Lipofectamine 2000 (Life Technologies) or Lipofectamine RNAi-MAX (Life Technologies) for 24 hr or 48 hr.

3.3.1.2 RT-PCR validations of RNA-Seq data

In current literature, RNA-*Seq* data is typically validated by another global gene expression technique, RT-PCR analysis. Thus, it was necessary to attempt to replicate our RNA-*Seq* data using RT-PCR. This was particularly imperative in this study considering RNA-*Seq* data was obtained from a single biological replicate and

thus artifacts of the technique could not be detected by any other means. RT-PCR validations in this study were performed on a select number of De*Seq* processed genes with coefficient of variation (CV) < 0.05 which qualifying them as housekeeping genes. This analysis was performed by Joana Cruz. Of the 22 genes in this category, only 4 genes were selected for RT-PCR validation (**Table 2**).

Table 2: Housekeeping mRNAs in RNA-*Seq* data

Gene	CV	De*Seq* counts								
			Lm-WT (min poi)				Lm-*Δhly* (min poi)			
		Uninfected	20	60	120	240	20	60	120	240
Yars	0.03	437.5	454.1	454.7	457.5	470.8	449.7	489.8	448.0	467.4
Mcfd2	0.04	356.0	353.7	380.2	362.1	377.2	391.4	379.4	380.4	379.2
Ppap2c	0.04	135.8	127.0	137.2	145.9	134.1	138.9	131.1	134.3	137.5
Crtap	0.04	639.8	613.4	643.8	577.2	622.5	604.1	590.4	585.4	627.6
Hnrpa1l3	0.04	747.6	731.0	693.8	724.1	727.5	688.6	716.5	775.0	690.2

A total of 500 ng of the RNA extracted from *Lm*-WT infected cells (0, 20 and 60 min poi) per sample was converted into cDNA using the Superscript II kit (Life Technologies) according to manufacturers specifications using random-hexamer primers during the reverse transcription reaction. All RT-PCR reactions were carried out using 1 ng of template cDNA in triplicate using Ssofast Evagreen supramix (BioRad) in a BioRad CFX96 thermocycler. Fold expression changes were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak *et al.* 2001) and the nonparametric paired Student's t-test was used for statistical analysis were p values < 0.05 were considered statistically significant.

3.3.2 Cell culture

3.3.2.1 HeLa cells

Hela cells were cultured in Dulbecco's modified Eagle's medium (Gibco), 10% fetal bovine serum (FBS), 2-4 mM L-glutamine (Lonza) and passaged every 2-3 days.

3.3.2.2 IPSc-derived reprogramming into macrophages

IPSc's (indcued pluripotent stem cells) were grown on 1% gelatin (Corning) in $dddH_2O$ coated (1hr at 37°C) plates seeded with Mitomycin C-inactivated mouse embryo fibroblasts (iMEF's) (1.5X10'6 cell density per well) for 24 hr. iPSCs were grown on iMEFs in hESc media (Dulbecco's modified Eagle's medium (GIBCO), 1% nonessential amino acids (GIBCO), 0.055 mM β-mercaptoethanol (GIBCO), 2 mM Glutamax (GIBCO) and 10ng/ml FGF (fibroblast growth factor, R&D) and feed daily for a 7 days before passaging by mechanical dissection. Mechanicially disscected iPSCs were then transitioned to coculture-free conditions over several passages as follows:

- Passage 1 : 75% hESc media + 25% StemPro hESC medium (Life technologies) + geltrex (GIBCO)*
- Passage 2 : 50% hESc media + 50% StemPro hESC medium + geltrex
- Passage 3 : 25% hESc media + 75% StemPro hESC medium + geltrex
- Passage 4 & thereafter: 100% StemPro hESC medium + geltrex

*Geltrex was prepared by slowing thawing on ice at 4°C overnight. The following morning, 100ul geltrex was diluted into chilled 10ml DMEM-F12 medium. DMEM-F12-geltrex medium was poured into growth plates and coated at 37°C for 60 min. Coated plates were incubated at room temperature for 60 min prior to cell plating.

Transitioning and transitioned iPScs were passaged by enzymatic dissocation with 500 ul StemPro accuatase per well (6 well plate) for 5 min at 37°C. Cells were then lifted by pipetting and spun at 200g for 5 min at 22°C before being seeded on geltrex-coated plates.

For embryoid body (EB) formation, iPScs were seeded into 6-well ultra-low adherence plates (Corning) in StemPro medium containing ROCK-1. The following day for the next 3 days, EBs were grown in StemPro containing bone

morphogenetic protein 4 (BMP-4), vascular endothelial growth factor (VEGF) and stem cell factor (SCF).

For factory formation, four day old EBs were transferred onto adherent plates and grown in factory medium (FM) (X-Vivo (Lonza), glutamax (GIBCO), β-mercaptoethanol, 100ng/ml M-CSF (macrophage colony stimulating factor), 25ng/ml IL 3) media containing MCSF and IL3 and left to adhere at 37°C for 7 days. Adhered factories were then fed FM medium on a weekly basis.

Approximately 30 days after adhering factories, monocytes were harvested and CD14 selected on a weekly basis for approximately 6 months in the following manner: Monocytes were collected in FM medium and spun down at 400 g at 22°C for 5 min then resuspended in 100ul per 10 million cells collected of Recollection medium (2% FBS (GIBCO) and 1 mM EDTA (Sigma) in PBS) and transferred into a polysterene tube for CD14 collection. Collected monocytes were then CD14 selected with the EasySep human CD14 selection kit (STEMCELL Technologies) as per manufacteurers specifications.

CD14-selected monocytes were then seeded in 6-well plates ($1X10^6$ cells per well) or on cleaned coverslips in 24 well plates ($2X10^5$ cells per well) in Macrophage medium (RPMI1640 (GIBCO), 10% FBS (GIBCO), 2 mM glutamax (GIBCO), 100ug/ml M-CSF) for 7-10 days to allow for macrophage differentiation. CD14+ monocytes/macrophages were feed every 3 days following seeding until complete differentiation was reached. Mature differentiated CD14+ iPSc monocyte-derived macrophages (MDMs) were used for *Listeria* infections after 7-9 days of seeding.

3.3.2.3 Listeria infections

Listeria monocytogenes EDGe-GFP was grown in tryptic soy broth (TSB) containing 30mg/ml chloramphenicol (Sigma) to an OD_{600}=0.7-0.9 and 1 ml aliquots were frozen in 50% glucose (Sigma Aldrich, Germany). Frozen aliquots of bacteria were thawed at 37°C for 1-2 min then washed twice in 1X PBS, followed by a wash in serum-free DMEM containing 4 mM L-glutamine and resuspended in DMEM containing 10% FBS and 4 mM L-glutamine. All bacteria washes were done at 900g for 1 min. Washed and spun down bacteria were grown on TSB-chloramphenicol

plates overnight and colonies were counted to determine the CFU/ml surviving freeze-thaw cycle per vial.

For infection experiments, HeLa cells were seeded at a density of approximately $5X10^{X4}$ cells per well on cleaned coverslips in 24 well plates overnight to achieve a cell confluency of approximately $2X10^{X5}$ per well prior to infections. iPSc-derieved macrophage infections were carried out on $2X10^{X5}$ seeded CD14+ selected monocytes on cleaned coverslips. Cells were washed serum free media infected at a multiplicity of infection (moi) of 50 for all infections in complete media unless otherwise stated. Cells were fixed in 3.7% formaldehyde (Ambion) in PBS for 10min, washed twice in PBS at the following time points after infection 0, 20, 60, 120 and 240 min.

3.3.2.4.1 RNA Extractions

Six well plate cultures of infected HeLa cells or iPSc-MDMs were lysed in 1ml TRIzol (Sigma) per well/dish and passed cell ysate several times through a pitpette. Homogenized cells were incubated for 5 min at room temperature for nucleoplasm complex dissociation and centrifuged at 12 000 x g for 1 min and the supernant was transferred into a new tube. RNA was then extracted using the Directzol™ RNA MiniPrep kit with in-column DNAse I treatment (ZymoResearch) as per manufacterers specifications. RNA was eluted in 50 ul water per sample. The concerntrations and intergrity of eluted RNA was then measured on the Nanodrop ND-1000 spectrophotometer (Nanodrop technologies).

3.3.3 Staining

3.3.3.1 Cleaning of coverslips for microscopy

Coverslips were cleaned by washing in acetone, methanol and $dddH_2O$ repeating cylce 3 times, sonicated in 1 M KOH (Sigma) for 1hr, sonicated twice in $dddH_2O$ for 15min and stored in 40% ethanol at room temperature until use.

3.3.3.2. Phalloidin staining

Fixed cells were permeabilized using 0.1% Triton X (Sigma) in 1X PBS for 10 min at room temperature, washed twice in PBS and stained with 1 nM Phalloidin 565 or 647N (Attotec, Germany). Stained coverslips were washed twice in PBS. Cells were stained with 1 nM Hoescht (Life Technologies) in PBS and mounted on slides in Vectashield (Vector Labs) for imaging.

3.3.3.3. Indirect Immunofluorescence

Fixed cells were permeabilized in 0.1% Triton X in 1X PBS for 10min at room temperature and washed X2 in PBS. Cells were incubated in blocking buffer (1% BSA in 1X PBS) for 30min at room temperature. Cells were then incubated in 0.3 ug primary antibody in blocking buffer for 1 hr at room temperature. Cells were washed X5 in wash buffer (0.05% Tween-20 (Sigma) in 1X PBS) and incubated with 1ug secondary antibody solution in blocking buffer. Cells were again washed X5 in wash buffer. Cells were stained with 1 nM Hoescht in PBS and mounted on slides in Vectashield for imaging.

3.3.4. Imaging

3.3.4.1 Brightfield Imaging

Stem cells were imaged on an inverted microscope in bright field using a smartphone camera at **10X and 20X** magnification.

3.3.4.2. Widefield Fluorescence Imaging

All fluorescence imaging was performed using a Nikon Eclipse Ti inverted fluorescent microscope. Imaging was carried out at 60x numerical aperture (NA) and 100x, 1.49 NA oil immersion objectives. Images were captured using the Andor 897 iXion EMCCD camera (Andor, Belfast, UK). The microscope was controlled using µManager open source microscope management software (UCSF and NIH, USA). Exposure times for DAPI signal was kept at 30 milliseconds, and ranged between 100 milliseconds and 2.0 seconds for other signals

3.3.5. Image Processing

3.3.5.1. Brightfield Images

Brightfield images were processed using ImageJ open source image software (NIH, USA) adjusted for brightness and contrast and scale bars added.

3.3.5.2. Fluorescent Images

Image processing and manual inspection was carried out in ImageJ open source image software (NIH, USA). A single 2D image of z-section stacks was produced by maximum intensity projection, which projects the brightest pixel value of each position throughout the z-stack onto a single composite image. This was followed by background subtraction and contrast adjustment. Pseudo-colors were applied to the different channels, and images of the same field of view in the appropriate channels were overlaid to show relative localization of signals from each channel.

Chapter 4: Results

4.1 RNA-*Seq* of *Listeria*-infected HeLa cells

(RNA Samples extracted by Dr. Y. Shibayama, Dr. Suraj Parihar & Dr. Reto Guler and RNA-Sequencing performed by BGIAmericas Technologies)

Quantification of the *Listeria*-infected HeLa transcriptome by RNA-*Seq* required the extraction of RNA from *Listeria*-infected HeLa cells. In order to obtain RNA for the RNA-*Seq* experiment, we began by optimizing *Listeria*-EDGe infections in order to accurately achieve the above-mentioned hallmarks of *Listeria* infection in epithelial cells at the relevant time points. We opted to infect HeLa cells with a relatively high mulitiplicity of infection (moi) approximated at 50 bacterial units per HeLa cell. The time points; 20, 60, 120 and 240 min post infection (poi) used in this study were selected in order to observe each of the hallmarks of *Listeria* infection.

The number of colony forming units (CFU) per ml of *Listeria*-infected cells was determined by extraction intracellular bacteria which was plated and CFU were counted (**Table 3**). From these numbers, we were able to calculate the percentage of bacterial uptake in infected cells as well as the rate of bacterial replication at set time points. These results show a similar percentage of bacterial uptake in both the EDGe (*Lm*-WT) and ΔLLO (*Lm-MUT*) *Listeria* strains at 20 min and a slightly higher percentage of bacterial uptake at 60 min poi and rate of bacterial replication of the *Lm*-MUT strain (**Table 3**).

Table 3: CFU counts

***Listeria* strain**	**Time point (min poi)**	**Average CFU/ml**	**Bacterial Uptake (%)**	**Rate of replication (min per doubling event)**
EDGe	20	$2.9x10^5$	23	
	60	$7.8x10^5$	62.4	
	120	$2.9x10^6$	n/a	36
	240	$2.6x10^7$	n/a	
Δ*LLO*	20	$3.1x10^5$	24.8	
	60	$1x10^6$	80	
	120	$1.5x10^6$	n/a	45
	240	$1.9x10^7$	n/a	

Table 4: RNA concentration and integrity

Listeria strain	Time point (min poi)	Total RNA available	260/280	260/230
Uninfected	0	133.9	2.08	2.09
EDGe	20	72.5	2.10	2.13
	60	153.7	2.06	2.14
	120	87.9	2.08	2.15
	240	59.7	2.07	2.09
Δ*LLO*	20	59.7	2.06	2.05
	60	69.0	2.09	2.09
	120	169.8	2.05	2.10
	240	35.2	2.01	1.98

Next, we infected HeLa cells with *Listeria* at an moi=50 and extracted RNA at the relevant time points, extracted RNA was then sent to BGIAmericas for RNA-*Sequencing*. Prior to sequencing, the concerntrations (ng/ul) and integrity (260/280 and 260/230 ratios) of extracted RNA samples were measured to ensure high sample purity (**Table 4**). In addition, RNA integrity was validated on a 12% urea gel (**Figure 4.1a**).

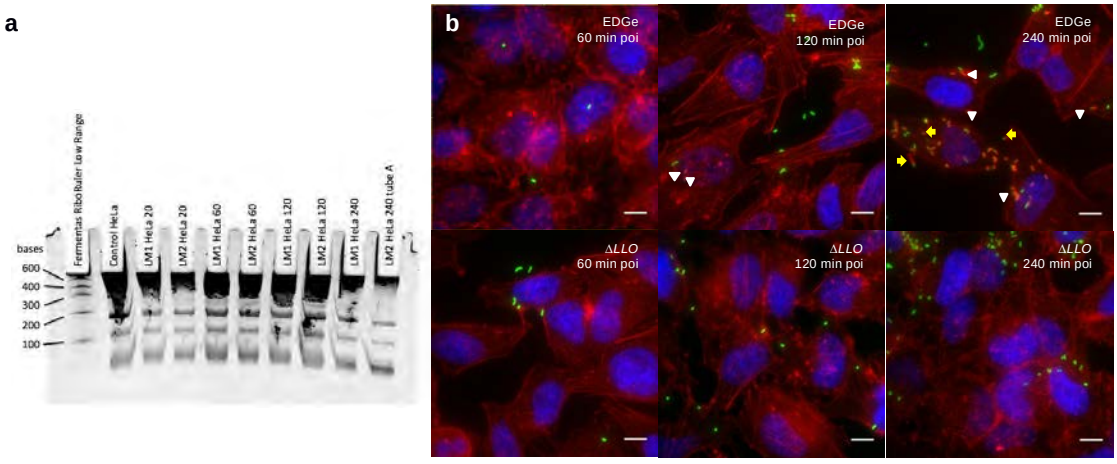


Figure 4.1**: RNA-*Seq* of *Listeria*-infected HeLa cells. a. RNA intergrity** 8 ul of each *Listeria*-infected HeLa extracted RNA sample was loaded onto 12 % urea gel. LM1= *Listeria* EDGe, LM2= *Listeria* Δ*LLO.* Numbers represent min poi (i.e. 20=20 min). **b. Microscopy analysis of infection time course.** The top panel shows wildtype EDGe infection and the bottom panel shows Δ*LLO* infection. White arrowheads=actin recruitment, Yellow arrows= comet tails, Blue=nuclei, Red= actin, Green= anti-*Listeria*. Scale bars = 5 um.

Simultaneously, HeLa cells grown on coverslips were also infected, fixed at the relevant time points then imaged using widefield microscopy for visual analysis and

validation of the infection time course (**Figure 4.1b**). The infection time course images revealed hallmark features of the *Listeria* infection cycle: bacteria internalization (60min poi), bacterial replication and actin recruitment (120min poi) and comet tail formation (240 min poi) (**Figure 4.1b**). Satisfied with the infection assay and the quality of RNA extracted from it, the extracted was sent for sequencing at BGIAmericas technologies.

4.2 *Listeria* infection optimizations

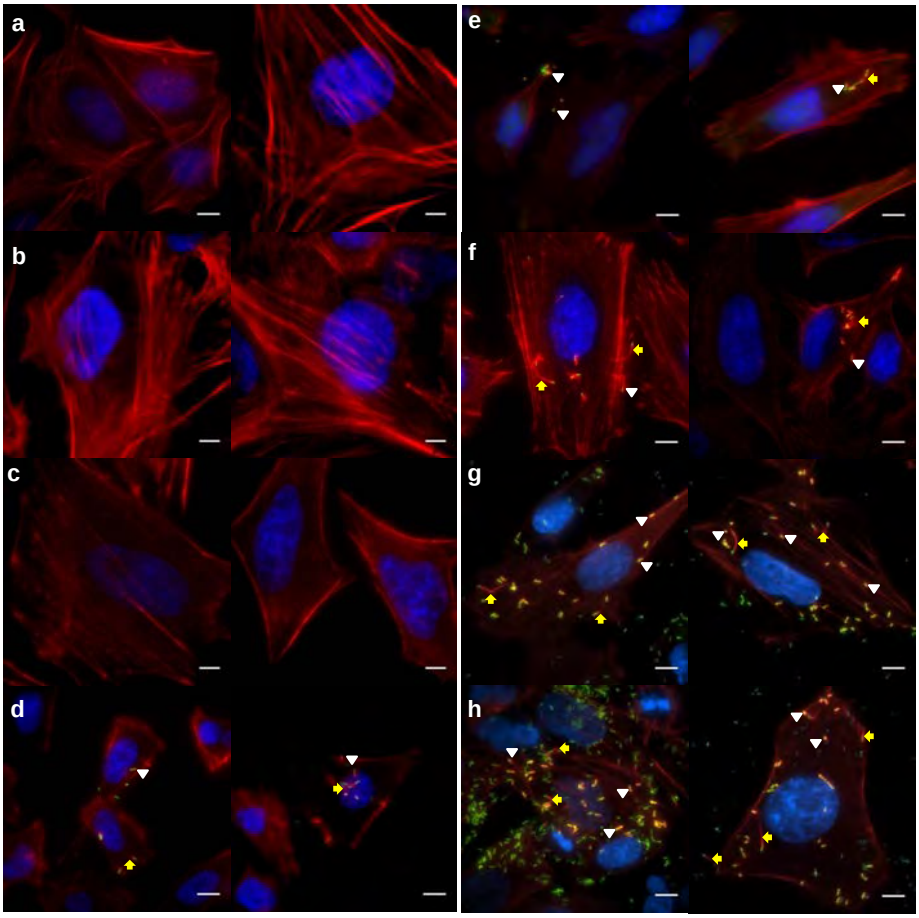


Figure 4.2**: *Listeria* infection optimatizations at moi=50.** Infections were performed with 0.25 ml infection media in 24 well plates at an incubator door temperature of 70% at room temperatures of **a.** 23°C **b.** 24°C **c.** 25°C. The rest of the optimizations were perforemed at room temperature of 24°C in **d.** DMEM F12 media, or in DMEM media with **e.**15 min bacterial settling at room temperature prior to infection time course. Infections were carried out with HeLa passage numbers less than 20 i.e. **f.** P6. Low passage HeLa cells were infected with infection media (DMEM) of **g.** 500ul or **h.** 1000ul per well. Nuclei=blue, Red=actin, Green= *Listeria* EDGe-GFP, White arrowheads= actin recruitment, Yellow arrows=comet tails. Scale bars=5 um.

For RNA-*Seq* data validation and lncRNA discovery experiments, *Listeria* infection time courses were optimized for infection with the wild type EDGe-GFP strain at the 4 hr poi time point using bacterial replication and actin recruitment/comet tail as indicators for successful infection. Several alterations to the initial infection protocol followed by Dr. Shibayama and Dr. Guler were made in order to achieve successful infection with the relevant hallmarks of the *Listeria* infection cycle in HeLa cells (**Figure 4.2**). A major difference in the infection protocol as compared to the initial

protocol was the use of the fluorescent *Listeria* EDGe-GFP strain instead the non-fluorescent strain used by Dr. Shibayama, Dr. Parihar and Dr. Guler. Optimal infection conditions yielded the relevant hallmarks of *Listeria* infection: actin recruitment and comet tail formation by 240 min poi. These conditons for moi=50 *Listeria* EDGe-GFP infections were obtained at an incubator door temperature of 70%, room temperature 24°C and infection media volume of 0.5 ml in 24 well plates (**Figure 4.2**). Subsequent infection time courses were carried out under these conditions.

4.3 Bioinformatics: LncRNA discovery

Having optimized the infection conditions to be used for RNA-*Seq* data validation and lncRNA discovery experiments, we then turned to the RNA-*Seq* data to investigate *Listeria*-induced alterations in the HeLa transcriptome. To this end, we used focussed on two strategies which involved the investigation of the mRNA and lncRNA data respectively for both the wild type and *Δhly* mutant infection data.

4.3.1 Strategy 1: Differential expression of mRNA during *Listeria* infection

4.3.1.1 mRNA heat map

In an effort to determine significantly differentially expressed genes in *Listeria*-infected cells the mRNA data was filtered according to the following parameters: *M-values* ≥0.5 and ≤-0.5 and *A-value* ≥7.8 for each infection condition. This filtration resulted in 133 differentially expressed genes, which were complied into a heat map (**Figure 4.3**), among which, 9 were unique to the wild-type infection and 16 unique to the *Δhly* infection (**Figure 4.3**, **Table 5**).

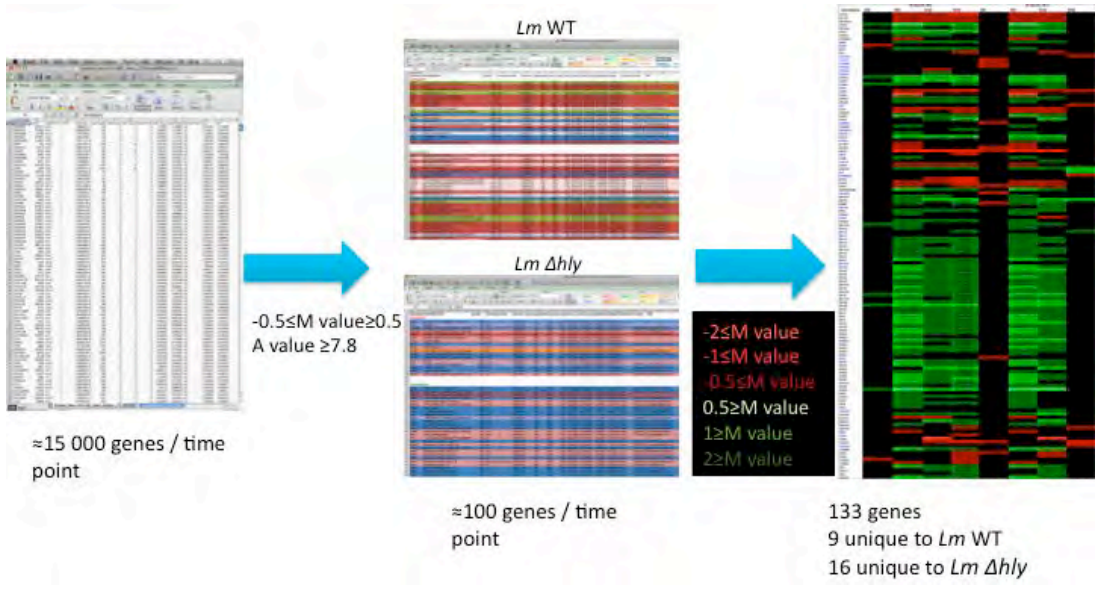


Figure 4.3***:*** **mRNA data filtration:** the poly-A tailed component of the RNA-*seq* data was filtered by (*M-values* ≥0.5 and ≤-0.5 and *A-value* ≥7.8) for each time point and infection condition resulting in a heatmap of 133 significantly downregulated and upregulated genes in *Listeria*-infected HeLa cells.

4.3.1.2 LLO -dependent and –independently regulated genes

Table 5: LLO- independent (unique to *Lm-Δhly*) and -dependent (unique to Lm-WT) significantly differen-tially expressed genes in *Listeria*-infected HeLa cells.

		M values LM1				M values LM 2			
	GeneName	T20	T60	T120	T240	T20	T60	T120	T240
Lm WT	BASP1	0,506239668747514							
	COL3A1			0,509506626765278					
	COX6A1			-0,554157236999401					
	CRIP1			-0,553214547183109				-0,553214547183109	
	HSP90AA1				-0,616547244473720				
	KRT8				0,549839609927226				
	PTMS							0,526905597394052	
	RPL27A			-0,548692147990016	-0,624392956289546				
	TGFBI				0,709573761007668				
Lm Δhly	C15orf63								0,592216675301889
	CCDC72					0,723950213738299			
	CDKN2A					0,959699606879945			
	CHCHD2					0,529598548275670			
	EIF1								0,659933736456636
	HMGN2					0,528748769789735			
	IFITM3						-0,521069362403325	-0,626715553635347	
	LGALS1							-0,847956245355550	
	MIF								-1,526662491553000
	MTRNR2L1								0,606788938874459
	NDUFA4					0,523330449618278			
	NOP10							-0,570414575424691	
	RPL14						-0,505367747771095		
	RPS2					0,594609438276647			
	S100A11							-0,563964426013053	
	TMSB10								0,507544053116745

Next, we explored the differentially expressed genes unique *LLO*-secreting (*Lm*-WT) and *LLO*-deleted (*Lm*-MUT) infection condtions. *LLO*-dependent differentially expressed genes (*BASP1, COL3A1, Cox6A1, CRIP1, HSP90AA1, KRT8, PTMS, RPL27A, TGFBI*) (**Table 5**) were involved in a variety of cellular functions including transcriptional regulation, extracellular matrix compostion, intracellular transport, mitochondrial functionality and translation.

To identify potential lncRNAs regulating the transcription of these LLO-dependent genes, the 80kb genomic regions surrounding these genes was scanned both in the UCSC genome browser (https://genome.ucsc.edu) and the RNASeq ncRNA data. Several potential lncRNAs regulating these genes were identified from the data. However, these ncRNAs were not investigated further due to one or more of the following reasons: uncorrelated expression patterns to LLO-dependent *cis* localized gene within 100 kb (-0.5>M values<0.5), lack of putative lncRNA-characteristic K4-K36 domain, support vector machine transcripts and/or >100 bp in length.

4.3.1.3 Ribosomal proteins: largest class of proteins in the mRNA heatmap

Unexpectedly, 40% of the significantly differentially expressed mRNAs in the data are transcribed from 52 of the 75 ribosomal protein (RP) genes (**Figure 3.4**).

Furthermore, all but one (*RPS2*) of these RP mRNAs are significantly downregulated during *Listeria* infection independent of *LLO*-secretion. Evidently, the downregulation of RP mRNAs generally occurs after 60 min poi, which coincides with the phagosomal escape infection stage of *Listeria* (**Figure 3.4**). This observation suggests *Listeria* infection, specifically bacterial entry into the cytoplasm, potentially leads to a global suprression of host translation by a widespread downregulation of RP gene expression.

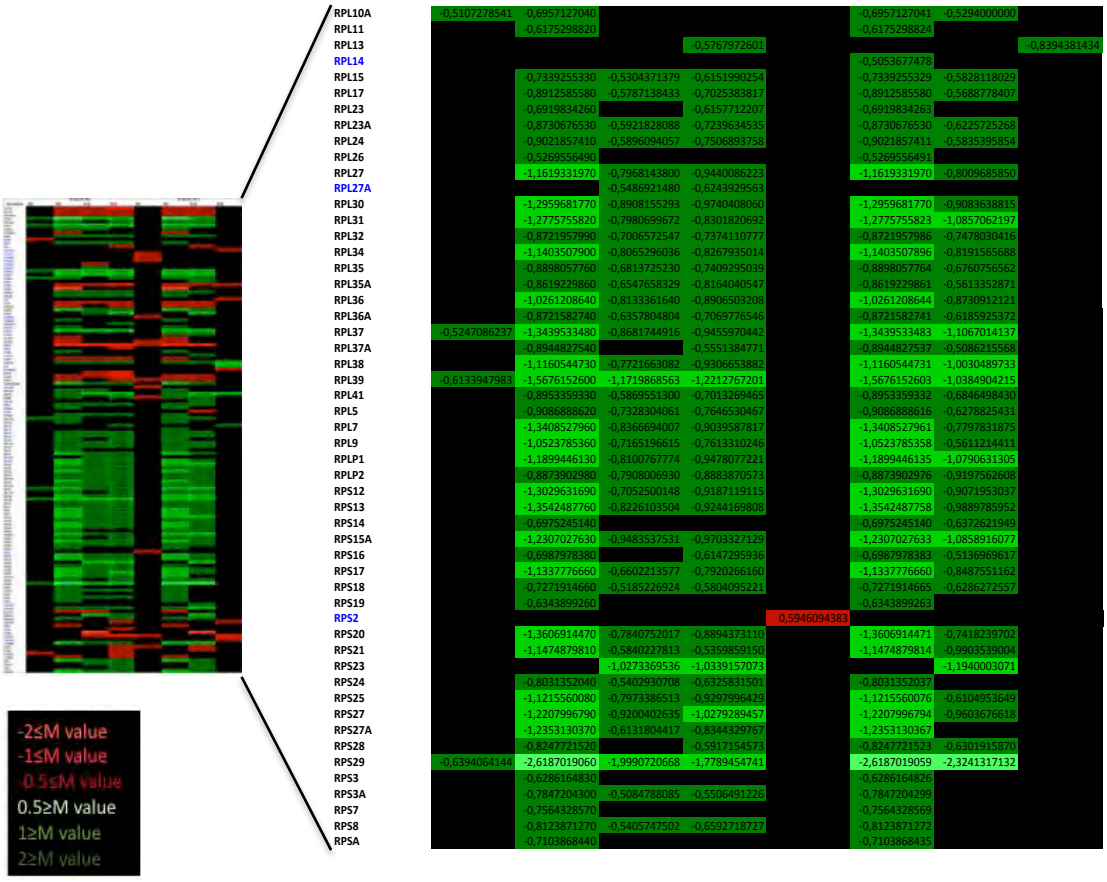


Figure 4.4**: Ribosomal proteins are downregulated during *Listeria* infection.** Of the 133 differentially expressed genes in *Listeria*-infected cells, 40% of the genes represent 52 of the 75 ribosomal proteins.

Extra-ribosomal translation-independent functions of RPs include transcriptional regulation (S3, L11) (Zang *et al.*, 2003, Wan *et al*., 2007) cell cycle control (L7, L13a, L11, S5) (Chen and Ioannou, 1999, Zang *et al.* 2003, Neumann *et al.* 2007, Matragkou *et al*., 2008) apoptosis (S3, S26) (Zang *et al*., 2003, Jang *et al.* 2004, Deisenroth and Zhang, 2010, Cui *et al.* 2014), splicing (S13) (Malygin *et al.* 2007), cytokine regulation (L13A) (Poddar *et al.* 2013), differentiation and development (L38, S5, S7, L22, L29) (Anderson *et al.* 2007, Matragkou *et al.*, 2008, Oristain *et al.*

2009, Kondrashov *et al.*, 2011, Duan *et al.* 2011), DNA repair (S3) (Jang *et al.* 2004). This assortment of extra-ribosomal functions, uncovered to date, may be specifically targetted by *Listeria* infection.

Notably, the only RP that is upregulated during *Listeria* infection, RPS2, a component of the 40S ribosome subunit, does not have known extraribosomal functions. To investigate the involvement of lncRNAs in the downregulation of RP transcription during infection, RNA-*Seq* data of these RPs were explored as in the previous strategy.

4.3.1.4 Differentially expressed genes during Listeria infection

Table 6: Top five significantly differentially expressed genes during *Listeria* infection

	M values LM1				M values LM 2			
GeneName	T20	T60	T120	T240	T20	T60	T120	T240
ATP5E	-0,6170092903	-2,0189401080	-1,4563212427	-1,5886739170		-2,0189401077	-1,6643898496	
COX7B	-0,6122993138	-1,9253035500	-1,3439199268	-1,1684161159		-1,9253035503	-1,5473122854	
RPS29	-0,6394064144	-2,6187019060	-1,9990720668	-1,7789454741		-2,6187019059	-2,3241317132	
THBS1			3,2645857920				2,8603353313	
CYR61		4,9013452880				4,9013452882		

An important observation from the mRNA heatmap is that two of the most signifcantly expressed genes (*M-values* ≥1 and ≤1) are involved in extra-cellular matrix interactions and signalling (*CYR61, THBS1*). Other genes in this filtered class are involved in mitochrondrial functioning (*ATP5E, Cox7b*), translation and apoptosis (*RPS29*) (**Table 6**).

Proximal lncRNA analysis at the genomic loci of these genes in the RNA-*Seq* data revealed an 187 bp ncRNA downstream of the ATP5E gene and antisense to the TUBB1 gene. Notably, the TUBB1 gene was found to be upregulated at 60 min poi in the *Lm*-WT data. However, the ncRNA had no characteristic chromatin marks associated to its genomic loci in HeLa S3 cells, and was not found in any of the lncRNA databases searched. The Cox7b locus yielded abundant and correlatively expressed ncRNA isoforms of the Cox7b genes with no distinct lncRNA chromatin marks. Examination of both our RNA-*Seq* data and various databases for lncRNAs within the 100kb THBS1 region in HeLa cells failed to produce a proximal lncRNA potentially regulating THBS1 expression. However, the 100 kb CYR61 locus revealed a conserved and abundant ncRNA locus, which was not found in our RNA-*Seq* data and thus was not futher analyzed.

4.3.1.5. A potential lncRNA regulating Filamin A expression

Although all 133 genomic loci of the filtered data where investigated for proximal lncRNAs and literature reviews of each of the genes were conducted, only a select few showed potentially infection-regulating functions, particulary with regards to proximal lncRNA characteristics and infection-related functions of the genes investigated. One such gene is actin binding protein Filamin A (FLNA), which was upregulated almost 2–fold in both infection conditions from 60 min poi infection (**Figure 4.4, 4.5**).

FLNA is a well-characterized cytoskeletal actin-binding protein regulating cell shape and cellular migration by crosslinking actin filaments (Stossel *et al.* 2001, Zhou *et al.* 2010, Nakamura *et al.* 2011). FLNA is involved in the stablization and reorganization of cortical F-actin at the lamellar membrane of the leading edge of motile cells. FLNA also functions as a scaffold for intergrins, small GTPAses and their upstream and downstream factors (Zhou *et al.* 2010). It has also been shown to be located within the nucleus were it interacts with transcription factors and binds BRCA2, forming a complex that participates in the DNA damage response (Yue *et al.* 2009) Recently, nucleolar located FLNA was shown to suppress ribosomal RNA expression by inhibiting pol I machinery to the rDNA promoter (Deng *et al*., 2012). In addition, FLNA can faciliate Arp2/3-independent *Listeria* comet tail elongation (Brieher *et al.* 2004). The involvement of FLNA in actin dynamics made it particularly interesting to further investigate during *Listeria* infection.

An RNA-*Seq* ncRNA data search yielded two ncRNAs within the 100 kb of the FLNA locus. The 196/7 bp ncRNA was selected for further investigation, as the expression. Intriguingly, this FLNA-proximal lncRNA candidate is located within the 3'UTR of emerging, EMD (**Figure 4.5 a, c**). Strikingly, the EMD-lncRNA's parental gene, EMD, showed a correlative differential expression pattern to that of the ncRNA in the RNA-*Seq* data. In fact, EMD expression decreases up to 2-fold from 20 min poi; inverse to the 2-fold decrease of the lncRNA (**Figure 4.5b**). EMD is a LEM-domain protein and integral component of the inner nuclear membrane that functions by tethering chromatin to the nuclear membrane by its LEM-domain mediated interaction with chromatin-protein "barrier-to-autointegration" factor (BAF) and lamin (Holaska *et al.*

2002, Margalit *et al*., 2007). EMD mutations resulting in EMD deficiency cause an X-linked form of the Emery-Driefuss muscular dystrophy (Emery and Dreifuss, 1966).

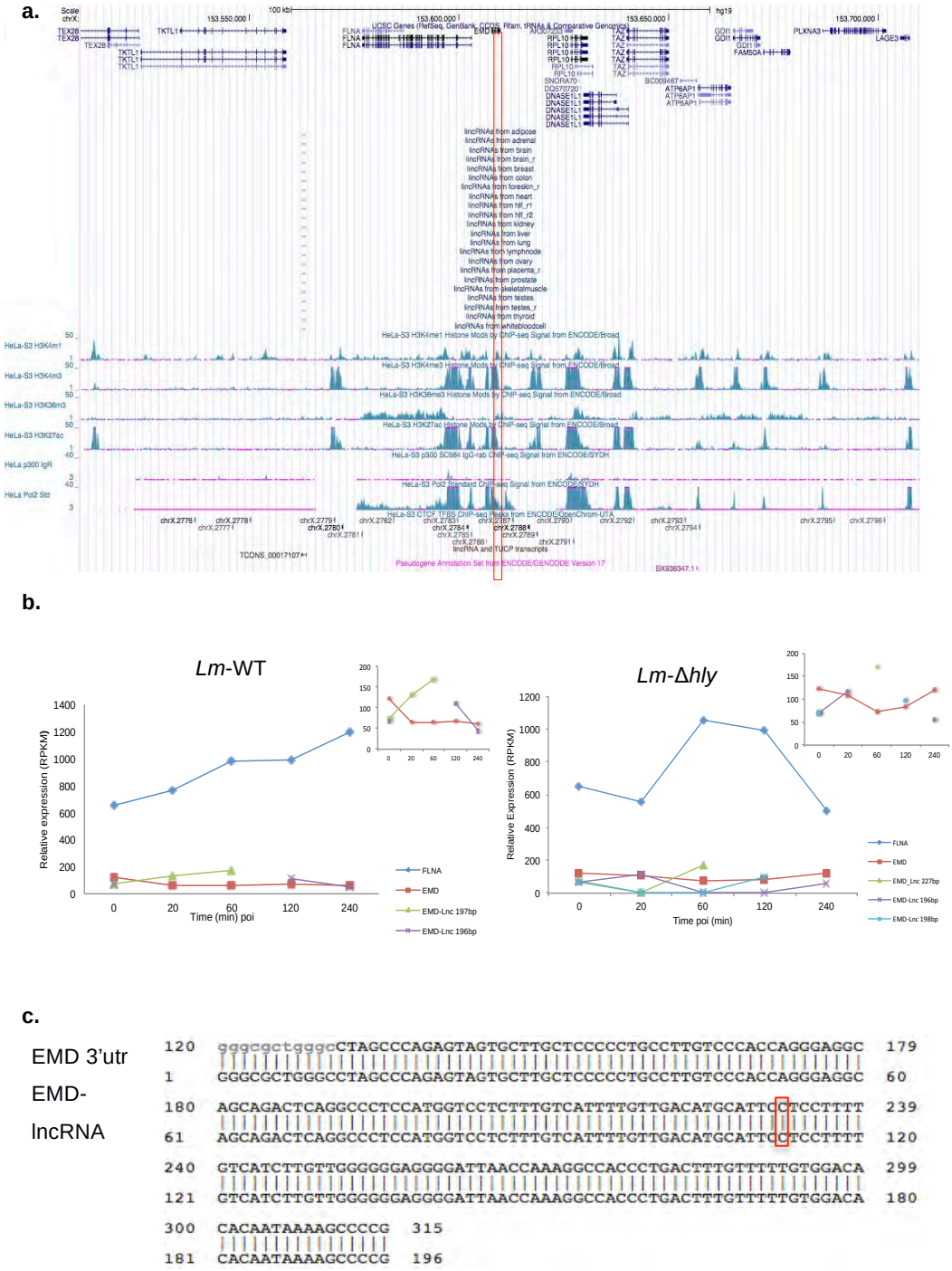


Figure 4.5: **3'UTR-derived ncRNA displays correlative expression to FLNA and EMD. a.** 100 kb region surrounding FLNA, red box indicates EMD-lncRNA location. Taken from (https://genome.ucsc.edu). **b.** Graphs displaying relative expression values of ncRNA isoforms; 197bp (green), 196bp (purple), 223(blue) and genes FLNA (dark blue) and EMD (red) in both infection conditions *Lm*-WT and *Lm*-MUT throughout infection time course. **c.** EMD-lncRNA aligns to a portion of EMD 3'UTR. Targeted TALEN cut site enclosed with red box.

The inversely correlated differential pattern of EMD-lncRNA and EMD suggests this intragenic ncRNA may be transcribed from an EMD-dependent promoter. Several lines of evidence have demonstrated transcriptional start sites (TSS's) within 3'UTRs of protein coding transcripts using CAGE, RACE and reporter gene methods (Carninci *et al.,* 2005, Mercer *et al.*, 2010). These 3'UTR-derived transcripts are independently transcribed from the parent transcript and are highly tissue specific (Mercer *et al.,* 2011). However, an independent TSS at the EMD-lncRNA locus could not be identified in the FANTOM5 (The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014) data.

Investigations on the chromatin signatures surrounding the EMD-lncRNA locus in HeLa S3 cells using the UCSC genome browser revealed a number of chromatin modifications (**Figure 4.5a**). These signatures infer that EMD-lncRNA is not a typical lncRNA with the K4-K36 consensus signature, which is consistent with active transcription, despite the presence of Pol2 occupancy at this locus. Furthermore, EMD-lncRNA does not display the eRNA consensus chromatin signature (high K4me1:K4me3 ratio) and therefore is not an enhancer eRNA, despite the enhancer associated mark H3K27Ac occupancy at this locus. The chromatin modifications at the EMD-lncRNA locus suggest that it is not transcribed and does not correspond to lncRNA or eRNA transcription in resting HeLa cells. The Pol II occupancy at this locus suggests transcription can occur at this locus, however, this may occur in different cellular contexts. Collectively, although conflicting, these observations warranted further investigation into this ncRNA particularly during *Listeria* infection.

4.3.2.1.1 EMD-lncRNA knockdown in HeLa cells

In an attempt to unravel the functional relevance of EMD-lncRNA during *Listeria* infection, we sought to knockdown EMD-lncRNAexpression using TALENs targeted to EMD-lncRNAgenomic locus. TALENs targeted to the 3'UTR of EMD were designed and produced according to the Zhang protocol (Sanjana *et al.* 2012). Functional characterization of TALEN pairs involves the transfection of TALENs in to cells. TALEN efficiacy and specificity can then be validated using a number of assays including H2AX staining, Surveyor assays, RNA FISH of targeted RNA, immnofluorescence or Western blots of targeted protien.

Transfected TALEN pairs targeted to a specific genomic loci discretly induce a double stranded break (DSB) with high single base pair specificity in their host's DNA. In response to DSBs histone variant H2AX is rapidly phosphorylated, resulting in the production of a chromatin domain of up to 1 Mb surrounding DNA lesions, which acts as a signal for the recruitment of DNA repair complexes (Rogakou *et al.*, 1999). Thus, to validate the efficacy of TALEN pair in producing DSBs, an immunofluorescence essay probing for H2Ax in the nucleus can be used. DSBs in mammalian DNA are partially repaired by the non-homologous end-joining pathway (NHEJ), which typically results in functional gene knockout caused by small sequence deletions (Huertas *et al*., 2010). Detection of this endogenous target cleavage by a TALEN pair can be carried out using Surveyor mutation detection assays, which carry the ability to quantify NHEJ. In this study, TALEN pairs have been successfully produced, transfected into HeLa cells using Lipofectamine RNAi-MAX (**Figure 4.6**) and - 2000 (**Figure 4.7**) and functionally characterized using the H2AX immunofluorescence assay (**Figure 4.6**).

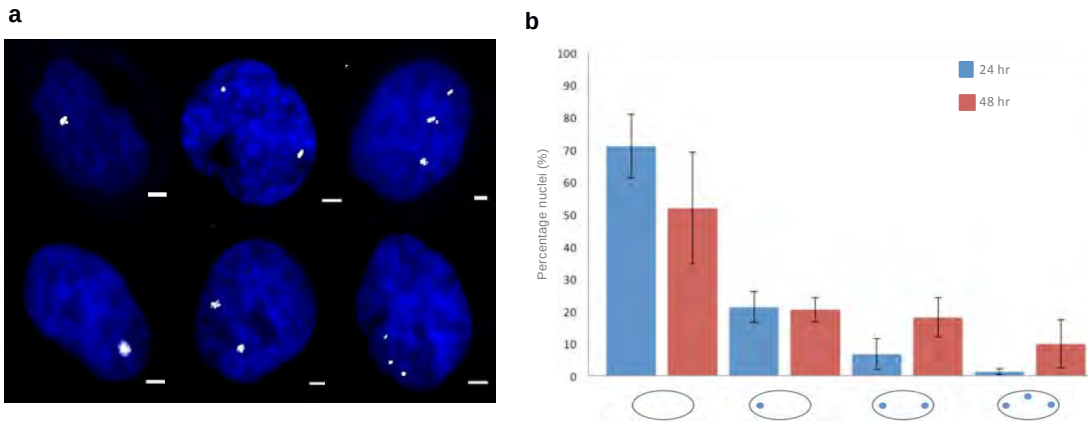


Figure 4.6**: EMD TALENs induce double-stranded breaks in HeLa cells at EMD 3'UTR locus. a.** HeLa nuclei fixed 24hr (top panel) and 48hr (bottom panel) post transfection and stained with H2AX indirect immunofluorescence. Scale bars=2 um **b.** Percentage HeLa nuclei with no, mono-, bi- and tri-allelic cuts after 24 and 48 hr EMD-TALEN transfections respectively (means ± SD of triplicate experiments).

Notably, this particular EMD-lncRNA is located within the 3'UTR of EMD, a protein-coding gene (**Figure 4.5b**), thus the TALEN pair transfection is likely to inhibit the expression of EMD, as well as the EMD-lncRNA. The inadvertent effect of inhibiting EMD expression is inconsequential to cell viability and morphology in HeLa cells (Harborth *et al*., 2001). TALEN-mediated EMD knockdown in this study showed no obivious effects on HeLa cell viability and morphology. Furthermore, HeLa cells have 3 copies of an X chromosome arm (Macville *et al.,* 1999, Livak *et al.* 2013),

although it is unclear whether this duplication encompasses the EMD-containing Xq28 arm. Nonetheless, in addtion to mono- and bi-allelic cuts, a significant portion of tri-allelic cuts were observed in EMD-TALEN transfected cells; 1.22 and 9.81% for 24 hr and 48 hr transfections respecitively (**Figure 4.6a**). The largest percentage of cells with cuts was observed 48 hr post transfection, except in the case of the mono-allelic cuts, determining this transfection condition as the standard for subsequent experiments (**Figure 4.6b**).

4.3.1.5.2. EMD-lncRNA in infection

To determine the effect of the EMD 3'UTR knockdown on *Listeria* infection, HeLa cells that were transfected with EMD-TALENS for 48 hrs, were then infected with the *Lm*-WT strain for 4 hrs (**Figure 3.7b**). A drastic decrease in the number of cells infected was seen in TALEN cut cells as compared to their uncut couter parts in the same well; 87,2, 12,8, 1.1, 0 % for no, one, two and three cuts per nucleus (**Figure 4.7a**). This data suggests that simultaneous EMD and EMD-lncRNA knockdown effectively inhibits *Listeria* infection.

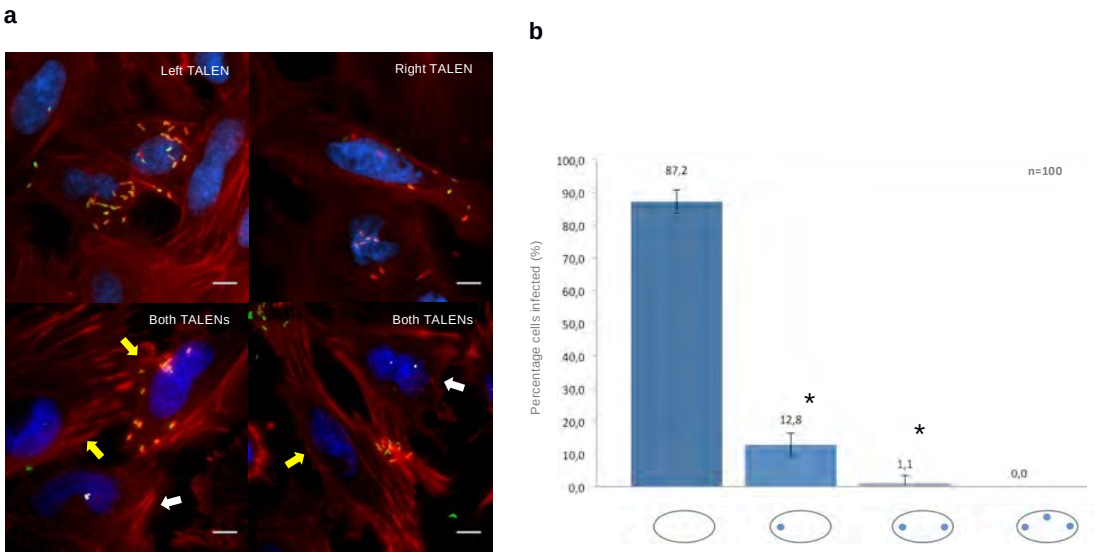


Figure 3.7**: Effect of EMD-3'UTR targetted TALENS in HeLa cells during *Listeria* infection. a.** Representative images of HeLa cells transfected with EMD-TALENs for 48 hr followed by 4 hr *Listeria* infection. Top panel: Left only and right only transfected control cells. Bottom panel: Cells transfected with both TALENs. Blue=nuclei, green=*Listeria*, red=actin, white=H2AX foci, white arrows=TALEN cut cells, yellow arrows=uncut cells. Scale bars= 5 um. **b.** Percentage of cells with no, mono-, bi- or tri-allelic cuts after 48 hr EMD transfections followed by *Listeria*-WT infection 4 hr poi (means ± SD of triplicate experiments).

Notably, this data is confounded by the simultaneous knockdown of EMD in addition to EMD-lncRNA knockdown. However, according to the RNA-*Seq* data, the EMD gene transcription is downregulated throughout the standard infection time course investigated here. Although the RNA levles from RNA-*Seq* have not been validated using microscopy-based analysis, the less than 200 RPKM reads of EMD throughout successful infection suggests that transcriptonal EMD knockdown does not inhibit infection (**Figure 4.1b**). Collectively, this suggests the infection inhibition observed in EMD-3'UTR knockdown cells may be solely mediated by EMD-lncRNA knockdown. However, this still needs to be confirmed with non-EMD-lncRNA targetted knockdown of EMD during *Listeria* infection.

3.3.2 Strategy 2: Differential expression of lncRNAs during *Listeria* infection

A major aim of this study is to identify potential lncRNAs from RNA-Seq data of *Listeria*-infected HeLa cells. To this end, we employed a number of filteration strategies on the RNA-*Seq* in order to hone in on potential *Listeria* infection-regulating lncRNAs. We began by creating a visual representations of the lncRNA data subset by plotting lncRNAs according to fold change (M value) and relative expression levels during the infection time course yielding MA (M value and A value) plots (**Figure 3.9**) and M (M value) plots (**Figure 3.8**). The MA plots were plotted per time point for the *Lm*-WT infection condition (**Figure 3.9**). The M plot shows the entire lncRNA dataset in both infection conditions across all time points, excluding all support vector machine (svm)-predicted transcripts (**Figure 3.8**).

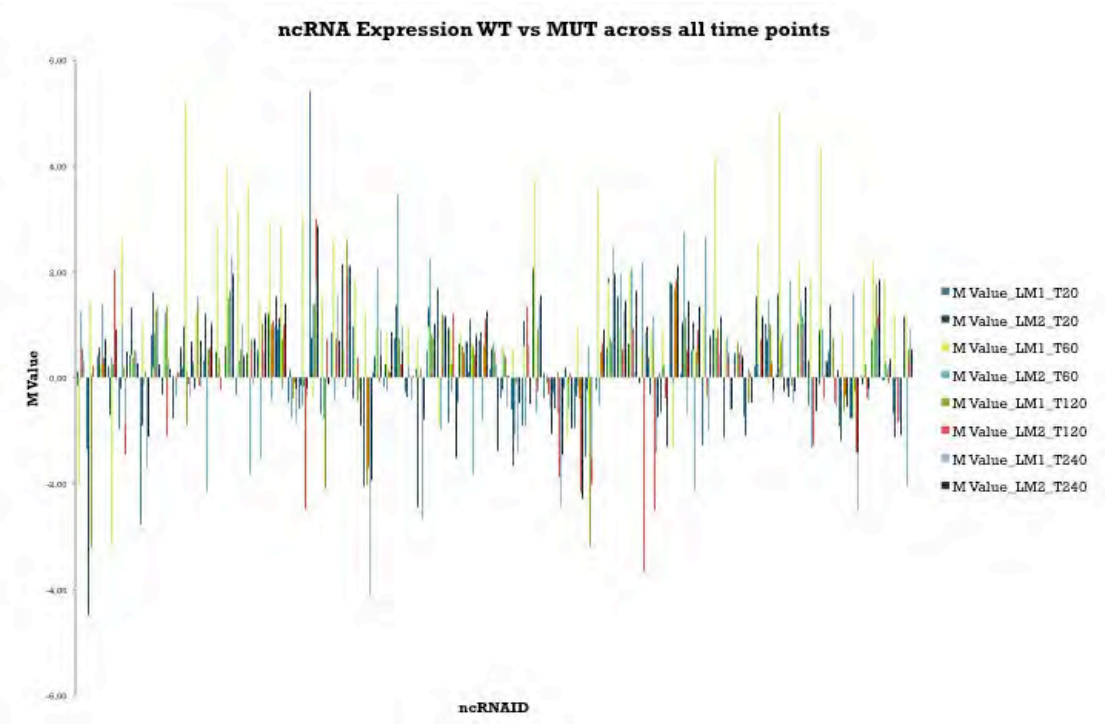


Figure 4.8**: ncRNAs in RNA-*Seq* data.** Graph *showing ncRNAs (x axis) and respective M-values(y axis) across all time points and infection conditions*

M value plots of the lncRNA dataset revealed the *Lm*-WT 60 min poi infection condition as the condition with the largest proportion of up-regulated ncRNAs and the largest fraction of down-regulated ncRNAs were found in the *Lm*-WT 240 mi poi condition (**Figure 4.8, Figure 4.9**). The highest number of ncRNAs within the *Lm*-WT lncRNA dataset is 8006 ncRNAs at 60 min poi as compared to 5218, 5274 and

5594 ncRNAs for 20, 120 and 240 min poi respectively (**Figure 4.9**). Furthermore, the *Lm*-WT 60 min poi infection condition also had the largest area of the graph with *M-values* ≥0.5 and ≤-0.5. Thus, subsequent analysis of the Lm-WT data was performed using this infection condition: *Lm*-WT 60 min poi.

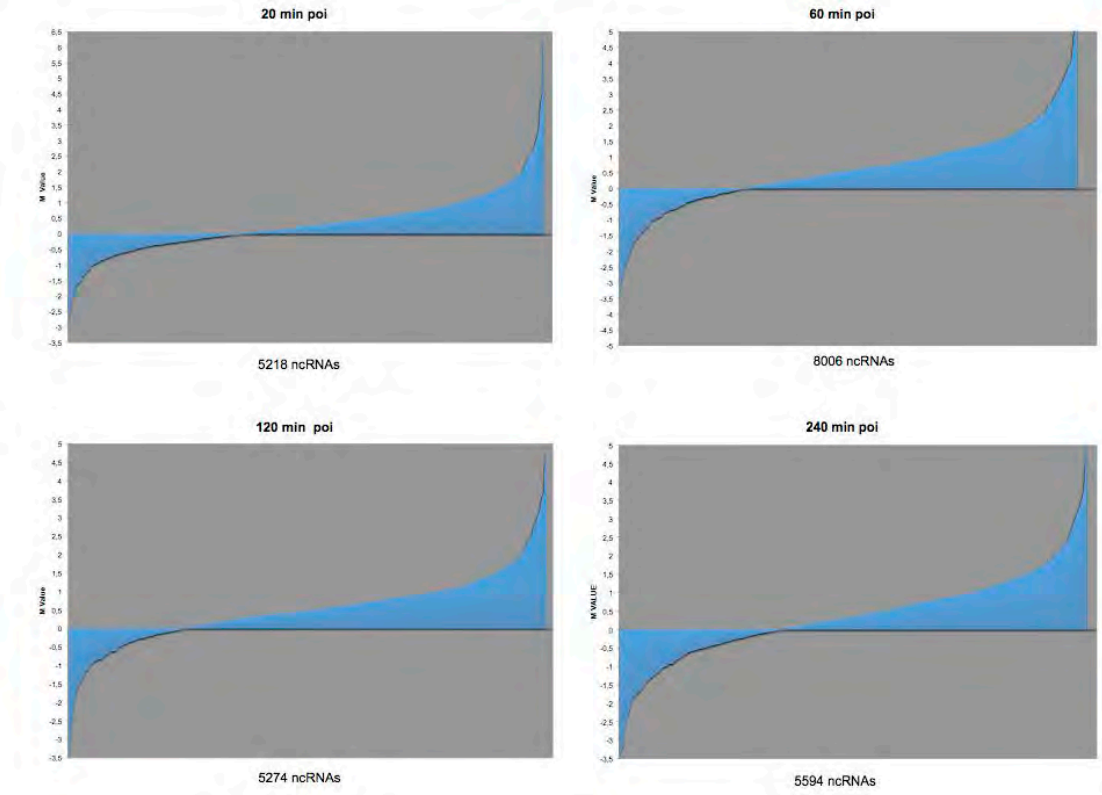


Figure 4.9**: MA plots of ncRNA data throughout *Lm*-WT infection.** Fold change (M value, y axis) and relative expression change (A value, x axis) plots showing the distributions of ncRNAs in the *Lm*-WT condition at different time points.

The distribution of the *Lm*-WT mRNA dataset shows that the 60 min time point, which is associated with the cytoplasmic entry of *Listeria* during infection, contains the largest changes in active host transcription (**Figure 4.4**). The non-coding portion of the data also shows sizeable increase in transcription at this particular time point compared to other time points (**Figure 4.9**).

Focusing on the ncRNA data subset in 60 min poi infection condition, we further filtered this data subset using the following parameters; *M-value* ≥2 and ≤-2, >100bp length and >100 RKPMs. This filtration process resulted in 5 and 26 ncRNAs, which were down- and up-regulated respectively. As with previous strategies, the 100 kb genomic loci surrounding the ncRNAs were investigated on the UCSC browser to recover chromatin signatures and proximal genes.

4.3.2.1 An intergenic lncRNA regulating global transcription factor ATF3?

Of the 26 upregulated lncRNAs determined by the pipeline described in Strategy 2, a 421 bp intergenic ncRNA was selected for further investigation. This ncRNA is located ~170 kb from and on the opposite strand of N-myc downstream regulated 1 (NDGR1). Thus, it will subsequently be referred to as linc-NDGR1+. This intergenic ncRNA was upregulated at 60 min poi with an M value of 4.33 (**Figure 4.10c**). The 421 bp linc-NDGR1+ is the consensus sequence for the following isoforms throughout the infection: 421 bp at 20 and 60 min poi, 511 bp at 120 min poi and 1144 bp 240 min poi in the *Lm*-WT data as well as a 494 bp at 60 min poi in *Lm*-MUT data. The differential expression pattern of the isoforms of this ncRNA peak at 60 min poi in the *Lm*-WT condition and decreases during the later time points (**Figure 4.10c**). In contrast to the *Lm*-WT, the expression pattern in the *Lm*-MUT peaks at an earlier time point 20 min poi before decreasing at the 120 min poi, then increasing slightly at the 240 min poi (**Figure 4.10c**).

Linc-NDGR1+ also shares a consensus sequence with the 1102 bp processed suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein) pseudogene 6 (ST13P6) (**Figure 4.10**) and appears to be an isoform of ST13P6. Notably, the annotated ST13P6 transcript was not found in our RNA-*Seq* mRNA and ncRNA data sets. Furthermore, linc-NDGR1+ and ST13P6 are transcribed on the opposite strand of the non-tissue specific 427/1434 bp lincRNA sequence TCONS_00015178 (**Figure 4.10**), which was also not found in our RNA-*Seq* data. Although functionally uncharacterized, the TCONS_00015178 lincRNA was found in the LNCipedia and lncRNASNP databases under the alias lnc-ST3GAL1-1 (1332 bp). According to LNCipedia data, the lnc-ST3GAL1-1 locus is highly conserved in human and mice. the apparent lncRNA isoforms; TCONS_00015178 (1434 bp) and lnc-ST3GAL1-1 (1332 bp) had independent TSS's. These transcripts were functionally uncharacterized in the literature with the exception of TCONS_00015178 lincRNA, whose transcription is independent of poly(A)-binding protein nuclear 1 (PABPN1) in HeLa cells (Beaulieu *et al.*, 2012). Collectively, this data suggests this particular locus to be bidirectionally transcribed to yield varied ncRNA isoforms from each strand.

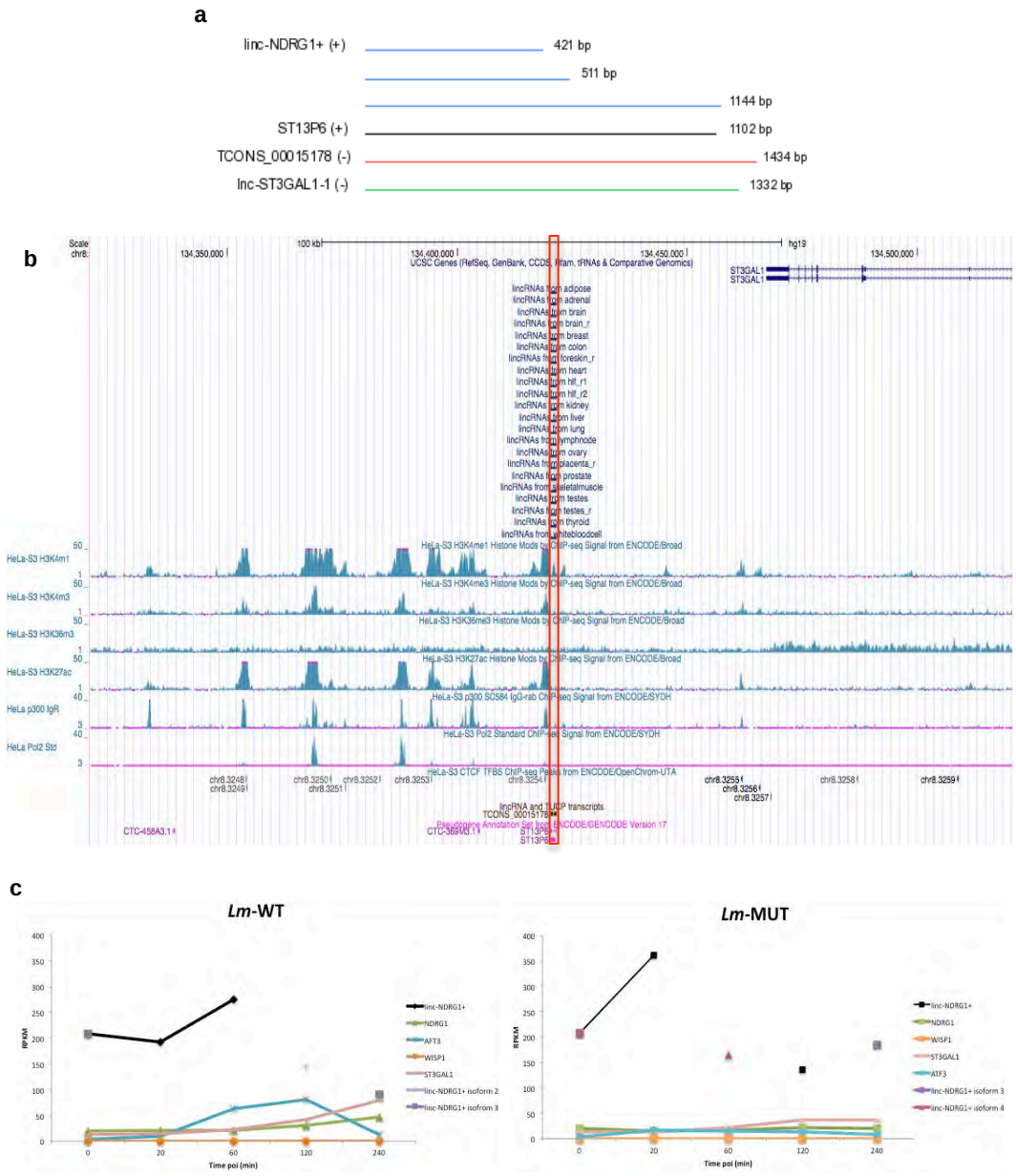


Figure 4.10: **Linc-NDRG1+ genomic properties. a.** Illustration of linc-NDRG1+ isoforms alignment to ST3P6 and lncRNAs TCONS_00015178 isoforms (encode) and lnc-ST3GAL1-1 (LNCipedia).
b. Screenshot of the 100 kb region surrounding the linc-NDRG1+ locus taken from the UCSC genome broswer. **c.** Graphs depiciting differential expression changes of linc-NDRG1+ and related genes during *Listeria* infection time course taken from the RNA-*Seq* data.

Intriguingly, linc-NDRG1+, its isoforms, the aligning ST13P6 showed varied CP scores. Linc-NDRG1+ had CP scores 2.67446, 3.62616 and 7.41965 for the 421, 511 and 1144 bp isoforms respectively suggesting this lincRNA is coding. Similiarly, the ST13P6, an apparent isoform of linc-NDRG1+, has a CP score of 6.85668 encoding a longer, yet similar open reading frame (ORF). However, ST13P6 has

been validated pseudogene with no known protein products. The CP scores of the lncRNAs on the opposite strand of linc-NDRG1+ were noncoding yielding scores of -1.0491 and -1.04127 for TCONS_00015178 and lnc-ST3GAL1-1 respectively. Together, the CP scores of the bidirectional ncRNAs found at this locus suggest that the forward strand encoding lncRNAs may yield shorter ST13 polypeptides, yet lncRNAs encoded fom the reverse strand lncRNAs are noncoding.

In resting HeLa S3 cells, the linc-NDRG1+ genomic locus displayed the following features: lack of a K4-K36 domain, a high K4me1:K4me3 ratio and was occupied by enhancer associated H3K27Ac and p300 chromatin signatures, but displayed no Pol II occupancy (**Figure 4.10b**). Despite the lack of a K4-K36 domain, this ncRNA locus displays eRNA-assoicated chromatin signatures. In additon, it displays high lincRNA conservation in a variety of human tissues (**Figure 4.10b**) and in mice (LNCipedia, Emsembl Compara API). Intriguingly, this locus is also located 1.175 kb upstream from a CTCF site (**Figure 4.10b**).

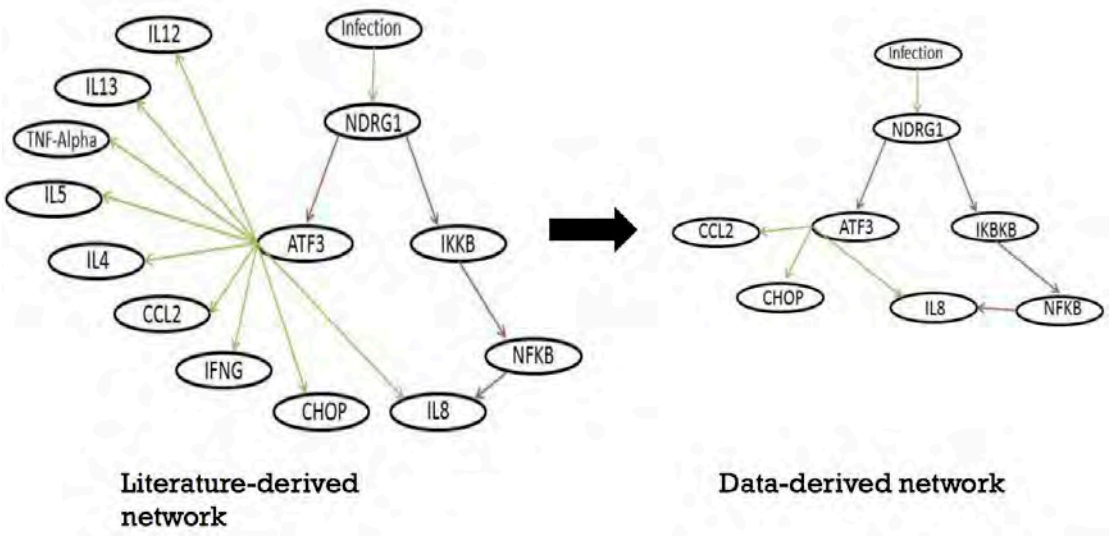


Figure 4.11: **Network analysis on NDRG1 gene network.** Green arrows show upregulated genes, Red arrows show downregulated genes. **Joana Cruz**.

As there are no genes located within 100 kb of the linc-NDGR+ locus other than ST3GAL1, the genomic region investigated was extended to 300 kb in order to identify potential linc-NDRG1+ regulated genes. Literature analysis was then employed on these proximal genes. Of particular interest is the NDRG1 gene located 110 kb away from linc-NDRG1+, which is a direct inhibitor of transcription factor ATF3 and NF-κβ inhibitor IKKβ. A literature-derived network of NDRG1

uncovered a possible network of protein-protien interactions involved in inflammation. This network was investigated in the DeSeq data to yield a data-derieved network with a high concordance to the the literature-derieved network (**Figure 4.11**). Together, this data uncovers a cascade of events within the innate response that may be regulated by linc-NDGR1+.

Collectively the RNA-*Seq* data, CP scores, chromatin marks in resting HeLa cells and the literature analysis of proximal genes yielded non-concordant evidence to the non-coding capacity of lincRNA-NDRG+. However, the 2-fold increase of this transcript from a conserved lincRNA locus and its genomic proximity to immune regulating genes warrented futher investigation.

4.3.2.1.1 Linc-NDGR1+ knockdown in HeLa cells

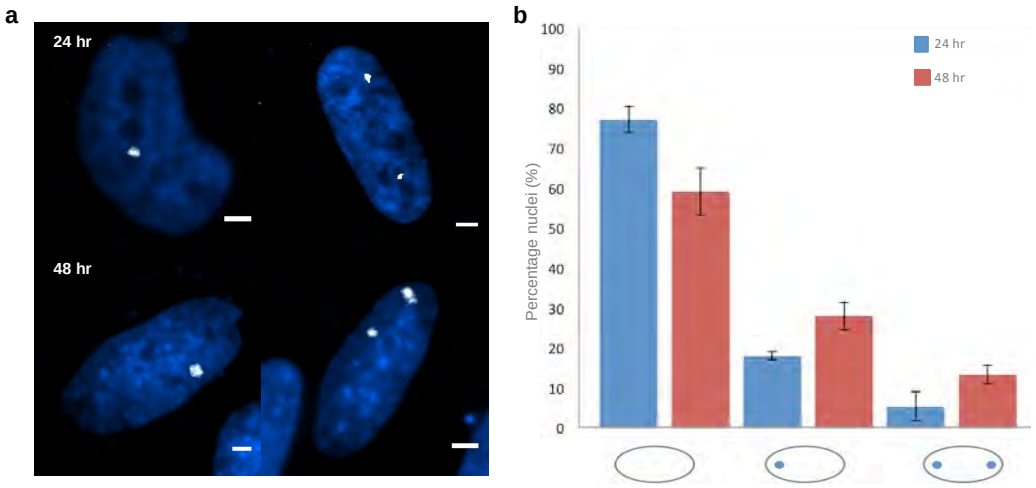


Figure 4.12**: linc-NDRG1+ TALENs induce double-stranded breaks in HeLa cells.** HeLas fixed at 24 hrs (top panel) and 48hr (bottom panel) post transfection and stained with Rb-H2Ax by indirect immunofluorescence. Blue=nuclei, white=H2Ax foci. Scale bars=2 um. **b.** Percentage HeLa nuclei with no, mono-, bi- and tri-allelic cuts after 24 and 48 hr lincNDRG+ TALEN transfections respectively (± SD of triplicate experiments).

We hypothesized that the downregulation of this linc-NDRG1+ would inhibit *Listeria* infection. To determine the effect of linc-NDRG1+ knockdown in *Listeria* infected HeLa cells, we produced TALENs against the linc-NDRG1+ locus. TALENs targetted to this locus consistently produced both mono- and bi-allelic DSBs in HeLa cells (**Figure 4.12a**). The highest percentages of HeLa nuclei with cuts were observed 48 hr post transfection as compared to 24 hr post transfection (**Figure 4.12**). Thus, the 48 hr transfection condition was used for subsequent experiments.

4.3.2.1.2. Effect of linc-NDGR1+ knockdown in Listeria infection

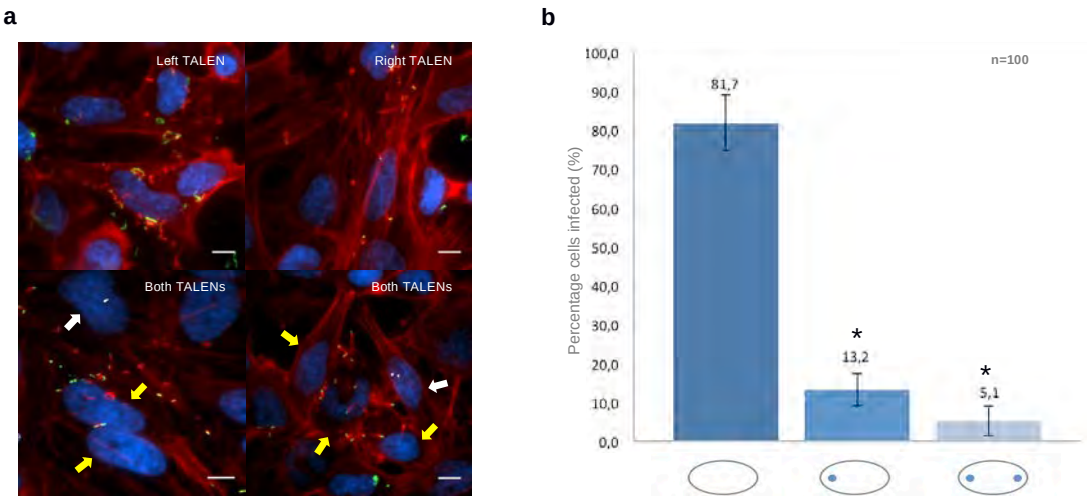


Figure 4.13**: linc-NDGR1+ targetted TALENs in HeLa cells during infection. a.** Representative images of HeLa cells transfected with 19215 TALENs for 48 hr followed by 4 hr *Listeria* infection. Blue=nuclei, green=*Listeria*, red=actin, white=H2AX foci, white arrows=TALEN cut cells, yellow arrows=uncut cells. Scale bars= 5 um. **b.** Quantitative representation of distribution of HeLa cells following 19215-TALEN transfection and *Listeria* infection. **b.** Percentage of cells with no, mono-, bi- or tri-allelic cuts after 48 hr linc-NDRG+ TALEN transfections followed by *Listeria*-WT infection 4 hr poi (± SD of triplicate experiments) * p value<0,05.

To determine the effect of linc-NDRG1+ knockdown during *Listeria* infection, TALEN transfected cells were infected with *Lm*-WT for 4 hrs (**Figure 4.13a**). As hypothesized, the percentage of cells with TALEN-mediated DSBs infected with *Listeria* for 4 hrs drastically reduced from 81,7% in uncut cells to 13,2% and 5,1% in cells with one and two DSBs per nuclei respectively (**Figure 4.13b**). This preliminary data suggests that transcriptional inhibition of the linc-NDRG1+ locus has a significant effect on *Listeria* entry into host cells which may be mediated, at least in part, by the immune response genes potentially regulated by linc-NDRG1+.

4.3.2.2 Microarray-based filtration of lncRNAs

Currently, there is no standarized filteration strategy or threshold parameters for lncRNA discovery analysis. Thus, we used standardized miRNA filtration parameters (*M-value* ≥0.5 and ≤-0.5 and *A-value* ≥10) to filter the *Lm*-WT lncRNA data. This resulted in a subset of data containing 51 highly abudant and significantly

differentially expressed ncRNAs, 24 of which were svm-predict ncRNAs. The 100 kb genomic loci surrounding these ncRNAs were explored for one or more of the following: conserved lincRNA loci, K4-K36 domain and proximal infection-related genes.

4.3.2.2.1 RCC1-lncRNA

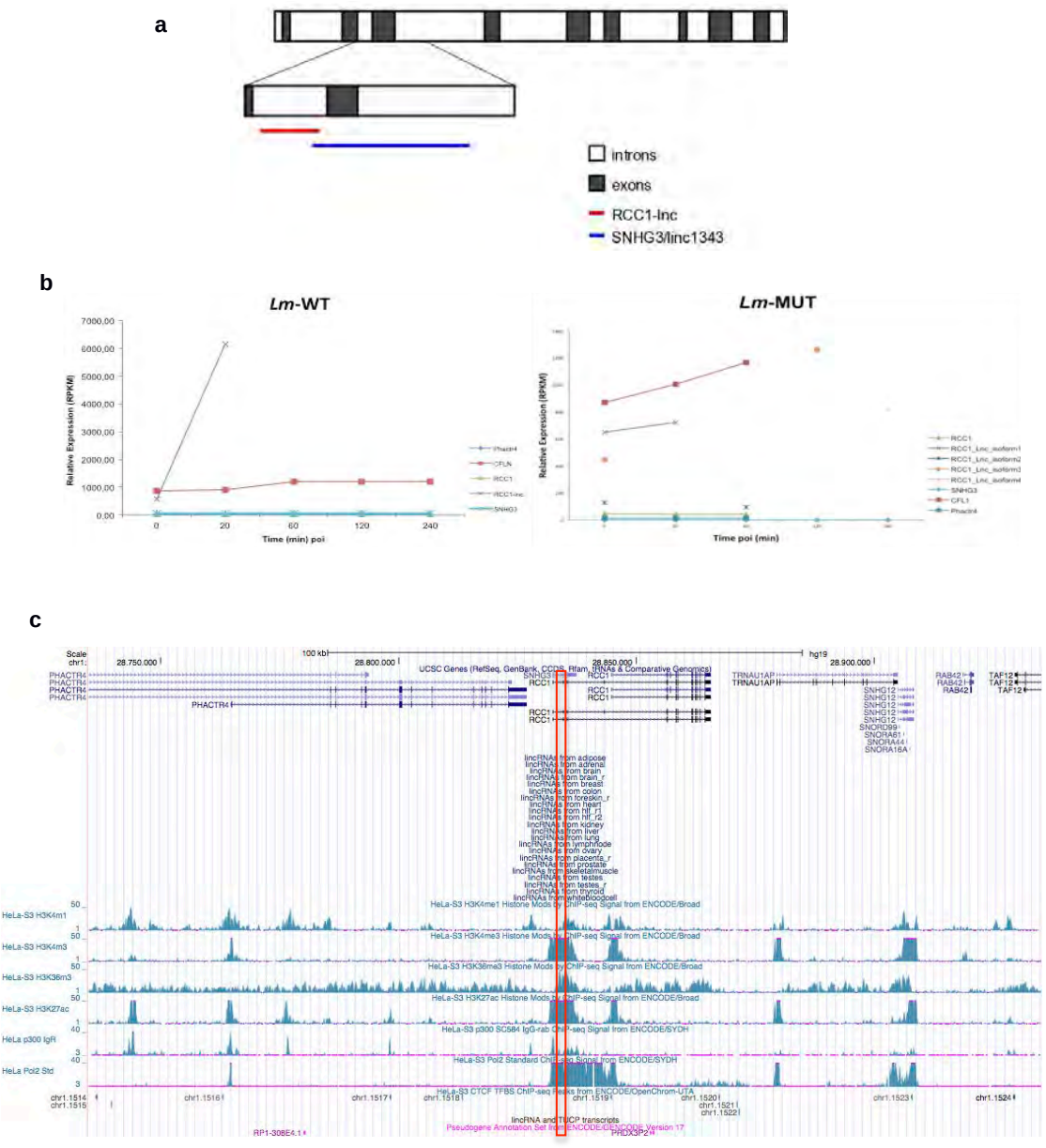


Figure 4.14*:* **Genomic and transcriptomic characteristics of RCC1-lncRNA. a.** Alignment of lnc-RCC1 and SNHG3/linc1343 to the RCC1 gene**. b.** Graphs showing differential expression patterns of RCC1-lncRNA variants and related genes, RCC1, CFLN and Phactr4 observed in RNA-*Seq* data. **c.** Screenshot of the 100 kb region surrounding the RCC1-lncRNA locus taken from the UCSC genome browser.

Analysis of the subset of data mentioned above lead to the discovery of a 556 bp intragenic lncRNA transcribed from the second intron of Regulator of Chromatin Condensation 1 (RCC1) (**Figure 4.14a,c**). This intronic lncRNA also overlaps with 5' terminus of SNHG3 (small nucleolar host gene 3). SNHG3 was found in the lncrna database (lncrnadb.org) under the alias linc1343. The ncRNA found in our RNA-*Seq* data, which will be subsequently be referred to as RCC1-lncRNA, is up-regulated 11-fold 20 min poi in the *Lm*-WT infection condition as compared to uninfected cells and undetected during later infection time points (60-240 min poi) (**Figure 4.14b**). Notably, several variants of this RCC1 intron 2 derived ncRNA exist within the RNA-*Seq Lm*-MUT data. However, these variants posses differential expression levels in the uninfected control data and are up-regulated up to 3-fold during *Lm*-MUT infection (**Figure 4.14b**). The RCC1-lncRNA is located 138 kb downstream from Phosphatase and Actin Regulator 4 (Phactr4) and 44 kb upstream from tRNA selenocysteine 1 associated protein 1 (TRNAU1AP) (**Figure 4.14c**). The presence of an actin regulator gene in close proximity to this differentially expressed lncRNA warranted further investigation.

A CP calculation of the RCC1-lncRNA sequence showed that the ncRNA was noncoding with a CP score of -1.16389. Although the RCC1-lncRNA aligns to SNHG3, several splice isoforms of SNHG3 exist ranging from ~0.9kb to ~2.6 kb. Notably, one of these variants, ~2.3 kb in length has been validated as a lincRNA and annotated as linc1343 (Guttman *et al.* 2011). This lincRNA shall subsequently be referred to as SNHG3/linc1343. Despite being validated as a lincRNA by K4-K36 domain (Guttman *et al.* 2009) and RNA-*Seq* (Guttman *et al*., 2010), a CP calculation of the SNHG3/linc1343 sequence displayed weak coding potential with a CP score of 0.527532. SNHG3/linc1343 interacts with repressive chromatin-binding proteins PRC1, PRC2, JAIRD1B and SUV38H1 in mouse embryonic stem cells (mESCs) (Guttman *et al.*, 2011). RNAi mediated knock- down of linc1343 in mESCs lead to down-regulation of stem cell markers Nanog, Sox2, Klf4 and Oct4 as well as alterations in mESC morphology, suggesting this lncRNA is involved in maintaining pluripotency in these cells (Guttman *et al.*, 2011). SNHG3/linc1343 was also shown to have a low (<2hr) half-life in mouse N2A (neuroblastoma) and human B cells (Friedel *et al.* 2009, Clark *et al.,* 2012). Notably, SNHG3/linc1343 is not highly conserved (40-50%) between mouse and human cells.

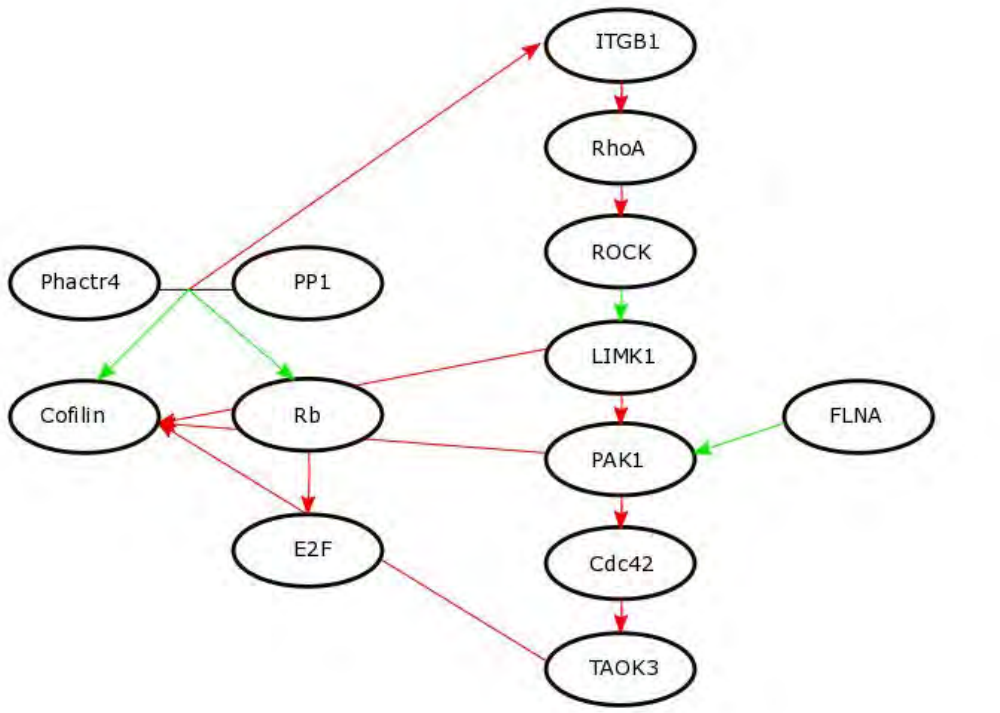


Figure 4.15: **Gene network derived from literature studies** showing the relationship between Phactr4, Filamin A and other proteins directly or indirectly regulating cofilin activity. Green arrows = up-regulated genes, red arrows = down-regulated genes.

Literature studies uncovered RCC1-lncRNA proximal gene, Phactr4, as a phosphatase that binds and dephosphorylates protein phosphatase 1 (PP1) (Allen *et al.*, 2004). The Phactr4-PP1 complex activates actin-severing protein, cofilin (CFL1), by dephosphorylating cofilin at Ser 3. The activation of cofilin results in the severing of actin filaments leading to an increase G actin monomer pool within the cell cytoplasm. Additionally, there is evidence indicating that cofilin is recruited by *Listeria* to the phagocytic cup during bacterial entry, suggesting a role for cofilin in *Listeria*-mediated phagocytosis (Bierne *et al.* 2001). The phosphorylation state of cofilin at Ser 6 regulated the formation or disruption of phagocytic cup (Bierne *et al.*, 2001). Cofilin activity has also been shown to regulate mitochondrial size, morphology, motility, location and dynamics in neuronal cells in an SSH1 (slingshot 1)-dependent manner (Beck *et al.*, 2012). Further investigation, suggested that cofilin deactivation is a downstream consequence of Filamin A binding to PAK1 (p21-activated kinase 1), a direct inhibitor of cofilin activity (Edwards *et al.*, 1999, Vadlamudi *et al.,* 2004). In addition, cofilin kinase LIMK1 inhibits cofilin activity resulting in the accumulation of F-actin fibers and inhibiting *Listeria* entry into Vevo

and Ref2 cells (Bierne *et al.* 2001). Other proteins were found to be involved in cofilin functioning including proteins within the Rho and Rac signaling pathways (Edwards *et al.*, 1999). This information was compiled into a cascade map, which also describes the differential expression pattern of the genes in the RNA-*Seq* data (**Figure 4.15**).

4.3.2.3 The hunt for enhancers

4.3.2.3.1 Chromatin modification marks

In the hunt for eRNAs, significantly up-regulated (M value>0.5) and downregulated (M value<-0.5) were identifed and their H3K4me1-H3K4me3 ratios were investigated on ENCODE. Notably, svm-predict transcripts were excluded from the analysis. This analysis did not yield workable ncRNAs with the necessary eRNA chromatin marks.

4.3.2.3.2 Gene ontology

From the 37 differentially expressed genes with the above-mentioned gene ontologies related to infection and inflammation 3, 10 and 2 genes had validated lincRNAs, pseudogenes and TUCPS, respectively, within a 100 kb from their genomic loci. Additionally, only 10 transcripts within 100 kb of the genomic loci of these genes were identified from the RNA-*Seq* data. Notably, all the lncRNAs found in the data were svm-predicted transcripts. This strategy was ineffective in discovering a workable number of eRNAs from the BGIAmericas data corresponding to high K4me1:K4me3 ratios in close proximity (within 100 kb) to significantly differentially expressed genes within the above-mentioned ontology classes.

4.3.2.4 lncRNA barcode

Stringent paramaters (1 < M value > -1, RPKM >100) were used in an attempt to identify a lncRNA barcode for early (60 min poi) *Listeria* infection from the RNA-*Seq* data. In addtion, several exclusion parameters were used to incorporate ease of wet lab analysis. This resulted in 4 downregulated (M value <-1, green rows in **Table 7**) and 42 upregulated lncRNAs (M value >1, red rows in **Table 7**) in early *Listeria* infection in HeLa cells. All the intragenic lncRNAs identified from this strategy showed differential expression patterns from their parental mRNAs. In additon, for wet lab validation assays, 6 “housekeeping” lncRNAs with minimal differential

expression patterns (0.2 < M value > -0.2, white rows in **Table 7**) in early *Listeria* WT infection were identified (white rows, **Table 7**).

Table 7: LncRNA barcode

ncRNAID	strand	chromosome	Start	End	geneLength	Control_HeLa-RPKM	LM1_HeLa_60-RPKM	M VALUE	Notes on genomic loci	Histone modification traces at loci
CUFF.4137.1	+	chr11	65190277	65192379	2103	321,7632842	99,89423228	-1,68752643	intergenic, NEAT1 (-1.0) isoform, high lncRNA abundance	K4m3, K4meo1,k27ac, POLR2A
CUFF.8851.1	+	chr16	22494372	22497475	674	106,0080651	39,12399038	-1,4380486	overlaps to SMG1P1 (n/a) pseudogene	K36me3
CUFF.13656.1	+	chr2	139046044	139046354	311	306,3212884	127,1844189	-1,2681236	intergenic	K4me3, K36me3, K27ac, p300, PolR2A
CUFF.19844.1	+	chr5	176730854	176731116	263	151,8160682	76,86946754	-0,98186191	exonic to PRELID1 (0.5)	K4me3, K27ac
CUFF.20096.1	+	chr6	26251838	26252285	448	248,6095129	215,8223696	-0,20	promoter upstream transcript from HIST1H2BH	K4me3, K27ac, p300, POLR2A
CUFF.4148.1	+	chr11	65266525	65273981	7457	1250,389081	1234,257576	-0,02	overlaps MALAT1 ncRNA	K4me3, K36me3, K4me1, K27ac, p300
CUFF.9026.1	+	chr16	34404139	34404557	419	270,8317799	270,6184938	0,00	antisense to UBE2MP1 pseudogene	
CUFF.17873.1	+	chr4	88812972	88813776	805	797,9434014	798,1844852	0,00	intergenic	
CUFF.4890.1	+	chr12	6976702	6978153	270	306,1377426	309,2727289	0,01	intronic to TPI1	K4me3, K36me3, K4me1, K27ac, POLR2A
CUFF.17219.1	+	chr3	184032367	184053127	5418	265,29984	303,2159802	0,19	within 3'utr of AHCYL1	K36me3
CUFF.14311.1	+	chr2	234197331	234197589	259	89,25094192	183,2634175	1,03797953	intronic ATG16L1 (0.1)	K36me3
CUFF.7476.1	+	chr15	21937777	21938024	248	59,31523376	124,0503942	1,06445174	antisense LOC646214 (-0.9) ncRNA, high linc abundance	
CUFF.5325.1	-	chr12	53433599	53435916	2318	132,0585844	276,8160159	1,06774933	intronic to EIF4B (-0.2) includes 3'utr	K36me3
CUFF.23468.1	-	chr9	21911113	21911364	252	83,33103159	177,8899531	1,09302089	intronic MTAP (-0.3)	
CUFF.11973.1	-	chr19	14625599	14629223	2238	76,87494891	171,6339273	1,12169569	overlaps 2 introns of DNAJB1 (0.5)	K4me3, K36me3, K4me1, K27ac, POLR2A
CUFF.5739.1	+	chr12	104559048	104655685	638	1690,825054	3763,932284	1,15451326	intronic to TXNRD1 (-0.01)	K36me3
CUFF.5415.1		chr12	56194345	56211510	280	100,0692379	226,0248816	1,17548305	overlaps 2 introns of SARNP (0.7), DNAJC14 (-0.1)	K4me3, K36me3, K4me1, K27ac, p300, POLR2A
CUFF.9907.1	+	chr17	19091322	19092085	764	743,5773084	1761,422815	1,24418662	overlaps DQ490867 (n/a novel nas)	K27ac
CUFF.20154.1	-	chr6	34392855	34393887	223	87,56323152	212,8485432	1,27501873	intronic RPS10 (0.4), RPS10-NUDT3 (n/a)	K4me3, K27ac
CUFF.9833.1	+	chr17	16284251	16285978	1011	59,59623266	144,3240534	1,27616082	promoter upstream transcript of UBB (2introns, 3'utr) (-0.3)	K4me3, K36me3, K4me1, K27ac, POLR2A
CUFF.14653.1	+	chr20	34637682	34638911	1230	128,2764382	313,7192688	1,27901479	antisense to LINC00657(n/a) ncRNA	K36me3, K27ac, POLR2A
CUFF.22085.1	+	chr7	128213439	128210678	240	186,7959108	454,1425861	1,28168246	intergenic, high linc abudance	
CUFF.3674.1	+	chr11	14619392	14620108	717	98,67357211	240,2805962	1,28398451	antisense to PSMA1 (0.7) intron	
CUFF.1301.1	-	chr1	113213012	113214188	1177	77,36873959	191,181246	1,30511833	overlaps CAPZA1 (-0.4) 3'utr	K36me3
CUFF.17088.1	+	chr3	160232744	160233008	265	113,6635695	285,2557206	1,32748615	antisense to KPNA4 (-0.02) intron, antisense to SCARNA7 (0.6 to 0)	K36me3
CUFF.4355.1	-	chr11	72287219	72289557	1565	66,69151127	170,1806084	1,35149161	overlaps 2 introns and 3'utr of PDE2A (0.9)	K36me3
CUFF.18650.1	+	chr5	34907003	34907603	601	45,45574367	118,4656201	1,38193393	antisense to 3'utr RAD1 (-0.4)	K36me3
CUFF.14857.1	-	chr20	48697962	48700977	1802	111,5644278	291,2066778	1,38416634	overlaps 2 introns of UBE2V1 (-0.6)	K36me3
CUFF.1766.1	+	chr1	160287065	160287527	463	152,8055181	408,1682827	1,41746751	antisense to COPA (-0.2) intron, overlaps with SUMO1P3 (n/a) pseudogene	K36me3
CUFF.24842.1	-	chrX	69506477	69507681	651	46,26857237	124,2191703	1,42478335	overlaps 2 introns of PDZD11 (-0.5)	
CUFF.3473.1		chr11	319720	320892	599	173,0746752	481,3143462	1,47558478	intronic to IFITM3 (-0.5)	
CUFF.4872.1	+	chr12	6619397	6619940	544	410,7621475	1142,358189	1,47563994	intronic to NCAPD2 (0.1), overlaps SCARNA10 (0.3-8rpkm) ncRNA	K36me3
CUFF.18007.1	+	chr4	114135198	114135559	362	766,2760427	2175,610892	1,50548445	intronic to ANK2 -0.1)	
CUFF.24323.1	-	chr9	134735667	134738483	427	47,57389999	135,8619507	1,51389927	intronic to MED27 (0.1)	
CUFF.1961.1	+	chr1	183595479	183596373	895	41,48121482	118,8349315	1,51842893	antisense to ARPC5 (-0.3) 3'utr	
CUFF.24776.1	+	chrX	53176425	53176828	404	138,7098347	472,1284311	1,76710928	intergenic	
CUFF.4049.1	+	chr11	62619469	62620133	665	48,45457835	166,5446496	1,78120411	antisense to SNHG1 (-0.5) 3'utr	K4me3, K36me3, K4me1, K27ac
CUFF.5597.1	+	chr12	69053099	69053444	346	64,78470893	223,5570441	1,78691776	exonic to RAP1B (-0.5)	
CUFF.1516.1	+	chr1	150546988	150549112	2125	99,00553181	225,9202398	1,87157636	antisense to MCL1 (1.5) 3'utr	K4me3, K4me1, K27ac
CUFF.11815.1	+	chr19	12902316	12904467	2152	224,2722744	890,4222668	1,98923839	exonic to JUNB (3.4)	K4me3, K4me1, K27ac, p300, POLR2A
CUFF.2526.1	+	chr10	3818984	3819903	920	28,17166589	115,6057214	2,03689602	antisense to KLF6 (2.0) promoter	
CUFF.9020.1	+	chr16	31580369	31580628	260	21,55337432	101,4214212	2,23437674	antisense to YBX3P1 (n/a) pseudogene	
CUFF.25157.1	+	chrX	134184976	134185546	571	23,30859657	112,3746979	2,26938319	antisense to FAM127B (0.2) 3'utr	K36me3
CUFF.13583.1	+	chr2	128951863	128052149	287	24,4071312	119,4440431	2,29096027	within 3'utr of UGGT1(-0.2)	
CUFF.7756.1	+	chr15	44084517	44087034	1301	25,30574886	131,7466041	2,38022761	overlaps 2 SERF2 (-0.6) introns	K4me3, K36me3, K4me1, K27ac
CUFF.14522.1	+	chr20	17943155	17943579	425	191,1911087	1050,646378	2,45618984	antisense to SNX5 (-0.5) intron	
CUFF.22309.1	+	chr7	23624740	23625045	306	13,73493344	100,5375744	2,87180666	antisense to CLK2P (-0.1) pseudogene	
CUFF.15265.1	+	chr21	46643461	46645743	283	12,37605416	105,6025163	3,09302089	exonic to ADARB1 (0.1)	K36me3
CUFF.18821.1	+	chr5	56218582	56218875	294	14,29560542	128,5591258	3,1687884	intronic to SETD9 aka C5orf35 (-0.2), antisense to MIER3 (-0.5) promoter	K36me3
CUFF.17029.1	+	chr3	151985861	151986066	206	17,00205498	174,943746	3,36311006	promoter derived from MBNL1 (-0.2)	K4me3, K4me1
CUFF.24044.1	+	chr9	120476929	120477220	292	7,19676026	105,3578681	3,87180666	exonic to TLR4 (0.4)	
CUFF.1881.1	+	chr1	174967037	174967268	232	0	106,0844751	#DIV/0!	intergenic, high linc abundance	

The epigenetic signatures of these lncRNAs were also identified from the UCSC genome browser. Many of them did not show any of the known lncRNA characteristic epigentic traces (K4-K36 domain and K4me1:K4me3 ratio) or active chromatin marks (K27ac, p300 and Pol2A occupancy). However, 14 of these differentially expressed lncRNAs identified from this strategy showed 2 or more active chromatin marks, as well as characteristic lncRNA epigentic marks. Only 7 of these were mapped to regions encoding validated lincRNAs and/or high lincRNA abundance across mammalian tissues. In addition, 3 of these lncRNAs overlapped with pseudogene encoding regions. Of the 46 lncRNAs identified, 6 lncRNAs displayed characteristic eRNA chromatin signatures (K4me1:K4me3 ratio) suggesting these lncRNAs may be eRNAs.

The final barcode, including 6 “housekeeping” lncRNas, encompassed an array of characteristic lncRNAs with 15 antisense ncRNAs, 5 mono-exonic ncRNAs, 8 intergenic ncRNAs, 11 intronic ncRNAs, 3 promoter-derived/upstream ncRNAs, 3 3’UTR derived ncRNAs and 7 multiexonic intragenic ncRNAs.

4.3.3 RT-PCR validations of RNA-Seq data

Currently, RT-PCR analysis used as a golden standard for validating RNA-*Seq* data. RT-PCR validation was particularly imperative in this study, as this RNA-*Seq* dataset only contained a single replicate. Before genes of interest or lncRNAs could be validated using this assay, it was necessary to identify a housekeeping or reference mRNA, to which the expression of transcripts of interest would be then compared.

4.3.3.1 Identification housekeeping genes

DeSeq analysis performed by on the mRNA data subset allowed for the identification of potential housekeeping genes within the RNA-Seq mRNA dataset. Housekeeping genes were qualified by CV values < 0.05. Of the 22 genes in this category, 5 genes were selected for RT-PCR validation (**Table 5**).

4.3.3.2 cDNA preparation

HeLa cells grown in flasks were infected with *Lm*-WT, followed by RNA extractions. RNA extractions yielded substantial RNA concentrations with relatively high purity, which was converted to cDNA for RT-PCR analysis (**Table 2**). In parallel, cells were

grown on coverslips, infected with *Lm*-WT and then samples were taken for microscopy analysis in order to characterize the infection time course (**Figure 4.16**).

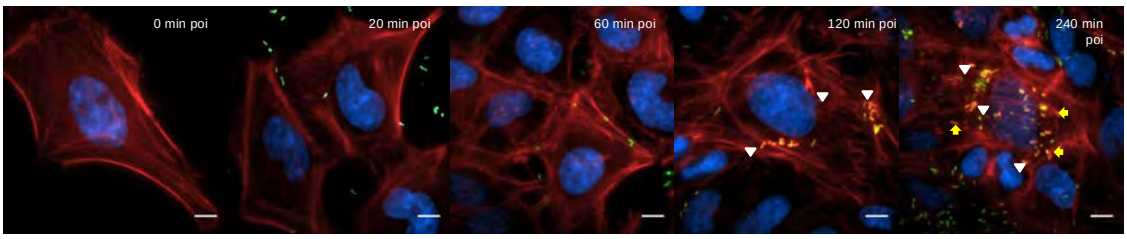


Figure 4.16**: RNA extractions from *Listeria*-infection time course in HeLa cells.** Representative images of infected cells. Nuclei=blue, Red=actin, Green= *Listeria* EDGe-GFP, White arrowheads= actin recruitment, Yellow arrows=comet tails. Scale bars=5 um.

Table 8: RNA extraction concentrations

Time point (min poi)	Total concentration (ug)	260/230	260/280
0	4.40	2.05	2.10
20	4.68	1.95	1.95
60	4.77	2.00	2.20
120	3.64	2.00	1.98
240	4.25	2.04	2.05

4.3.3.2.1 RT-PCR on “housekeeping” genes

In an attempt to validate RNA-*Seq* data and identify a “housekeeping” gene, DeSeq identified the following “housekeeping” genes (**Table 2, Figure 3.17a**): Yars, Mcdf2, Ppap2c Crtap and Hnrpa1l3. Their expression values were validated by RT-PCR analysis across the infection time course.

With regards to variance, the DeSeq and RT-PCR datasets apear to show concordance in that the expression of these genes remains relatively unchanged throughout early wild type *Listeria* infection. Furthermore, no statistically significant differences in expression data were observed with any of these genes in the conditions investigated using a paired Student’s t-test (p value<0,05). However, the expression patterns, i.e. whether expression is increasing or decreasing between time points, of Crtap and Hnrpa1l3 do not show concordance between the two types of analysis. Furthermore, Ppap2c was undetected in all biological and technical RT-PCR replicates. This may be to the low copy number of this gene with just over 100 RPKMs at throughout early infection.

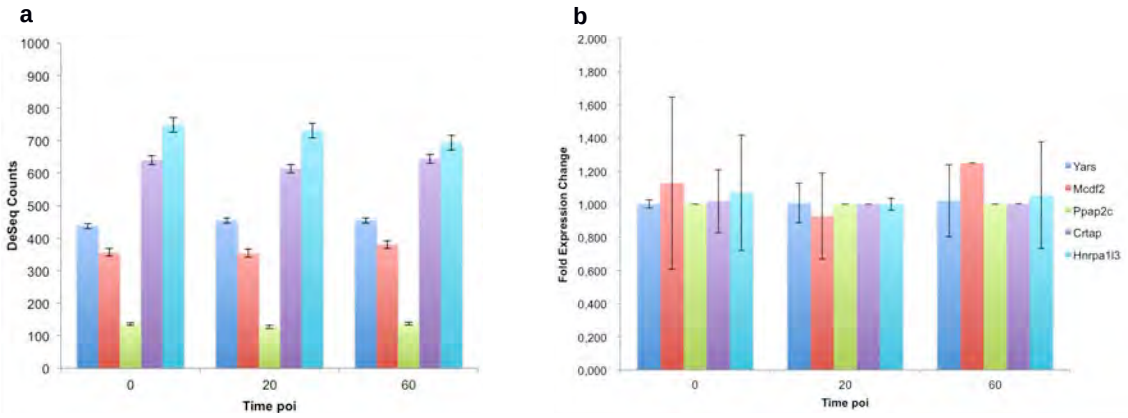


Figure 4.17: **Housekeeping genes analysis. a.** Graph displaying DeSeq data counts of "housekeeping genes" from RNA-*Seq* data. **b.** Fold expression changes of "housekeeping genes" as measured by RT-PCR. Statistical significance measured by paired Student's t-test, $p<0,05$.

Based on the experiments described above, Yars appears to be an optimal housekeeping gene since it displays the least variance across all infection conditions in the DeSeq and RT-PCR analysis (**Table 2, Figure 4.17**). However, the gene with the least fold expression change and variance according to the RT-PCR data is Crtap (**Figure 4.17**). Thus, we chose Crtap to be used as the housekeeping gene in our work.

4.4 iPSc-derived macrophages

Infection biology has become largely dependent on using cell lines to model human disease and infection. Although, cell lines have resulted in countless therapeutic molecular and cellular advances, outcomes using these models must be interpreted with caution. This is mainly due to limited physiological relevance as a result of these lines being immortal and tumorigenic, making them prone to genetic malformations including inter-chromosome translocations (Landry *et al.* 2013). Thus, the genetic landscape of cell lines is vastly different to that of a primary cell. To circumvent this constraint particularly in cellular and genetic biology, primary cells such as bone marrow derived macrophages and hepatocytes have been widely used to better represent normal physiology. However, these techniques result in limited cell numbers, introduce patient sample bias and are comparatively expensive.

Since the successful derivation of stem cells from patient cells by Takahashi and Yamanaka in 2006, iPSc reprogramming technologies have imploded cellular biology. These technologies have since made it possible for cellular biologists to effectively produce any cell type to use as physiologically relevant models in studying human biology. The use of iPSc technologies circumvents the limitations caused by cell line, mice and primary cell models by providing a highly relevant physiological model with genetic and cellular preservation and homogeneity, yet still maintaining the convenience of large numbers without having repeated biopsies for patient samples. Thus, iPSc technologies have become widely used in every sector of disease and academic research.

Activation of host macrophages is of extreme importance for mounting an innate response in efforts to clear a variety of viral and bacterial pathogens including *Listeria*. TNF-α and IFN-γ activated macrophages in particularly are important for bacterial clearance and the priming of the adaptive immune response (Pamer, 2004). However, macrophages also play an important role in *Listeriosis* functioning as cellular motors for the bacterium between organ systems during the infection cycle. *Listeria* induces phagocytosis and even efferocytosis (Czuczman *et al*., 2014) by invading macrophages where it can replicate within the phagosome (Birmingham et al., 2008) or eventually escape the phagosome, spreading intracellularly by hijacking cellular machineries such as IFN-γ response (Rayamajhi *et al.* 2010), actin

machinery (Lam *et al.* 2013). Thus, macrophages serve as an excellent study model for *Listeria* infection and indeed, several studies have been published using this model.

In this study, the oldest cell line, HeLa, was used to determine the subtle transcriptional variations caused by bacterial infection and their physiological and pathological effects. However, recently Landry *et al.* (2013) characterized the genomic landscape of HeLa cells relative to the human reference genome and identified approximately 4.5 million single nucleotide variants, 0.5 million indels and 3000 structural variants within the HeLa cell line, using deep DNA and RNA sequencing. These large numbers in genetic aberrations leave no room for any confidence in the physiological relevance of the HeLa cell line particularly for genetic and transcriptomic studies. Therefore, iPSc-derived macrophages were explored as an alternative to validate the HeLa-derived sequencing data in this study.

4.4.1 iPSc-programming

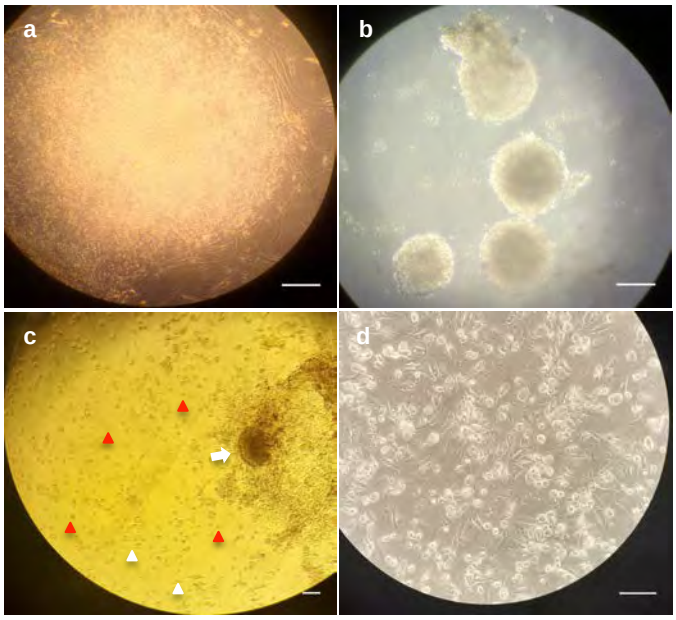


Figure 4.18: **iPSc reprogramming a.** IPSc's grown on iMefs. **b.** Embryoid bodies. **c.**Twelve-week old monocyte factory (white arrow). White arrowheads = monocytes, red arrowheads = macrophages. **d.** Mature (day 9) CD14 selected iPSc-MDM. Scale bars = 100 um.

An adapted Karlsson *et al.* 2008 iPSc-MDMs (Induced pluripotent stem cell monocyte-derived macrophages) production protocol was performed under the supervision of Dr. J. Scholefield. IPSc and IPSc MDMs characterizations used in

this study were conducted by Dr. J. Scholefield using the protocols described in Bock *et al.* (2011), Marti *et al.* (2013), excluding the teratoma assay, and the TaqMan hPSc Scorecard assay (data not shown). The morphologies of the different cell types and transitions of the protocol were used to ascertain successful iPSc reprogramming into endothelial MDMs (**Figure 4.18**). IPSc colonies grown on imbeds had distinct colony boarders and resultant embryoid bodies ranged from 80 to 200 um in diameter. Monocyte factories were harvested on a weekly basis, producing an average of 4-500 000 monocytes per well (6 well plate) per harvest. CD14+ differentiated iPSc-MDMs produced from this protocol showed a similar morphology to that described in Karlsson *et al.* (2008).

4.4.2 Listeria infection of iPSc-MDMs

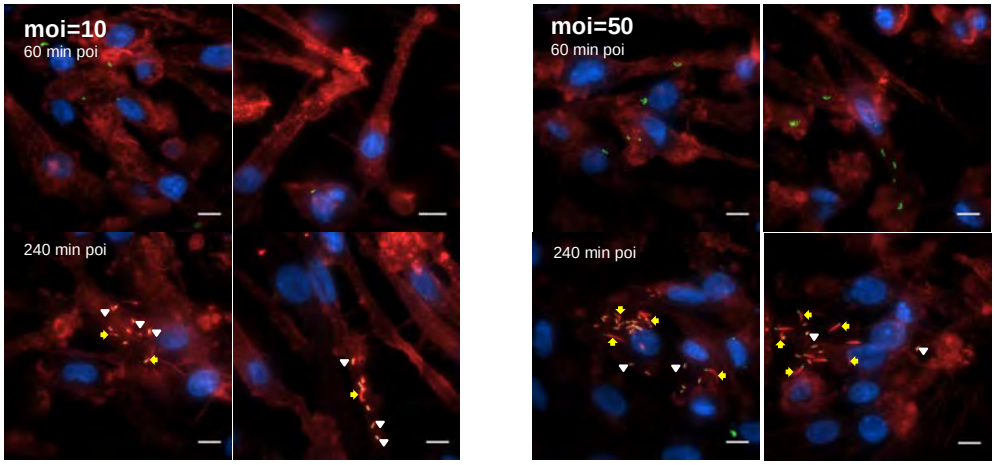


Figure 4.19**: Preliminary iPSc-MDM *Listeria* infection optimizations.** Nuclei=blue, Red=actin, Green= *Listeria* WT-GFP, White arrowheads= actin recruitment, Yellow arrows=comet tails. Scale bars=5 um.

To produce a more physiologically relevant model of *Listeria* infection in humans, several *Listeria* preliminary infection optimization experiments were performed on iPSc-MDMs. Using only the wild type strain of the bacterium; infections were conducted as per HeLa infections in the relevant medium. Initially, it was thought that due to the phagocytic nature of macrophages; infections may progress faster with higher bacterial loads compared to their HeLa counterparts thus iPSc-MDM infection optimizations began with decreasing infection mois (**Figure 4.19**). Preliminary infections with varied mois have not provided conclusive evidence for increased bacterial loads in iPSc-MDMs as compared to HeLa cells. Further

experiments are required before RNA-*Seq* infections can be mimicked in iPSc-MDMs including infection time courses and more infection moi conditions.

Chapter 5: Discussion and Recommendations

Long noncoding RNAs have been identified as key molecular players mediating an array of regulatory interactions governing a variety of cellular functions (Huarte *et al.* 2010, Lee *et al.* 2010, Clemson et al, 2010, Hung *et al.* 2011). In recent years, a number of studies have presented the exciting capability of lncRNAs in mediating the host response and immune functioning against microbial infection (Illot *et al.* 2013, Rapicavoli *et al.* 2013, Krawczyk *et al.* 2014, Li et al., 2014). Extensively studied in cancer research several lncRNAs have been identified as important diagnostic markers and therapeutic targets. However these aspects of lncRNA functioning have been neglected in infection biology highlighting the cavities our understanding of the biological significance of lnRNAs in infection biology. Currently, only a handful of lncRNAs have been discovered, mechanistically and functionally characterized in specifically pathogenic infection including linc-Cox2 (Carpenter *et al.* 2013, Heward and Lindsay, 2014) and NeST (Gomez *et al.* 2013).

In this study, we sought to identify potential lncRNAs involved in promoting or inhibiting *Listeria* infection in the HeLa cell model. Through a combination of bioinformatic, genomic, biochemical and cell biological approaches we identified a select number of lncRNAs from RNA-*Seq* data of *Listeria*-infected HeLa cells. RNA-*Seq* was conducted on samples from wild type and LLO-deficient *Listeria* EDGe infected HeLa cells for 0, 20, 60, 120 and 240 min poi. To our knowledge this is the first publication of RNA-*Seq* performed on this infection model.

The RNA-*Seq* data was provided in three data subsets: mRNAs, ncRNA and small RNAs (miRNA). For the purposes of this study, only the mRNA and ncRNA data subsets were evaluated. In efforts to identify *Listeria* infection-regulating lncRNA candidates, the RNA-*Seq* data was subjected to several lncRNA discovery strategies. As lncRNA characterization largely requires functional validation assays, we attempted to focus on identified lncRNAs that were easily amenable to wet lab validation assays. LncRNA identification strategies in this study generally followed the Rinn & Chang (2012) lncRNA discovery pipeline, with a few exceptions, which included literature analysis of candidate lncRNA proximal genes (typically within 100 kb) as a “guilt by association” approach.

Candidate lncRNAs selected for further analysis and preliminary validation studies in this study were proximal to innate immunity and infection-related genes in fulfillment of our aim to identify potential infection-regulating lncRNAs. Furthermore, lncRNA discovery was performed based on lncRNA differential expression patterns, length, lncRNA-characteristic genomic chromatin signatures including the putative "K4-K36" domain (Guttman et al., 2009, Khalil *et al.* 2009), high H3K4 me1:me3 ratio (De Santa *et al.* 2010, Marques *et al.* 2013) and coding potential calculations (http://cpc.cbi.pku.edu.cn/).

Filtration of the RNA-*Seq* mRNA dataset based on differential gene expression and transcript abundance yielded a heat map with 133 differentially expressed genes across the 4 hour infection time course and WT and MUT infection conditions. Less than 30% of the significantly differentially expressed and high abundance mRNA transcripts were up-regulated, with the majority of genes being down-regulated during *Listeria* infection. Of the differentially expressed and abundant mRNAs observed in this study, 9 were LLO-dependent and unique to wild type infection and 16 were unique to *Lm*-MUT infection. LLO-dependent genes had varied cellular functions that were not directly involved in infection. These genes were shown to be involved in transcriptional regulation, mitochondrial function, extracellular matrix composition and translation. LLO-dependently up-regulated KRT8 and Cox6AA1 have roles in mitochondrial functioning. Recent studies have revealed drastic alterations in mitochondrial dynamics, including mitochondrial fragmentation, at the onset pathogenic *Listeria* infection (Stravu *et al.* 2011). To test this hypothesis, lncRNA knockdown assays must be performed to determine their effect on the progression and physiological responses of *Listeria* infection. Although not obviously related to infection, with the exception of mitochondrial functioning genes, the functioning of these cellular processes may indirectly affect *Listeria* infection.

In the HeLa cell model, *Listeria* infection induced a global reduction in the transcription of ribosomal protein (RP) genes, which account for ~40% of the significantly differentially expressed genes during the infection time course investigated here. This down-regulation of ribosomal genes suggests that *Listeria* may induce global translation inhibition by inhibiting RP gene transcription and subsequently limiting ribosomal machinery available within the host cell. However, it is important to note that several RPs have been shown to perform extra-ribosomal functions and these functions may be targeted by the bacterium. In addition, the down-regulation of genes implicated in protein synthesis including RPs is correlated

with cellular stress responses induced by bacterial pathogens (Chakrabarti *et al.*, 2012, Fontana *et al.*, 2012, McEwan *et al.* 2012), suggesting that the global down-regulation of RP genes observed in this study may be a result of cellular stress. Notably, a macroarray transcriptional profile of *Listeria* infected mouse spleens by Camejo *et al.* (2009) resulted in the up-regulation of a significant portion of RP genes. However, in our RNA-*Seq* data only one RP gene, RPS2, was significantly up-regulated. Despite the different infection models and techniques used in these studies, these conflicting results warrant further investigation into the global and specific functions of RPs during *Listeria* infection.

In this study, our focus was on the discovery of lncRNAs involved in *Listeria* infection. Using different lncRNA discovery strategies, we discovered a number of lncRNAs potentially involved in infection regulation. Three of these identified lncRNA candidates were intensively investigated, two of which were subject to preliminary cell biological validation assays. We have shown that the transcription of two lncRNA candidates, possibly induced by the bacterium, promotes *Listeria* infection in the HeLa cell model. Literature analysis of "guilty by association" genes proximal to the lncRNA candidates led to the formulation of hypothesis of the functional mechanisms of these lncRNA candidates.

The first lncRNA candidate of interest, termed EMD-lncRNA, was identified from genomic loci analysis of heat map genes. The EMD-lncRNA was found ~32kb downstream from the up-regulated actin regulating gene FLNA transcriptional start site (TSS) within 3' UTR of nuclear actin regulating gene emerin (EMD). The correlative expression patterns of EMD-lncRNA with proximal gene FLNA, the contrasting expression patterns of the lncRNA compared to its parental gene EMD and a low CP score suggest this candidate lncRNA is a noncoding transcript that regulates FLNA expression in *cis* during *Listeria* infection. The proximity of actin binding and regulating genes, FLNA and EMD, to this lncRNA made it an exciting lncRNA given the well-established alteration of actin dynamics during *Listeria* infection particularly for intra- and inter-cellular spread.

Furthermore, the contrasting expression patterns of EMD-lncRNA and its parental gene, EMD, suggest this lncRNA may posses an EMD-independent TSS. However, this hypothesis was disproved by lack of a detected TSS at the EMD-lncRNA locus in the FANTOM5 CAGE-*Seq* data (The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014). Nonetheless, TALEN-mediated knockdown of EMD-

lncRNA, and consequently its parental gene EMD, performed in this study resulted in the infection inhibition in both mono and bi-allelically TALEN-cut cells. The RNA-*Seq* data suggests EMD down-regulation may be a consequence of *Listeria* infection, thus exacerbated down-regulation of EMD is not likely to inhibit infection. This suggests that the phenotypic response observed in TALENs targeted to the EMD 3' UTR locus is consequence of the EMD-lncRNA knockdown only. However, this needs to be confirmed by a control experiment where a region further upstream of the EMD locus not encompassing the EMD-lncRNA is targeted by a separate set of TALENs to ensure the infection inhibition response observed in this study is solely due to EMD-lncRNA knockdown.

A major limitation on the possible functional validation approaches of EMD-lncRNA functioning arises from its short length. For this reason, RNA localization assays such as smFISH cannot be performed for this lncRNA. Future work in this study will begin with TALEN the biochemical detection and validation of the endogenous targeted cleavage by the TALEN pair using a Surveyor mutation detection assay. RT-PCR expression analysis before and after *Listeria* infection, as well as in EMD-lncRNA knockdown cells, EMD-lncRNA-independent TALEN-mediated EMD knockdown cells will then be performed. Following which, the cellular and genomic localization of EMD-lncRNA will be determined using tyramide signal amplification fluorescent *in situ* hybridization (TSA-FISH) and ChIRP (chromatin isolation by RNA purification) respectively. In addition, EMD-lncRNA interacting proteins will be identified using RNA-pull down assays and mass spectrometry.

Current literature does not contain set parameters for lncRNA identification from sequencing data. Thus, the filtration strategies used in this study involving the ncRNA data subset were specific for this dataset. A filtration strategy using consensus miRNA data thresholds yielded abundant and highly differentially expressed lncRNA candidates. Of particular interest was an intronic lncRNA candidate located ~140 kb downstream from the Phactr4 TSS termed RCC1-lncRNA. This candidate was up-regulated 11-fold at 20 min poi and undetectable in the later time points in the *Lm*-WT data. RCC1-lncRNA contained the putative lncRNA consensus K4-K36 domain, had noncoding CP score, and was transcribed from the SNHG3 locus known to encode several SNHG3 splice isoforms including a characterized lncRNA, linc1343 (Guttman *et al.* 2011). Collectively, this data strongly suggests this lncRNA candidate is a noncoding isoform of SNHG3. "Guilt by association" and literature analysis suggests RCC1-lncRNA may regulate Phacrt4

transcription or Phactr4-PP1 binding in order to shift the cofilin-mediated G actin pool within host cells. Phactr4-PP1 binding results in the dephosphorylation and activation of filamentous actin severing protein cofilin that increases the G actin pool available for comet tail formation during *Listeria* infection. Additionally, cofilin has been implicated in *Listeria* mediated phagocytosis (Bierne *et al.* 2001). Collectively, the up-regulation of EMD-lncRNA exclusively at the onset of *Listeria* infection and literature studies suggest EMD-lncRNA may indirectly regulate cofilin-mediated actin dynamics during *Listeria* infection.

Validation studies of the potentially infection-regulating RCC1-lncRNA have not been attempted as of yet, however further work will involve RT-PCR and/or single molecule FISH assays to validate RNA-*Seq* identified expression patterns during *Listeria* infection. Then, to determine the involvement of the RCC1 lncRNA during *Listeria* infection, siRNA knockdown of the RCC1-lncRNA will be performed as well as RNA-pull down assays to validate potential RCC1-lncRNA and Phactr4 binding. Furthermore, transcriptional expression, protein translation and location of proximal RCC1-lncRNA genes will be explored using RT-PCR and, immune-precipitation and immunofluorescence respectively. The effect of RCC1-lnc*RNA* on cofilin activity can be validated using immune-precipitation to determine the proportion of active (dephosphorylated cofilin) and inactive cofilin (phosphorylated) in RCC1-lnc*RNA* knockdown cells particularly during *Listeria* infection.

A filtration strategy focused on identifying highly differentially expressed lncRNAs in the 60 min poi *Lm*-WT condition was employed to increase the lncRNA candidate pool. Strikingly, this filtration strategy yielded among others, a 421 bp lincRNA candidate antisense to validated lincRNA lnc-ST3GAL1 and apparent ST13P6 isoform transcribed ~170 kb from the NDRG1 TSS termed linc-NDRG1+. However, ST13P6 and lnc-ST3GAL1 were undetected in our RNA-*Seq* data. CP scores for lnc-NDRG1+ and ST13P6 suggested these transcripts were coding, yet the sense lnc-ST3GAL1 was noncoding. Furthermore, the lnc-NDRG1+ locus did not posses the characteristic lncRNA consensus K4-K36 domain in resting HeLa S3 cells, however several enhancer-associated chromatin marks were observed including a high K4me1-K4m3 ratio, high p300 and H3K27Ac occupancies. H3K27 acetylation in combination with H3K4Me1 is correlated with enhancers near active genes (Bulger and Groudine, 2011). Furthermore, the transcription of sense and antisense transcripts at this locus suggests this locus may be an enhancer region.

We hypothesized that transcriptional inhibition of this locus in HeLa cells would inhibit *Listeria* infection. To this end, we transfected HeLa cells with a TALEN pair targeted to the linc-NDRG1+ locus followed by 4 hour *Listeria* infection. This allowed us to determine the effect of transcriptional inhibition at this locus in *Listeria*-infected cells. These preliminary results established the importance of the transcription of this locus during *Listeria* infection. Transcriptional inhibition of this locus resulted in significant *Listeria* infection inhibition. In the model used for this study, ST13P6 and lnc-ST3GAL1 were undetected suggesting that TALEN-mediated transcriptional inhibition of this locus in this model was specifically targeted against lnc-NDRG1+. Collectively, this data establishes a role for the transcription of this lncRNA in promoting *Listeria* infection in HeLa cells. Further experiments are required to conclusively determine the importance of linc-NDRG1+ transcription during *Listeria* infection. Future work will involve the validation of TALEN-mediated knockdown using Surveyor assays, RNAi-mediated knockdown and RT-PCR analysis of linc-NDRG1+ and its "guilty by association genes" during infection. Linc-NDRG1+ interacting genomic regions and protein will be determined using ChIRP and RNA-pull down assays respectively.

In addition to identifying lncRNA candidates using a modified Rinn and Chang (2012) pipeline, we attempted to identify lncRNA and eRNA barcodes for *Listeria* infection from the RNA-*Seq* data. The lncRNA barcode was produced by filtering the data according to the following parameters: differential expression change of -1>M value<1, RPKM >100 and length > 200 bp. In order to make the final lncRNA barcode amenable to wet lab validation experiments intragenic ncRNAs with similar expression patterns to their parental transcripts, as well as svm-predict ncRNAs and ncRNAs spanning more than 2 exons of their parental genes were excluded. The final barcode included 46 differentially expressed ncRNAs and 6 "housekeeping" ncRNAs spanning the array of genomic lncRNA characteristics: intronic, mono- and multi-exonic, intergenic, antisense, 3'UTR derived and promoter-derived/upstream. Of the 52 lncRNAs identified in this barcode, 2 aligned with known validated lncRNAs: Neat1 and Malat1 and only one aligned with a pseudogene (SMG1P1). Furthermore, only 6 lncRNAs within the lncRNA barcode exhibited the consensus K4-K36 domain chromatin signature in resting cells. Future work will validate this lncRNA barcode using an RT-PCR and RNAi-mediated knockdown screen.

We also attempted to identify eRNAs from the RNA-*Seq* data using the consensus high H3K4me1:H3K4me3 ratios, H3K27Ac and p300 enhancer- and eRNA-

associated chromatin signatures of resting HeLa cells. This strategy was employed to ncRNAs proximal to infection related gene ontologies. However, it did not yield a substantial number of eRNA candidates for further validation. Furthermore, the majority of eRNA candidates identified by this strategy did not represent true reads but were support vector machine predicted transcripts. The unsuccessful results of this strategy intensify the requirement for ChIP-*Seq* assays for chromatin modifications, specifically H3K4me1:H3K4me3 ratios, p300, H3K27Ac and Pol II, in combination with RNA-*Seq* to identify eRNAs.

The revelation of the highly mutated genomic landscape of HeLa cells plagued by millions of genetic aberrations by Landry *et al.* (2013) drastically emphasized the physiological irrelevance of the HeLa cell model particularly for genetic and transcriptomic studies. Thus, the deep sequencing data in this study may not accurately represent the transcriptomic responses during *Listeria* infection in a physiological context. To mitigate this we devised a strategy to validate the RNA-*Seq* data in the highly physiologically relevant iPSc-derived monocyte-derived macrophage cellular model. To this end, we have adapted a protocol for the culturing of and differentiating iPSCs into fully functional macrophages at high yields, which will to be used in future infection and validation studies (Karlsson *et al.* 2008).

In this study lncRNA discovery was performed using a single RNA-*Seq* biological replicate dataset. The insufficient number RNA-*Seq* biological replicates places low statistical power in the data set used for lncRNA discovery. Recently Liu and colleagues (2013) demonstrated that performing more biological replicates, regardless of sequencing depth increased statistical power and accuracy in RNA-*Seq* studies. Thus, the results of RNA-*Seq* data in this study must be interpreted with caution and require extensive validation studies before deemed conclusive. To this end, RT-PCR validation assays were employed on De*Seq* identified “housekeeping” genes and the genomic loci of lncRNA candidates were subjected to TALEN-mediated knockdown. The high concordance of RT-PCR data to the RNA-*Seq* dataset, suggest the RNA-*Seq* dataset used in this study accurately detected the transcriptional patterns of transcripts greater than 100 RPKMs. However, further RT-PCR validation assays of differentially expressed RNAs are required to determine the accuracy and statistical power of the RNA-*Seq* dataset.

Collectively, our findings have provided novel insight into the alterations of the transcriptional landscape during *Listeria* infection. It raises important questions that warrant further investigation in regards to both gene and lncRNA expression changes observed in the HeLa infection model. Furthermore, it emphasizes the importance of applying integrative approaches during lncRNA discovery. Identifying lncRNAs involved in regulating *Listeria* infection will provide a better understanding of the molecular environment during *Listeria* infection and provide novel targets for infection therapies. This work proposes a lncRNA barcode that can be used as a molecular diagnostic code particularly for the *Listeria*-infected HeLa cell model and potentially for other infection models. It also provides novel insights into the molecular mechanisms potentially governing specific transcriptional responses, which may be applicable to different infection models and cellular contexts. Importantly this work provides a framework which can be used by biologists to identify and characterize lncRNA functioning using simple filtration strategies and a guilt by association approaches which led to wet lab amenable and testable hypothesis'.

Chapter 6: References

1. **Alberti-Segui *et al.*** 2007. Differential function of *Listeria* monocytogenes listeriolysin O and phospholipases C in vacuolar dissolution following cell-to-cell spread. *Cell. Microbiol.* **9**(1):179.
2. **Allen *et al.*** 2004. Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**(18):7187.
3. **Anderson *et al.*** 2007. Ablation of ribosomal protein L22 selectively impairs alphabeta T cell development by activation of p53-dependent checkpoint. *Immunity*. **26**(6):759.
4. **Andersson *et al.*** 2014. An atlas of active enhancers across human cell types and tissues. *Nature*, **507:**455.
5. **Arbibe *et al.*** 2007. An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses. *Nat. Immunol.* **8**:47.
6. **Archambaud *et al.*** 2006. Control of *Listeria* superoxide dismutase by phosphorylation. *J. Biol. Chem.* **281**(42):31812.
7. **Beaulieu *et al.*** 2012. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS. Genet.* **8**(1): e1003078.
8. **Beck *et al.*** 2012. Serum Response Factor (SRF)-cofilin-actin signaling axis modulates mitochondrial dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **109**(38):e2423.
9. **Bentley *et al.*** 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 456, 53.
10. **Bertone *et al.*** 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science*. **306**:2242.
11. **Bhavsar *et al.*** 2007. Manipulation of host-cell pathways by bacterial pathogens. *Nature.* 449:827.
12. **Bi and Zigmond**. 1999. Actin polymerization: Where WASP stings. *Curr. Biol.* **9**:160.
13. **Bierne *et al.*** 2001. A role for cofilin and LIM kinase in *Listeria*-induced phagocytosis. *JCB***. 155**(1)101.
14. **Birmingham *et al.*** 2008. Listeriolysin O allows *Listeria* monocytogenes replication in macrophage vacuoles. *Nature.* **451**:350.
15. **Bock *et al.*** 2011. Reference Maps of Human ES and iPS Cell Variation Enable High-Throughput Characterization of Pluripotent Cell Lines. *Cell* **144(3)**:439-52.
16. Boneca *et al.* 2007. A critical role for peptidoglycan N-deacytelation in *Listeria* evasion from the host immune system. *Proc. Natl. Acad. Sci.* **104**(3):691.
17. **Braslavsky *et al.*** 2003. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci.* **100**:3960.
18. **Brieher *et al.*** 2004. Fascin-mediated propulsion of *Listeria* monocytogenes independent of frequent nucleation by the Arp2/3 complex. *JCB.* **165**(2):233.
19. **Bulger and Groudine *et al.*** 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* **144**(3):327.

20. **Bullard *et al.*** 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* **11**:94.
21. **Camejo *et al.* 2009.** In vivo transcriptional profiling of *Listeria* moncytogenes and mutagenesis identify new virulence factors involved in infection. *PLoS Pathog.* **5**(5): e1000449.
22. **Cameron *et al.*** 2000. Secrets of actin-abased motility revealed by a bacterial pathogen. *Nat. Rev. Mol. Cell. Biol.* **1**:110.
23. **Carninci, P. *et al.*** 2005. The transcriptional landscape of the mammalian genome. *Science*. **309**:1559.
24. **Carpenter *et al.*** 2013. A long noncoding RNA mediates both activation and repression of immune response genes. *Science*. **341**:789.
25. **Chakrabarti *et al.*** 2012. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. *Cell Host Microbe.* **12**:60.
26. **Chen and Ioannou**.1999**.** Ribosomal proteins in cell proliferation and apoptosis. *Int. Rev. Immunol.* **18**:429.
27. **Clark *et al.*** 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**(5):885.
28. **Clemson *et al.* 2009.** An architectural role for nuclear noncoding RNA: NEAT 1 RNA is essential for the structure of paraspeckles. *Mol. Cell.* **33**(6):717.
29. **Cloonan *et al.*** 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature Methods **5**:613.
30. **Cossart and Toledo-Arana.** 2008. *Listeria* monocytogenes, a unique model in infection biology: an overview. *Microbes Infect.* **10**:1041.
31. **Cossart**. 2001. The use of host cell machinery in the pathogenesis of *Listeria* monocytogenes. *Curr. Opin. Immunol.* **13**(1):96.
32. **Cuellar-Mata *et al.*** 2002. Nramp1 modifies the fusion of Salmonella typhimurium-containing vacuoles with cellular endomembranes in macrophages. *J. Biol. Chem.* **277**:2258.
33. **Cui *et al.*** 2014. The human long noncoding RNA lnc-IL7R regulates the inflammatory response. *Eur. J. Immunol.* **44**:2085.
34. Cui *et al.* 2014. The ribosomal protein S26 regulates p53 activity and response to DNA damage. *Oncogene*. **33**:2225.
35. **Czuczman *et al.*** 2014. *Listeria* monocytogenes exploits efferocytosis to promote cell-to-cell spread. *Nat.* **509**: 230.
36. **Dabiri *et al.*** 1990. L*isteria* monocytogenes moves rapidly through the host-cell cytoplasm by inducing direction actin assembly. *Proc. Natl. Acad. Sci. U.S.A.* **87**(16):6068.
37. **De Santa *et al.*** 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**:e1000384.
38. **Deisenroth and Zhang**. 2010. Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. *Oncogene*. **29**:4253.
39. **Deng *et al.*** 2012. Cytoskeletal protein filamin A is a nucleolar protein that suppresses ribosomal RNA gene transcription*. Prot. Natl. Acad. Sci. U.S.A.* **109** (5):1524.
40. **Derrien *et al.*** 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. **22**:1775.

41. **Ding *et al.*** 2010. Helicobacter pylori-induced histone modification associated gene expression in gastric epithelial cells, and its implication in pathogenesis. *Plos One.* **5(4):**9875**.**

42. **Doerr.** 2012. Predicting PPIs. Nat. Methods. **9**(12):1139.

43. **Duan *et al.* 2011.** Knockdown of ribosomal protein S7 causes developmental abnormalities via p53-dependent and independent pathways in zebrafish. *Int. J. Biochem. Cell. Biol.* **43**(8):1218.

44. **Edwards *et al.*** Activation of LIM-kinase by Pak1 couples Rac/Cdc42 GTPase signaling to actin cytoskeletal dynamics. *Nature Cell Biol.* **1**:253.

45. **Emery and Dreifuss.** 1966. Unusual type of benign x-linked muscular dystrophy. *J. Neurol. Neurosurg. Psychiatr.* 29:338**.**

46. **Engström *et al.*** 2006. Complex loci in human and mouse genomes. *PLOS Genet.* **2**(4):47.

47. **Farber and Peterkin**. 1991. *Listeria* monotcytogenes, a food-borne pathogen. *Microbiol. Rev.* **55**(4):752

48. **Fontana *et al.*** 2012. Activation of host mitogen-activated protein kinases by secreted Legionella pneumophila effects that inhibit host protein translation. *Infect. Imunol.* **80**(10):3570.

49. **Friedel *et al.*** 2009. Conserved principles of mammalian transcription regulation revealed by RNA half-life. *Nucleic Acids Res.* **37**(17):e115.

50. **Gedde *et al.*** 2000. Role of listeriolysin O in cell-to-cell spread of *Listeria* monocytogenes. *Infect. Immun.* **68**(2):999

51. **Gomez *et al.*** 2013. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the Interferon-γ Locus. *Cell*. **152**:743.

52. **Gouin *et al.*** 2005. Actin-based motility of intracellular pathogens. *Curr. Opin. Microbol.* **8**(1):35.

53. **Griffith *et al.*** 2010. Alternative expression analysis by RNA Sequencing. *Nat Methods.* **7:**843.

54. **Guttman *et al.*** 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nat.* **458**:223.

55. **Guttman *et al.*** 2010. Ab initio reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotech.* **28**(5):503.

56. **Guttman *et al.*** 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nat.* **477**:295.

57. **Hamon and Cossart.** 2008. Histone modifications and chromatin remodeling during bacterial infections. *Cell*. 4:100-9.

58. **Hangauer *et al.*** 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **9**: e1003569.

59. **Harborth *et al.*** 2001. Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J. Cell. Sci.* **114**(24):4557.

60. **Harris *et al.*** 2008. Single-molecule DNA sequencing of a viral genome. *Science*. **320**:106.

61. **Hasegawa *et al.*** 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell.* **19**(3):469.

62. **Henn *et al.*** 2012. Whole genome deep sequencing of HIV-1 reveals the impact of minor variants upon immune recognition during acute infection. *Plos Pathog.* **8**(3):e1002529.

63. **Heward and Lindsay.** 2014. Long noncoding RNAs in the regulation of the immune response *Trends. Immunol.* **35**(9):408.

64. **Hirose *et al.*** 2014. Neat1 long noncoding RNA regulates transcription via protein sequestration within subunuclear bodies. *Mol. BIol. Cell.* **25**:169.

65. **Holaska *et al.*** 2002. The nuclear envelope, lamins and nuclear assembly, *Curr. Opin. Cell Biol.* **14**:357.

66. **Huarte *et al.*** 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**:409.

67. **Huertas *et al.*** 2010. DNA resection in eukaryotes: deciding how the fix the break. *Nat. Struct. Mol. Biol.* **17**(1):11.

68. **Hung *et al.*** 2011. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet.* **43**(7):621.

69. **Hung *et al.*** 2011. Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol. **7**:582.

70. **Illot *et al.*** 2014. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nat. Comm.* **5**:3979.

71. **Jang *et al.*** 2004. RPS3, a DNA repair endonuclease and ribosomal protein, is involved in apoptosis. *FEBS Lett.* **560**(1-3):81.

72. **Kapranov *et al.*** 2010.The majority of total nuclear- encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. *BMC Biol.* **8**:149.

73. **Karlsson *et al.*** 2008**.** Homogeneous monocytes and macrophages from human embryonic stem cells following coculture-free differentiation in M-CSF and IL-3. *Exp. Hematol.* **36**(9):1167.

74. **Khalil *et al.*** 2009. Many human large intergenic noncoding RNAs associated with chromatin-modifying complexes and affection gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **106**:11667.

75. **Kim *et al.*** 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010; 465: 182

76. **Kondrashov *et al.*** 2011. Ribosome-mediated specificity in HOX mRNA translation and vertebrate tissue patterning. *Cell.* **145**93):383.

77. **Kotake *et al.*** 2011. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**:956.

78. **Krawczyk and Emerson**. 2014. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-kappaB complexes. *Elife* **3**:e01776.

79. **Lam *et al.*** 2013. Host and bacterial factors that regulate LC3 recruitment to Listeria monocytogenes during the early stages of macrophage infection. *Autophagy* **9(7)**:985-95.

80. **Landry *et al.*** 2013. The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda).* **3**(8):1213.

81. **Langmead *et al.*** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):25.

82. **Lebreton *et al.*** 2011. A Bacterial Protein Targets the BAHD1 Chromatin Complex to Stimulate Type III Interferon Response. Science. **331**:1319-21.
83. **Lee *et al.*** 2010. The X as model for RNA's niche in epigenomic regulation. *Cold Spring Harb. Perspect. Biol.* **2**:a003749.
84. **Li & Dewey**. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **12**:323.
85. **Li *et al.*** 2007. The phosphothreonine lyase activity of a bacterial type III effector family. *Science*. 315:.1000
86. **Li *et al.*** 2009. SOAP2: an improved ultrafast tool for short read alignment *Bioinformatics* **25**(15):1966.
87. **Li *et al.*** 2014. The long noncoding RNA THRIL regulates TNF-α expression through its interaction with hnRNPL. *Proc. Nat. Acad. Sci. U.S.A.* **111**(3):1002.
88. Lisnic *et al.* 2013. Dual analysis of the murin cytomegalovirus and host cell transcriptomes reveal new aspects of the virus-host cell interface. *Plos. Pathog.* **9**(9):e10003611.
89. **Liu *et al.*** 2013. RNA-Seq differential expression studies: more sequence, or more replication. *Bioinformatics* [Epub ahead of print]
90. **Livak *et al.*** 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the **2** $\Delta\Delta$**C**T Method Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-$\Delta\Delta$**C**T Method. *Methods*. **25**:402.
91. **Macville *et al.*** 1999. Comprehensive and definitive molecular cytogenic characterization of HeLa cells by Spectral karyotyping. *Cancer Res.* **59**(1):141.
92. **Malygin *et al.*** 2007. Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism. *Nucl. Acids. Research.* **35**(19): 6414.
93. **Margalit *et al.*** 2007. Barrier-to-autointegration factor-a BAFfling little protein. Trends Cell Biol. **17:**202**.**
94. **Margulies *et al.*** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* **437**:376.
95. **Marques *et al.*** 2013. Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.* **14**:131.
96. **Marquis and Hager.** 2000. pH-regulated activation and release of a bacteria-associated phospholipase C during intracellular infection by *Listeria* monocytogenes. *Mol. Microbiol.* **35**(2):289.
97. **Marquis *et al.*** 1995. The broad-range phospholipase C and a metholloprotease mediated listeriolysin O-independent escape of *Listeria* monocytogenes from a primary vacuole in human epithelial cells. *Infect. Immunol.* **63**:4531.
98. **Marti *et al.*** 2013. Characterization of pluripotent stem cells. *Nat. Prot.* **8**:223-253.
99. **Martin & Wang.** 2011. Next-generation transcriptome assembly. *Nat. Review.* **12**:671.
100. **Matragkou *et al.*** 2008. The potential role of ribosomal protein S5 on cell cycle arrest and initiation of murine erythroleukemia cell differentiation. *J. Cell. Biochem.* **104**(4):1477.

101. **McEwan *et al.*** 2012 Host translation inhibition by Pseudomonas aeruginose exotoxin A triggers an immune response in Caenorhabditis elegans. *Cell Host Microbe* **11**:364.

102. **Mercer *et al.*** 2009. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**:155.

103. **Metzker**. 2010. Sequencing technologies-the next generation. *Nature Reviews Genetics* **11**:34-46.

104. **Mortazavi *et al.*** 2008. Mapping and quantifying mammalian transcriptomes by RNA –Seq. *Nat. Methods*. **5**:621-8

105. **Nagalakshmi *et al.*** 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*.320:1344.

106. **Nagano *et al.*** 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**:1717.

107. **Nakamura *et al.*** 2011. The filamins: Organizers of cell structure and function. *Cell Adh. Migr.* **5**:160.

108. **Neumann *et al.*** 2007. Constitutive expression of human ribosomal protein L7 arrests the cell cycle in G91) and induced apoptosis in Jurkat T-lymphoma cells. *Exp. Cell Res.* **230**:252.

109. **Okoniewski & Miller.** 2006. Hybridization interactions between probe sets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7**:276-.

110. **Okuda *et al.*** 2005. Shigella effector IpaH9.8 binds to a splicing factor U2AF35 to modulate host immune responses. *Biochem. Biophys. Res. Commun.* **333**:531.

111. **Oristain *et al.*** 2009. Ribosomal protein L29/HIP deficiency delays osteogenesis and increase fragility of adult bone in mice. *J. Orthop. Res.* **27**(1):28.

112. **Pamer.** 2004. Immune responses to *Listeria* monocytogenes. *Nat. Rev. Immunol.* **4**:812.

113. **Pang *et al.*** 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* **22**(1):1.

114. **Pathak *et al.*** 2006. TLR4-dependent NF-kappaB activation and mitogen- and stress-activated protein kinase 1-triggered phosphorylation events are central to Helicobacter pylori peptidyl prolyl cis-, trans-isomerase (HP0175)-mediated induction of IL-6 release from macrophages. *J Immunol* **177(11):** 7950.

115. **Pennini *et al.*** 2006. Mycobacterium tuberculosis 19-kDa lipoprotein inhibits IFN-gamma-induced chromatin remodeling of MHC2TA by TLR2 and MAPK signaling. *J. Immunol.* **176**:4323.

116. **Pennini *et al.*** 2007. CCAAT/enhancer-binding protein beta and delta binding to CIITA promoters is associated with the inhibition of CIITA expression in response to Mycobacterium tuberculosis 19-kDa lipoprotein. *J. Immunol.* **179**:6910.

117. **Plath *et al.*** 2002. Xist RNA and the mechanism of X chromosome inactivation*. Annu. Rev. Genet.* **36**:233

118. **Podder *et al.*** 2013. AN extraribosomal function of ribosomal protein L13a in macrophages resolves inflammation. *J. Immunol.* **190**(7):3600.

119. **Poliseno *et al.*** 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature **465**:1033.

120. **Pust *et al.*** 2005. *Listeria* monocytogenes exploits ERM protein functions to efficiently spread from cell to cell. *EMBO.* **24**(6):1287.
121. **Radtke *et al.*** 2011. *Listeria* monocytogenes exploits cystic fibrosis transmembrane conductance regulator (CFTR) to escape the phagosome. *Proc. Natl. Acad. Sci.* **108**(4): 1633.
122. **Rapicavoli *et al.*** 2013. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife*. **2**:e00762
123. **Rayamajhi *et al.*** 2010. Induction of IFN-αβ enables Listeria monocytogenes to suppress macrophage activation by IFN-γ. *J. Exp. Med.* **15**;**207(2)**:327-37.
124. **Rinn & Chang.** 2012. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**:145.
125. **Rinn *et al.*** 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* **129**:1311.
126. **Robbins *et al.*** 1999. *Listeria* monocytogenes exploits normal host cell process to spread from cell to cell. *JCB*. **146**(6):1333.
127. **Robinson *et al.*** 2009. edge R: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139.
128. **Rogakou *et al.*** 1999. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *JCB.* **146**(5):905.
129. **Royce & Gerstein**. 2007. Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res.* **35**:e99.
130. **Saccani *et al.*** 2002. p38-Dependent marking of inflammatory genes for increased NF-kappa B recruitment. *Nat. Immunol.* **3**:69.
131. **Sanjayna *et al.*** 2012. A transcription activator-like effector toolbox for genome engineering. *Nat. Prot.* **7**(1):171.
132. **Schadt *et al.*** 2004. A comprehensive transcript index of the human genome generated using microarrays and computational approaches. Genome Biol. **5**:R73
133. **Schena *et al.*** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 270:467.
134. **Schlech**. 2002. Epidemiology and clinical manifestations of *Listeria* monocytogenes infection in Gram-Positive Pathogens. V.A. Fischetti, editor. *American Society for Microbiology Press*, Washington, D.C. 473
135. **Schnupf and Portnoy.** 2007. Listeriolysisn O: a phagosome-specific lysin. *Microbes. Infect.* **9**(10):1176.
136. **Schroeder *et al.*** 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**:3.
137. **Schubert *et al.*** 2002. Structure of internalin, a major invasion protein of *Listeria* monocytogenes, in complex with its human receptor E-cadherin. *Cell*. **111**(6):825.
138. **Schuerch *et al.*** 2005. Molecular basis of listeriolysin O pH dependence. *Proc. Natl. Acad. Sci. U.S.A.* **102**(35):12537.
139. **Selinger *et al.*** 2000. RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat. Biotechnol.* **18**:1262.
140. Shen et al., 2000. InlB-dependent internalization of *Listeria* is mediated by the Met receptor tyrosine kinase. *Cell.* **103**(3):501.

141. Skoble *et al.* 2000. Three regions within Acta promote Arp2/3 complex-mediated actin nucleation and *Listeria* monocytogenes motility. *JCB* **150**(3):527.

142. Smith *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. Immunol.* **63**(11):4231.

143. Stolc *et al.* 2004. A gene expression map for the euchromatic genome of Drosophila melanogaster. *Science*. **302**:655.

144. Stossel *et al.* (2001) Filamins as integrators of cell mechanics and signaling. Nat Rev Mol Cell Biol 2:138–145.

145. Stravu *et al.* 2011. *Listeria* monocytogenes transiently alters mitochondrial dynamics during infection. *Proc. Natl. Acad. Sci U.S.A.* **108**(9): 3612.

146. Takahashi and Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:**663.

147. Takenawa *et al.* 2007. The WASP-WAVE protein network connecting the membrane to the cytoskeleton. *Nat. Rev. Mol. Cell. Biol.* **8**(10):37.

148. Tang *et al.* 2010. RNA-Seq analysis to capture the transcriptome landscape of a single cell**.** Nat*. Protoc.* **5:**516.

149. The FANTOM Consortium and the RIKEN PMI and CLST (DGT), 2014. A promoter-level mammalian expression atlas. *Nat.* **507**:462.

150. Tilney and Portnoy. 1989. Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria* monocytogenes. *JCB.* **109**(4): 1597.

151. Trapnell *et al.* 2010. Transcript assembly and quantification by RNA-*Seq* reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**:511.

152. Trapnell *et al.* 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols.* **7**:562.

153. Tripathi *et al.* 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell.* **39**(6): 925.

154. Tsai *et al.* 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science.* **329**:689.

155. Tweten *et al.* 2005. Cholesterol-dependent cytolysisns, a family of versatile pore-forming toxins. *Infect. Immunity.* **73**(10):6199.

156. Vadlamudi *et al.* 2002. Filamin is essential in actin cytoskeletal assembly mediated by p21-activated kinase 1. *Nat. Cell Bio.* **4**:681.

157. Valouev *et al.* 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**:1051–63

158. Vigneau *et al.* 2001. Homology between a 173-kb Region from Mouse Chromosome 10, Telomeric to the Ifnγ Locus, and Human Chromosome 12q15. *Genomics.* **78**(3): 206.

159. Wan *et al.* 2007. Ribosomal protein S3: a KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. *Cell*. **131**:927.

160. Wang and Chang. 2011. Molecular mechanisms of long noncoding RNAs. Mol. Cell **43**:904.

161. **Wang *et al.*** 2005. Mycobacteria inhibition of IFN-gamma induced HLA-DR gene expression by up-regulating histone deacetylation at the promoter region in human THP-1 monocytic cells. *J. Immunol.* 174:5687.

162. **Wang *et al.*** 2009. Regeneration, repair and remembering identity: the three Rs of Hox gene expression. *Trends Cell Biol.* **19**:268.

163. **Wang *et al.*** 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**:57-63.

164. **Wang *et al.*** 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* **26**:136.

165. **Weinmann *et al.*** 1999. Rapid and selective remodeling of a positioned nucleosome during the induction of IL-12 p40 transcription. *Immunity* **11**:665–75.

166. **Weinmann *et al.*** 2001. Nucleosome remodeling at the IL-12 p40b promoter is a TLR-dependent, Rel-independent event. *Nat. Immunol.* **2**:51.

167. **Welch *et al.*** 1998. Interaction of human Arp2/3 complex and the *Listeria* moncytogenes ActA protein in actin filament nucleation. *Science.* **281**(5373):105.

168. **Yamada *et al.*** 2003. Empirical analysis of transcriptional activity in the Arabidopsis genome, *Science.* **302**:842.

169. **Yap *et al.*** 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell.* **38**:662.

170. **Yuan *et al.*** 2014. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell.* **25**(5):666.

171. **Yue *et al.*** 2009. The cytoskeleton protein filamin-A is required for an efficient recombinational DNA double strand break repair. *Cancer Res.* **69**:7978.

172. **Zang *et al.*** 2003. Ribosomal protein L11 negatively regulates oncoprotein MDM2 and mediates p53-dependent ribosomal-stress checkpoint pathway. *Mol. Cell. Biol.* **23**:8902.

173. **Zhong *et al.*** 2002. The phosphorylation status of nuclear NF-κβ determines its association with CBP/p300 or HDAC-1. *Mol. Cell* **9**: 625.

174. **Zhou *et al.*** 2010. Filamins in cell signaling, transcription and organ development. *Trends Cell Biol.* **20**(2):113.

175. **Zurawski *et al.*** 2009. Shigella flexneri type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interations with retinoblastoma protein. *Mol. Microbiology.* **71(2):**350.