ABSTRACT

Title of Dissertation: TRANSCRIPTOMIC PROFILING OF *LEISHMANIA* PARASITES AND HOST MACROPHAGES DURING AN INFECTION

Laura A. L. Dillon, Doctor of Philosophy, 2015

Dissertation Directed by: Dr. Najib M. El-Sayed, Ph.D.
Department of Cell Biology and Molecular Genetics

*Leishmania* parasites cause leishmaniasis, a group of diseases that range in manifestations from skin lesions to fatal visceral disease. The parasite's life cycle is divided between its insect vector and its mammalian host, where it resides primarily inside of macrophages. Once intracellular, *Leishmania* parasites must avoid being killed by the innate and adaptive immune responses. We performed transcriptomic profiling using RNA-seq to simultaneously identify global changes in gene expression in *Leishmania* parasites across multiple lifecycle stages and in infected macrophages from both murine and human hosts. Using a novel approach based on a dual statistical test to identify genes that were differentially expressed relative to both uninfected macrophages and macrophages that had ingested inert particles, we were able to filter out genes that were differentially regulated as part of a general phagocytic response and thereby select genes specific to *Leishmania* infection. The most substantial and dynamic *Leishmania*-specific differential expression responses were observed during early infection, while changes observed later were common to

phagocytosis more generally. An evaluation of RNA processing events within the parasite revealed precise UTR boundaries for a majority of genes and widespread alternative *trans*-splicing and polyadenylation. Collection of data from multiple biological replicates, the use of matched host control samples, careful statistical analysis of variation, and removal of batch effects enabled the detection of biological differences between samples and timepoints with high confidence and sensitivity. Pathway and gene ontology analyses provided insights into the higher level processes activated across parasite developmental stages and during intracellular infection to reveal signatures of *Leishmania* differentiation and infection.

# TRANSCRIPTOMIC PROFILING OF *LEISHMANIA* PARASITES AND HOST MACROPHAGES DURING AN INFECTION

By

Laura A. L. Dillon

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Najib M. El-Sayed, Chair
Professor Hector Corrada Bravo
Professor Sridhar Hannenhalli
Professor David M. Mosser
Professor Michael Cummings, Dean's Representative

## Dedication

I dedicate my degree to my husband, Sean, and our beautiful children, Luke and Juliet, who bring me immeasurable joy and continually remind me that the best path in life is not always the most direct.

## Acknowledgements

I would like to thank my mentor, Najib El-Sayed, for allowing me the opportunity to join his lab and for constant support and guidance on all things, work-related and not. I was nervous about whether I would be able to fit back into the academic setting (especially in the new-to-me field of bioinformatics) or to achieve a satisfying work-life balance; I feel extremely lucky to have found an environment where I could learn from the ground up and I am appreciative-beyond-words for the flexibility you've allowed me to find ways to thrive at both work and home. I have learned so much from working with you and by virtue of the experiences afforded me as a member of your lab – for my own project and multiple exciting collaborations - and I am exceedingly grateful.

I would like to thank my committee members – Drs. Dave Mosser, Hector Corrada Bravo, Sridhar Hannenhalli, and Michael Cummings, for your time, insights, and encouragement. It's been a real pleasure interacting with each of you through our collaborations, in class, and just generally out and about around BRB and CBCB.

I would like to acknowledge the co-authors of my "LMinfectome" project publications - from the El-Sayed lab: Cecilia Fernandes for rigorous experimental and analytical work on the human infection project, Keith Hughitt for developing innovative methods to identify and quantify *L. major* RNA processing sites, Yuan Li for laying the groundwork for the *trans*-splicing and polyA analyses and for robust tools to interpret the RNA processing sites, and Trey Belew for supplying reference datasets for novel ORFs, *L. major* `GOseq`, and ontology mapping and

for helpful data transformation scripts. I would also like to acknowledge Rahul Suresh from the Mosser lab for preparing samples for the procyclic-metacyclic *L. major* and mouse experiments and Kwame Okrah from the Corrada Bravo lab for developing tools to evaluate relationships between samples, to evaluate and account for batch effects, and to test for differential expression. I'm so proud of what we were able to accomplish and pretty thrilled that we were able to have some fun along the way too.

I would also like to thank other members of the El-Sayed lab, past and present, for their discussions of data and methods, as well as for their camaraderie – April Hussey, Ginger Houston-Ludlam, Saloe Bispo, Jung Min Choi, Ramzi Temanni, MariaSanta Mangione, Eva Morgun, Rachel Evans, Rondon Neto, and Wanderson Da Rocha. I especially thank April, Trey, Yuan, and Keith for working so diligently to keep the lab running by handling the logistics related to ordering, server maintenance, dataset updates, etc.

I would like to thank my parents, Julina and Jim Liefer, for their constant support and encouragement throughout my schooling, and for all of their help with the kids, dog, and house when we needed advice or an extra set of hands. I would like to thank my husband, Sean, for being my sounding board for all manners of wild ideas, plans, and frustrations, for encouraging me to return to graduate school even though it meant giving up a degree of financial security and delaying some of our long-term goals. Thank you for being such a capable spouse and father and a true friend and partner. And I would like to thank my little loves – Luke and Juliet – for bring all manners of cheerfulness to everyday

moments and for offering true perspective on what is important and meaningful in life (and at times making the difficulties of graduate school pale in comparison). I love you both more than I ever imagined possible.

# Table of Contents


Dedication ..... ii

Acknowledgements ..... iii

Table of Contents ..... vi

List of Tables ..... ix

List of Figures ..... x

List of Abbreviations ..... xii

Chapter 1 ..... 1
- Introduction ..... 1
  - *Leishmania* lifecycle ..... 1
  - Trypanosomatid gene expression ..... 3
  - Previous studies of leishmaniasis ..... 4
  - Study overview ..... 5
  - Dissertation overview and contributions ..... 6

Chapter 2 ..... 9
- RNA processing events during *Leishmania* differentiation ..... 9
  - Introduction ..... 9
  - Identification of transcript boundaries ..... 12
  - Gene structure features ..... 13
  - Alternative RNA processing events ..... 17
  - Conclusion ..... 24

Chapter 3 ..... 26
- Transcriptomic profiling of *Leishmania* metacyclogenesis ..... 26
  - Introduction ..... 26
  - Experimental design ..... 27
  - Statistical evaluation of biological replicates and batch effects ..... 28
  - Identification of genes differentially expressed between stages ..... 33
  - Examination of gene lists and gene ontology-based enrichment analyses ..... 34
  - Conclusion ..... 39

Chapter 4 ..... 41
- Transcriptomic profiling of *Leishmania*-infected murine macrophages ..... 41
  - Introduction ..... 41
  - Infection dynamics and global transcription patterns ..... 42
  - Statistical assessment of biological replicates and batch effects ..... 45
  - Differential expression and pathway enrichment analyses in murine macrophages ..... 49

Differential expression and gene ontology enrichment analyses in *L. major* ..... 62
Conclusion ..... 69

Chapter 5 ..... 71
Transcriptomic profiling of *Leishmania*-infected human macrophages ..... 71
Introduction ..... 71
Study design ..... 73
Global transcriptome profiles of *Leishmania* and its human macrophage host cell ..... 76
Discriminating between parasite- and phagocytosis-driven changes in human macrophages ..... 79
*Leishmania*-induced remodeling of gene expression in human macrophages ..... 82
Towards defining the signature of mammalian macrophage infection by *Leishmania* ..... 85
Identification of differentially expressed genes in *Leishmania* parasites ..... 90
Towards defining the signatures of *Leishmania* differentiation and intracellular survival ..... 95
Conclusion ..... 101

Chapter 6 ..... 103
Future Directions ..... 103
Extend *L. major* RNA processing/gene regulation analysis ..... 103
Deepen analysis of gene expression from 0-24 hpi ..... 104
Expand infection systems and conditions studied ..... 105
Detect parasite-host protein-protein interactions to derive biological mechanisms ..... 107
Establish new host-pathogen interaction resources ..... 108

Chapter 7 ..... 110
Experimental Procedures ..... 110
*Leishmania* Culture ..... 110
Mouse infection ..... 111
Human macrophages and infection ..... 111
RNA isolation and cDNA library preparation ..... 112
RNA-seq data generation, pre-processing, and quality trimming ..... 113
Mapping cDNA fragments to the reference genome, abundance estimation, and data normalization ..... 113
Data quality assessment by statistical sample clustering and visualization ..... 114
Differential expression analysis ..... 115
Ortholog mapping ..... 116
Gene ontology (GO) analysis ..... 116
KEGG pathway analysis ..... 116
*Trans*-splicing site detection and 5´ UTR analysis ..... 117
Polypyrimidine tract characterization ..... 118
Polyadenylation site detection and 3´ UTR analysis ..... 118

Alternative RNA processing site analysis ..... 119
Data access ..... 120
Appendices ..... 121
Appendix 1: Procyclic-metacyclic experimental design ..... 121
Appendix 2: Coordinates of novel ORFs ..... 122
Appendix 3: Mouse infection experimental design ..... 136
Appendix 4: Human infection experimental design ..... 137
References ..... 140

# List of Tables

Table 1: Dinucleotide acceptor site usage frequency. ..... 19
Table 2: Top differentially expressed genes in the *L. major* procyclic to metacyclic promastigote transition. ..... 35
Table 3: Gene ontology (GO) categories enriched during the *L. major* procyclic to metacyclic promastigote transition. ..... 36
Table 4: KEGG pathways enriched in *L. major*-infected murine macrophages at 4 hpi. ..... 56
Table 5: KEGG pathways enriched among genes upregulated in *L. major*-infected murine macrophages at 24, 48, and 72 hpi. ..... 59
Table 6: KEGG pathways enriched among genes downregulated in *L. major*-infected murine macrophages at 24, 48, and 72 hpi. ..... 60
Table 7: Gene ontology categories enriched across *L. major* stages/timepoints. ..... 65
Table 8: KEGG pathways enriched in genes common to human and murine macrophages infected with *L. major* at 4 hpi. ..... 87
Table 9: GO categories enriched among *L. major* genes in common to human and mouse infections. ..... 97
Table 10: GO categories enriched among *L. major* genes unique to human and mouse infections. ..... 100
Table 11: Procyclic-metacyclic experimental design. ..... 121
Table 12: Coordinates of Novel ORFs. ..... 122
Table 13: Mouse infection experimental design. ..... 136
Table 14: Human infection experimental design. ..... 137

# List of Figures

Figure 1: The lifecycle of *Leishmania* parasites. ..... 2
Figure 2: *Trans*-splicing and polyadenylation in Trypanosomatids ..... 10
Figure 3: Length and position distribution of gene structure components. ..... 15
Figure 4: UTR length distribution by developmental stage. ..... 16
Figure 5: Characterization of alternative RNA processing. ..... 19
Figure 6: Preferential usage of primary site across developmental stages. ..... 22
Figure 7: Visualization of changes in primary *trans*-splicing sites across developmental stages. ..... 23
Figure 8: Sample characterization. ..... 28
Figure 9: Sample diagnostics to globally assess data similarities and identify outliers. ..... 30
Figure 10: Global gene expression profiles of the procyclic and metacyclic promastigote forms of *Leishmania major*. ..... 32
Figure 11: MA plot of differentially expressed genes in the *L. major* procyclic to metacyclic promastigote transition. ..... 34
Figure 12: Characterization of L. major intracellular growth and proportion of RNA from the parasite. ..... 44
Figure 13: Global assessment of data distribution. ..... 45
Figure 14: Principal Component Analysis (PCA) and hierarchical clustering analysis prior to accounting for batch effects. ..... 47
Figure 15: Principal Component Analysis (PCA) and hierarchical clustering analysis after accounting for batch effects. ..... 48
Figure 16: MA plot of differentially expressed genes in murine macrophages upon infection by *L. major*. ..... 51
Figure 17: Differentially expressed genes in L. major parasites and their murine macrophage host cells. ..... 52
Figure 18: Expression patterns for the top up- and down-regulated genes in *L. major*-infected murine macrophages. ..... 53
Figure 19: MA plot of differentially expressed genes in murine macrophages over time upon infection by *L. major*. ..... 54
Figure 20: MA plot of differentially expressed genes in *L. major* across developmental stages and intracellular timepoints. ..... 63
Figure 21: Study design, donors, and timepoints. ..... 74
Figure 22: Dynamics and characterization of CSF1-induced human macrophage infection with *Leishmania*. ..... 75
Figure 23: Global expression profile of human macrophages and *Leishmania* parasites during infection. ..... 78
Figure 24: Differentially expressed genes in macrophages infected with *Leishmania* species. ..... 80
Figure 25: Response of human and murine macrophages to *L. major* infection at 4 hpi after accounting for the phagocytosis effect. ..... 86

Figure 26: Differentially expressed genes between *Leishmania* parasite developmental stages. ..... 91
Figure 27: Comparison of *Leishmania* transcriptomes between species..... 93
Figure 28: Comparison of the *L. major* transcriptome upon infecting human and murine macrophages. ..... 96

# List of Abbreviations

| | |
|---|---|
| A | adenine |
| AAH | aminohydrolase |
| AAT | aspartate aminotransferase |
| Akt | protein kinase B |
| ALD | aldolase |
| Amast | Amastigote |
| ATG8 | autophagy-related protein 8 |
| ATP | adenosine triphosphate |
| C | cytosine |
| $Ca^{2+}$ | calcium ion |
| CARD9 | caspase recruitment domain-containing protein 9 |
| CCL20 | chemokine (C-C motif) ligand 20 |
| CD14 | cluster of differentiation 14 |
| CDS | coding sequence |
| CMIP | c-Maf inducing protein |
| CR3 | complement receptor 3 |
| CSF1 | colony stimulating factor 1 |
| CSF2 | colony stimulating factor 2 |
| CSF3 | colony stimulating factor 3 |
| CXCL1 | chemokine (C-X-C motif) ligand 1 |
| CXCL2 | chemokine (C-X-C motif) ligand 2 |
| CXCL3 | chemokine (C-X-C motif) ligand 3 |
| CXCL5 | chemokine (C-X-C motif) ligand 5 |
| CYP51 | lanosterol 14 alpha-demethylase |
| DE | differentially expressed/differential expression |
| DOT1 | histone-lysine N-methyltransferase |
| EGFR | epidermal growth factor receptor |
| ENOL | enolase |
| ErbB | erythroblastic leukemia viral oncogene |
| EuPathDB | Eukaryotic Pathogen Database |
| FBS | fetal bovine serum |
| FC | fold change |
| FIGF | C-fos-induced growth factor |
| FoxO | forkhead box |
| FPPS | farnesylpyrophosphate synthase |
| G | guanine |
| GABA | gamma-aminobutyric acid |
| GABRE | gamma-aminobutyric acid receptor subunit epsilon |
| GAPDH | glyceraldehyde 3-phosphate dehydrogenase |
| GO | gene ontology |
| GPI | glycosylphosphatidylinositol |
| GSS | genomic survey sequence |
| HASP | hydrophilic acylated surface protein |
| HIF-1 | hypoxia-indicuble factor 1 |

| | |
|---|---|
| HMGR | HMG-CoA reductase |
| HMOX1 | heme oxygenase 1 |
| HPGL | host-pathogen genomics laboratory |
| hpi | hours post infection |
| hr | hour |
| HSP | heat shock protein |
| HTLV | human T-lymphotrophic virus |
| IL1B | interleukin 1 beta |
| IL1RAP | interleukin 1 receptor accessory protein |
| IL1RN | interleukin 1 receptor antagonist |
| IL6 | interleukin 6 |
| IL10 | interleukin 10 |
| IL11R | interleukin 11 receptor |
| IL18 | interleukin 18 |
| IL18R1 | interleukin 18 receptor 1 |
| Inf | infected |
| infLA | *Leishmania amazonensis*-infected |
| infLM | *Leishmania major*-infected |
| ino1 | inositol-3-phosphate synthase |
| JAK | janus kinase |
| logFC | log(2) fold change |
| LPG | lipophosphoglycan |
| MAP | mitogen-activated protein |
| MAPK | mitogen-activated protein kinases |
| Metac | Metacyclic promastigote |
| mit1 | *myo*-inositol |
| mpc | median pairwise correlation |
| NF-kappa B | nuclear factor kappa-light-chain-enhancer of activated B cells |
| NLRP | NOD-like receptor pyrin domain |
| NOD | nucleotide-binding oligomerization domain |
| NOS2 | nitric oxide synthase, inducible |
| NS | not significant |
| NSDHL | sterol-4-alpha-carboxylate 3-dehydrogenase |
| nt | nucleotide |
| ORF | open reading frame |
| PI3K | phosphatidylinositol-4,5-bisphosphate 3-kinase |
| polyA | polyadenine/polyadenylation |
| polyPy | polypyrimidine |
| P | primary |
| PAMP | pathogen-associated molecular pattern |
| PC | principal component |
| PCA | principal component analysis |
| PCR | polymerase chain reaction |
| PNA | peanut agglutinin |
| Proc | Procyclic promastigote |
| PSG | promastigote secretory gel |

| | |
|---|---|
| PTGS2 | prostaglandin-endoperoxide synthase 2 |
| RIG | retinoic acid-inducible gene |
| S | secondary |
| SCG | phosphoglycan beta 1,3 galactosyltransferase |
| SHERP | small hydrophilic endoplasmic reticulum protein |
| SL | spliced leader |
| SNARE | soluble NSF attachment protein receptor |
| SOCS3 | suppressor of cytokine signaling 3 |
| SOD2 | superoxide dismutase 2 |
| SRA | Short Read Archive |
| STAT | signal transducer and activator of transcription |
| SYT | synaptotagmin |
| T | thymine |
| TGF | transforming growth factor |
| TLR | toll-like receptor |
| TM-HMM | transmembrane hidden Markov model |
| TNF | tumor necrosis factor |
| TNFSF9 | TNF superfamily member 9 |
| U | uracil |
| Uninf | uninfected |
| UTR | untranslated region |
| VAMP | vesicle-associated membrane protein |
| VEGF | vascular endothelial growth factor |
| XRPT | xanthine phosphoribosyltransferase |
| Y2H | yeast-2-hybrid |

# Chapter 1

## Introduction

### *Leishmania* lifecycle

*Leishmania major* and related species are the causative agents of leishmaniasis, a group of diseases that vary in severity from self-healing skin lesions to disfiguring mucosal manifestations to fatal visceral disease. More than a million new cases are reported annually, mostly concentrated in the Middle East and Central and South America (Alvar, 2012). The *Leishmania* life cycle is divided between its insect vector, the phlebotomine sand fly, and its mammalian host, where it resides primarily inside of macrophages, although neutrophils, dendritic cells, and fibroblasts have also been implicated at various stages of infection (Bogdan, 2000; Laufs, 2002; Moll, 1995; Peters, 2008; Sarkar, 2013) (see Figure 1). The developmental stages of *Leishmania* parasites can be distinguished by cell size and shape, position of the nucleus, proliferation state, and the presence and length of an external flagellum. Upon being bitten by a sand fly, non-proliferating, flagellated metacyclic promastigotes are introduced into the bloodstream of the mammalian host and are quickly phagocytosed by host immune cells. They come to reside inside of macrophage lysosome-derived parasitophorous vacuoles where they undergo differentiation into the small, aflagellated amastigote stage. Amastigotes divide rapidly inside of the phagolysosomes of infected macrophages and are then released from the cells to infect other host macrophages. When an infected host is bitten by a sand fly,

amastigotes are transmitted into the abdominal midgut of the sand fly where they differentiate into proliferative procylic promastigotes and then further into large, slender (nectomonad) promastigotes that are cell cycle-arrested. The nectomonad promastigotes migrate to the thoracic midgut of the sand fly where they change into proliferative (leptomonad) promastigotes, which secrete a promastigote secretory gel (PSG) that causes the sand fly to regurgitate them during its subsequent blood meal. The leptomonad promastigotes differentiate into the metacyclic promastigotes that are transmitted back to the mammalian host, thus perpetuating the cycle (Bates, 2007; Kramer, 2012).

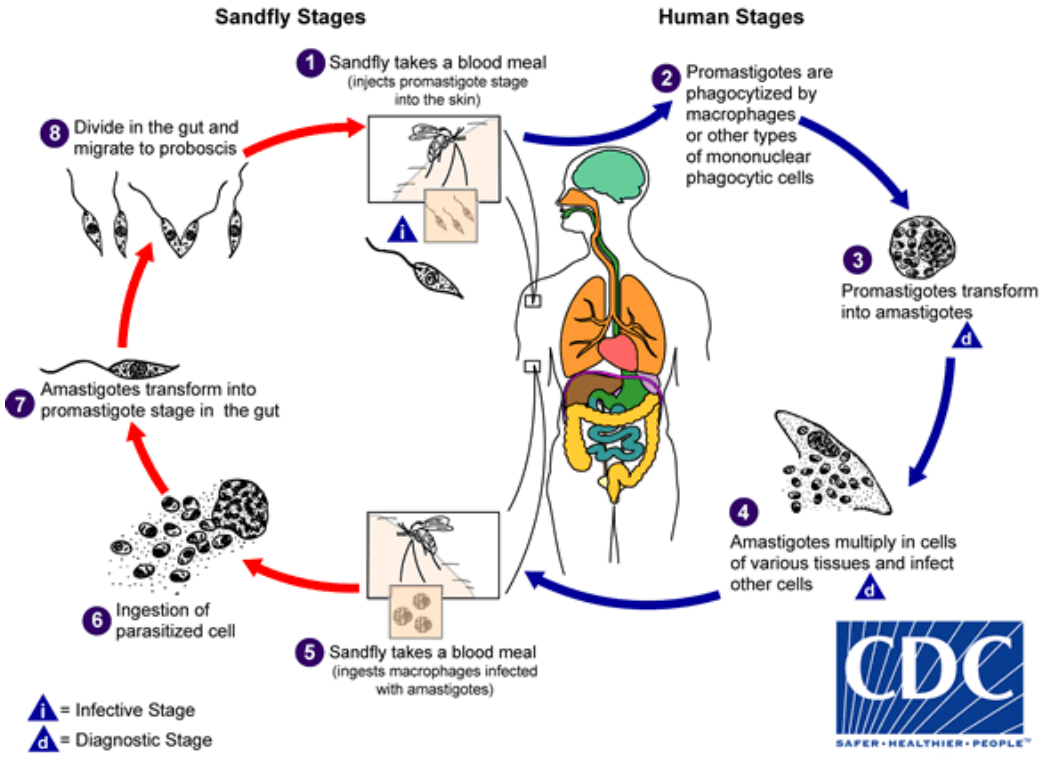


**Figure 1: The lifecycle of *Leishmania* parasites.**

The lifecycle of *Leishmania* parasites is divided between its insect vector, the female phlebotomine sand fly (left) and its mammalian host (right), where it resides inside of macrophages [from http://www.cdc.gov/parasites/leishmaniasis/biology.html].

When responding to changes in the environment as it moves through its lifecycle, such as upon leaving the sand fly vector and infecting host cells, the parasite must alter its gene expression to adapt to the new surroundings. In particular, the parasite must respond to changes in temperature, pH, and osmolarity and withstand assault from the host immune system. Many of the changes that the parasite undergoes are reflected in changes in its morphology (size, shape, position of organelles) and variations in cell surface components (Ambit, 2011; Beverley, 1998; Wheeler, 2011). Less is known about changes that take place at the transcriptional level.

**Trypanosomatid gene expression**

Unlike most other eukaryotes, *Leishmania* and other trypanosomatids, including *Trypanosoma brucei* and *Trypanosoma cruzi*, do not regulate the expression levels of individual genes by the differential recruitment of RNA polymerase II influenced by cellular transcription factors. Rather, their genes are arranged as polycistronic clusters of tens to hundreds of functionally unrelated genes that are transcribed at roughly the same rate across the genome. The long, polycistronic pre-mRNAs are then split into component mature mRNAs through the processes of *trans*-splicing and polyadenylation (Sutton, 1986). While the genome sequence of *L. major* was completed in 2005 (Ivens, 2005), much remains unknown about gene boundaries, the mechanisms directing the expression levels of individual genes, or how *Leishmania* genes influence or are influenced by host gene expression during an infection.

**Previous studies of leishmaniasis**

Studies using microarrays or SAGE tags have started to elucidate changes that occur within the parasite or within the host as infection occurs (Akopyants, 2004; Almeida, 2004; Buates, 2001; Chaussabel, 2003; Cohen-Freue, 2007; Depledge, 2009; Ettinger, 2008; Giraud, 2012; Gregory, 2008; Guerfali, 2008; Holzer, 2006; Lahav, 2011; Leifso, 2007; McNicoll, 2006; Novais, 2015; Osorio y Fortéa, 2009; Probst, 2012; Rabhi, 2012; Ramírez, 2012; Rochette, 2008; Rochette, 2009; Rodriguez, 2004; Saxena, 2003; Saxena, 2007), but have so far not sought to look at the transcriptomes of both simultaneously and over the course of an intracellular infection. While very informative, microarray-based approaches have several inherent limitations such as hybridization and cross-hybridization artifacts, the restriction on genes interrogated to probes included on the array (inhibiting the identification of previously unannotated genes), dye-based detection issues, the need for large amounts of input RNA, and the inability to detect 5´ and 3´ UTRs boundaries.

Previous studies have identified genes that are differentially regulated upon infection with various *Leishmania* species, sometimes with opposing results. Some of these differences may be attributable to the parasite species and host systems used, the severity of the resulting infection, and the timepoints/ developmental stages examined. Additionally, studies of *Leishmania* amastigotes have often used axenic cultures (Holzer, 2006; Lahav, 2011; McNicoll, 2006; Rochette, 2009; Saxena, 2007) or lesion-derived amastigotes (Akopyants, 2004; Almeida, 2004; Holzer, 2006; Leifso, 2007; Maretti-Mira, 2012; Rochette, 2008).

The former have been shown to significantly differ from the intracellular biological state (Holzer, 2006; Rochette, 2009) while the latter contain a mixture of amastigotes at various timepoints post-infection, making it difficult to differentiate between changes that take place during early versus later intracellular infection. These limitations likely resulted in the identification of an incomplete list of genes that are up- or down-regulated in the parasite and host prior to and during an infection. RNA-seq, which enables a precise and sensitive measurement of mRNA transcript abundance, has begun to be applied to this problem (Rastrojo, 2013), and additional, comprehensive, well-replicated studies examining gene expression across multiple conditions and in multiple host systems are needed to more fully understand both the gene expression signatures of the parasite and host during an infection and how these change over time.

**Study overview**

Transcriptomic profiling using RNA-seq was performed to interrogate gene structure and expression signatures within *Leishmania* developmental stages and to concurrently identify transcriptomic changes in the parasite's mammalian hosts. Specifically, we were able to precisely delineate the 5´ and 3´ UTR boundaries for a majority of *Leishmania* genes and to detect widespread alternative *trans*-splicing and polyadenylation. We were able to identify global changes in gene expression that occur as *Leishmania major* undergoes metacyclogenesis from the non-infective procyclic promastigote form to the infective metacyclic promastigote form, and to simultaneously identify global changes in host and parasite gene expression over the first 72 hours of a

synchronized intracellular infection of murine peritoneal macrophages or primary human macrophages by either *Leishmania major* or *Leishmania amazonensis*. Subsequent pathway and gene ontology analyses provided insights into the higher level processes activated as the parasite achieves infectivity and during an infection.

A paired-end mRNA sequencing approach was used to allow high confidence read mapping and transcript assembly. Careful study design which included the collection of data from multiple biological replicates, consistency between biological systems evaluated, the use of matched control samples, robust statistical analysis of replicates, and consideration of technical sources of variation (batch effects) enabled the detection of biological differences between samples and timepoints with high confidence and sensitivity. This work builds on and improves existing expression datasets and gene structure annotations and allows insights into how *Leishmania* is able to evade host defenses and cause modulations in the host transcriptome in order to survive in the mammalian intracellular environment. The datasets generated provide a substantially more detailed interpretation of *L. major* biology that will inform the field and potentially provide additional data for drug discovery and vaccine development efforts.

**Dissertation overview and contributions**

This dissertation describes analyses done to identify RNA processing sites in *Leishmania major*, to characterize changes in gene expression throughout the *Leishmania* lifecycle, and to describe changes in gene expression in *Leishmania*-infected mammalian macrophages. Chapters 2-3 describe analyses of the

procyclic and metacyclic promastigote stages of *L. major* to identify and characterize RNA processing sites (Chapter 2) and to quantify differential expression (Chapter 3). Chapter 4 describes differential expression analyses for *L. major*-infected murine macrophages. Chapter 5 describes similar analyses in human macrophages and additionally describes *L. amazonensis* infection of macrophages and changes that occur when macrophages ingest latex beads. Chapter 6 describes proposed future directions for this research.

Cell samples for the promastigote analyses and murine infections (Chapters 2-4) were prepared by Rahul Suresh. For these samples, total RNA was isolated and cDNA libraries were prepared by Laura Dillon. For the human experiments (Chapter 5), all samples were generated by Cecilia Fernandes (cell culture, RNA isolation, and cDNA library preparation). The pipeline used to identify *trans*-splicing and polyadenylation sites (Chapter 2) was developed and implemented by Keith Hughitt. The scripts used to characterize alternative RNA processing were originally conceived and written by Yuan Li with additional adaptations by Laura Dillon for the specific purposes of the *Leishmania* analysis. The novel ORF dataset (Chapters 2-3), *Leishmania* GO mapping reference dataset (Chapters 3-5), and human-mouse/*L. major*-*L. amazonensis* orthology tables (Chapter 5) were provided by Trey Belew. Imaging and quantification of the metacyclic stage enrichment was done by Cecilia Fernandes (Chapter 3). The pipelines for sample diagnostics, batch effect detection, and differential expression analysis (Chapters 3-4) were developed in collaboration with Kwame Okrah and Hector Corrada Bravo, and were run by Laura Dillon. The dual

statistical test used to test for differential expression after accounting for the phagocytosis effect was developed by Hector Corrada Bravo and implemented by Laura Dillon. `GOseq` analyses (Chapters 3-5) and KEGG analyses (Chapters 4-5) were done by Laura Dillon. The results of the analyses, including the differentially expressed gene lists and GO and KEGG enrichments, were interpreted by Laura Dillon (Chapters 2-5), Cecilia Fernandes (Chapter 5), Najib El-Sayed (Chapters 2-5), and David Mosser (Chapters 3-5).

The analyses described in Chapters 2-3 were published in *Nucleic Acids Research* in July 2015 (Dillon et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania* major differentiation. 2015. PMID: 26150419). A manuscript describing the analyses in Chapter 4 is under revision at *BMC Genomics* (October 2015). A manuscript describing the analyses in Chapter 5 has been submitted (October 2015).

# Chapter 2

## RNA processing events during *Leishmania* differentiation

### Introduction

The genomes of *Leishmania* and other kinetoplastid organisms – including *T. cruzi* and *T. brucei* – are composed of polycistronic clusters of functionally-unrelated genes that are transcribed at roughly the same rate across the genome. Transcription initiation occurs at divergent 'strand switch regions' where polycistronic units originate in opposite directions on opposing DNA strands (El-Sayed, 2005; Ivens, 2005; Martinez-Calvillo, 2003; McDonagh, 2000; Myler, 1999). The *trans*-splicing of a capped 39-nucleotide (nt) spliced leader (SL) mini-exon sequence to the 5´ end of each nuclear mRNA and the polyadenylation of the 3´ end are used to separate each polycistronic pre-mRNA transcript into its component mature mRNAs (Sutton, 1986) (see Figure 2). *Trans*-splicing and polyadenylation events are coupled temporally and spatially such that the SL acceptor site of the downstream gene determines the location of the polyadenylation site of the upstream gene and both modification events occur simultaneously during post-transcriptional processing (LeBowitz, 1993; Matthews, 1994; Ullu, 1993). Spliced leader acceptor sites contain a consensus AG dinucleotide that is preceded by polypyrimidine-rich sequence and a G nucleotide excluded from the -3 position (Kolev, 2010; LeBowitz, 1993; Matthews, 1994). Polyadenylation sites do not appear to contain a specific signal sequence and have been reported to occur about 500-600 nucleotides upstream

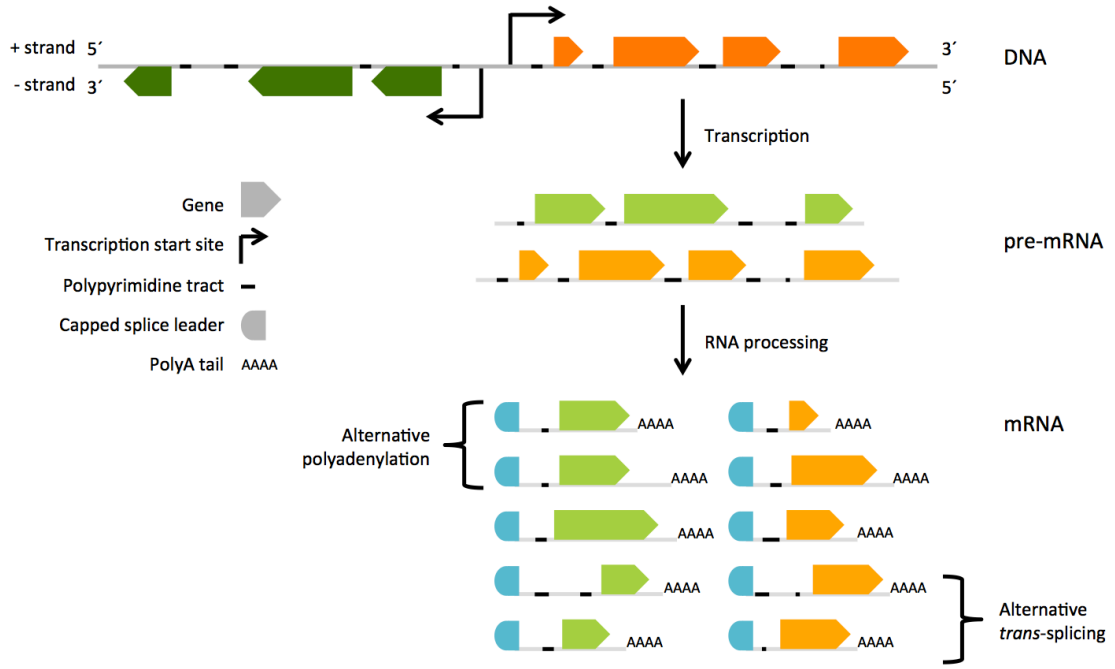


**Figure 2: *Trans*-splicing and polyadenylation in Trypanosomatids**

*Leishmania* and related species utilize polycistronic transcription followed by the *trans*-splicing of a capped 39-nt spliced leader sequence to the 5′ end of each component RNA coupled with polyadenylation of the 3′ end of the upstream gene to separate pre-mRNAs into their component mature mRNAs. The illustration shows the transcription and subsequent RNA processing of two polycistronic units (green and orange) by the addition of the capped spliced leader (blue) and polyA tail (AAAA). A stretch of pyrimidine residues, known as the polypyrimidine tract (black) is involved in the regulation of RNA processing. The use of different nucleotide positions for the *trans*-splicing and polyadenylation events leads to alternative RNA processing.

of the coupled *trans*-splicing acceptor site (LeBowitz, 1993). *Trans*-splicing of a specific SL sequence is not itself unique to trypanosomatids, but has evolved in parallel in a range of diverse organisms including Euglenozoa and dinoflagellates, *C. elegans* and related nematodes, Platyhelminthes, and primitive chordates (Derelle, 2010). However, trypanosomatids are distinct in their reliance on *trans*-splicing to express all genes transcribed by RNA polymerase II and because they do so without the use of any sequence-specific transcription factors.

Since the transcription of *Leishmania* genes occurs fairly uniformly across the genome, steady-state mRNA levels for individual genes are largely dependent on gene copy number and the rate of mRNA degradation, with mRNA deadenylation preceding degradation for most mRNAs. Sequence motifs contained in the 3´ UTRs greatly influence mRNA stability and the recruitment of the cellular degradation machinery (Brittingham, 2001; Charest, 1996; Fadda, 2014; Folgueira, 2005; Manful, 2011; Michaeli, 2011; Peacock, 2007; Rogers, 2011; Zilka, 2001; Coughlin, 2000; Quijada, 2000). Since kinetoplastids lack introns (with very few exceptions), they do not control gene expression by alternative *cis*-splicing (Ivens, 2005; Mair, 2000). Gene expression is thus predominantly controlled, not at the transcriptional level through the developmental regulation of RNA polymerase II activity, but by gene copy number, post-transcriptional mRNA processing, rates of mRNA degradation, and translational efficiency (see (Clayton, 2007) for review).

The genome of *Leishmania major*, which defined the boundaries of coding sequence for the large majority of *L. major* genes, was completed in 2005 (Ivens, 2005). While it has been an invaluable resource for researchers in the field, the lack of defined UTR boundaries has hampered the ability to look for sequence motifs contained in the UTRs that may be involved in directing the post-transcriptional regulation of individual mRNAs - including degradation, storage, and translation rate. Deep sequencing of *L. major* procyclic and metacyclic promastigote samples by RNA-seq presented an opportunity to comprehensively annotate transcript boundaries, thereby enhancing the structural annotation of *L.*

*major* genes and providing a substantial additional resource for the *Leishmania* research community. We exploited the signal sequences generated by *trans*-splicing and polyadenylation events to accurately map the 5′ and 3′ UTR boundaries of transcripts by comparing reads containing these signals to the reference genome sequence. The UTR boundaries identified were used to identify alternative RNA processing events within and between *L. major* developmental stages.

**Identification of transcript boundaries**

Distinct transcript boundaries were determined for a large majority of previously annotated protein-coding genes and novel genes for which there was evidence of translation by ribosome profiling (Table 12). RNA was isolated from cultured *L. major* grown to log phase (procyclic form) or enriched for metacyclic forms using: 1) a Ficoll gradient or 2) negative selection using peanut agglutinin (PNA). PolyA enriched cDNA libraries were generated using the Illumina TruSeq protocol and 100-bp paired end sequences were generated. RNA-seq reads which did not map to the *L. major* genome due to RNA processing events were examined separately for evidence of spliced leader (SL) sequence and a polyA tail. Of the ~960 million reads from 13 *L. major* samples (5 procyclic promastigote samples and 8 metacyclic promastigote samples), ~3.9% contained evidence of *trans*-splicing and ~0.05% contained evidence of polyadenylation (Table 11). Once the SL and polyA sequences were removed, the remainders of the reads were mapped to the genome, allowing the identification of coordinates for at least

one *trans*-splicing site for 8,981 genes (94.2% of a total of 9,530 genes) and at least one polyadenylation site for 8,841 genes (92.8%).

A sampling of the *trans*-splicing and polyadenylation sites identified here was compared to existing data in TriTrypDB (Peter Myler's group, Seattle Biomed) that were generated using an RNA-seq method that specifically enriched for SL-containing sequences (biological sample type unknown). Our *trans*-splicing site data were highly concordant with these previously reported data. This high degree of agreement is remarkable given the differences in sample type, culture, and preparation across different labs and may potentially indicate that the usage of *trans*-splicing sites in *Leishmania* is fairly consistent across various biological conditions. The observed variability is likely attributable to the differences in coverage, RNA-seq approach, and data analysis methodology. Our polyA site data did not generally match the existing data on TriTrypDB down to the specific nucleotide. This could be due to the extreme heterogeneity of these sites (previously reported for *T. brucei* (Kolev, 2010)), differences in the biological samples studied, or differences in the methods used to identify and assign sites.

**Gene structure features**

We sought to determine the length distribution of the elements of each gene - 5´ UTR, coding sequence (CDS), and 3´ UTR - as well as the intergenic region, including the polypyrimidine (polyPy) tract, for previously annotated protein-coding genes and the novel ORFs. Start and stop coordinates for *L. major* genes were used to determine a median CDS length of 1,241 nt with a

range from 64 to 52,178 nt (Figure 3A). The boundaries of 5′ UTRs were defined using the coordinates of the SL addition sites and start codon annotations and a similar analysis was done to determine the lengths of 3´ UTRs using stop codon and polyadenylation site coordinates. The median length of all identified 5′ UTRs (not including the 39 nt SL sequence) and 3´ UTRs was 547 and 729 nt, respectively (Figure 3B and Figure 3C). When only the most-utilized (primary) *trans*-splicing or polyadenylation site for each gene was considered, these values were reduced to 233 and 517 nt, respectively (see alternative RNA processing section below). The distribution of both the 5′ and 3´ UTR lengths was similar in both stages (Figure 4) and there did not appear to be a correlation between CDS length and either UTR length or between corresponding UTR lengths.

The length distribution analysis was extended to examine the polyPy tract, which is known to be involved in the regulation of RNA processing events in trypanosomatids (Günzl, 2010; Huang, 1991; Siegel, 2005). In this analysis, the polyPy tract was identified as the longest stretch of pyrimidine residues located upstream of each of the (primary) *trans*-splicing sites and interrupted by no more than one purine. PolyPy tracts ranged from 7 to 123 nt in length, with a median value of 21 nt (Figure 3D) and a clear usage preference for cytosine (54%) over thymine (42%) residues. This observation, which runs counter to what has been found in related species (Greif, 2013; Kolev, 2010) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision), is unsurprising given the higher GC content of *Leishmania* relative to the other trypanosomatids (El-Sayed, 2005). The median distance between each polyPy tract and its

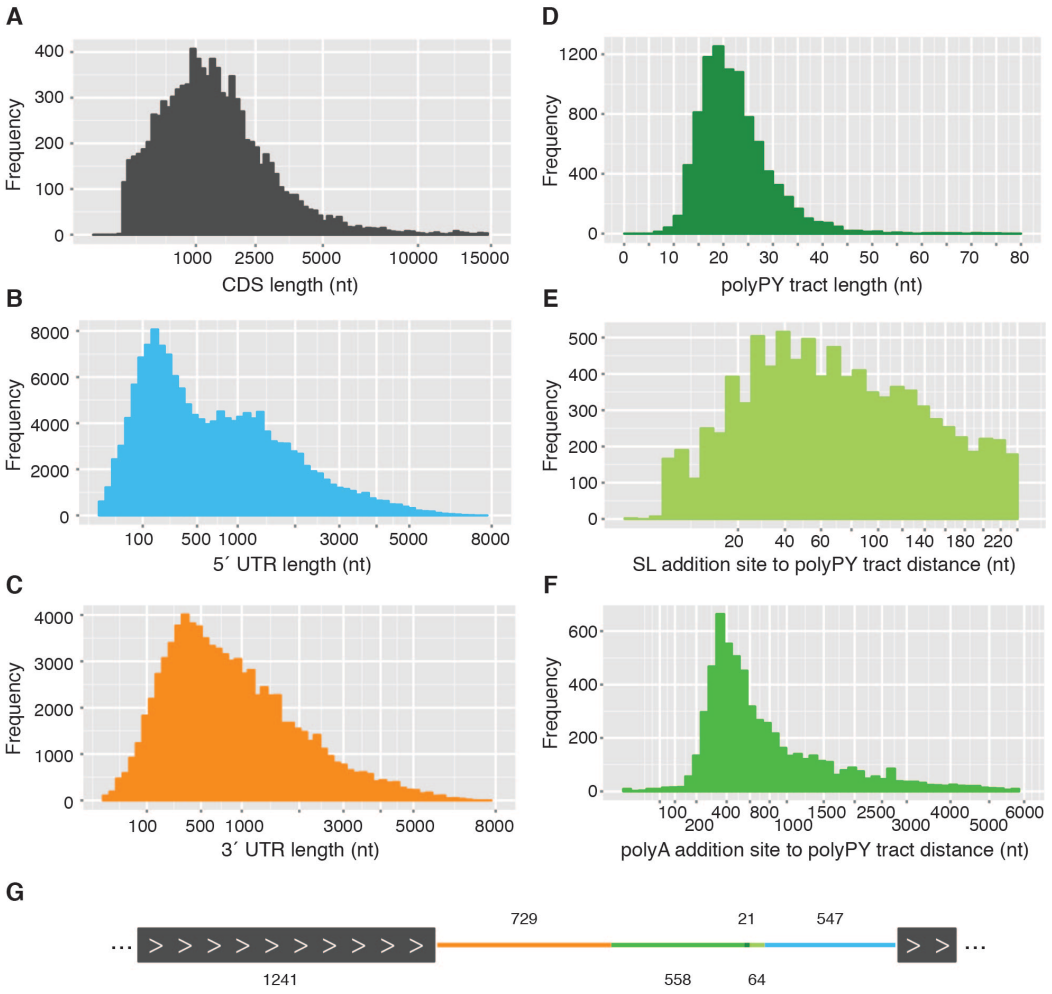


**Figure 3: Length and position distribution of gene structure components.**

Data from procyclic and metacyclic promastigote samples were combined to describe the gene structure elements of *L. major*. (A) Distribution of CDS lengths. Start and stop coordinates for coding sequences for previously annotated protein-coding genes (TriTrypDB v. 6.0) and novel ORFs were used to compute CDS lengths. For genes with multiple isoforms, the first isoform listed in TriTrypDB was included in the analysis. (B) Distribution of 5′ UTR lengths. The exact *trans*-splicing sites associated with each CDS were used to determine the coordinates and lengths of 5′ UTRs. A total of 154,046 *trans*-splicing sites were identified, corresponding to 5′ UTRs ranging from 0 to 7,252 nt in length. (C) Distribution of 3′ UTR lengths. An analysis of polyadenylation sites associated with each CDS was performed to determine 3′ UTR coordinates and lengths. A total of 84,331 polyadenylation sites were identified, corresponding to 3′ UTRs ranging from 0 to 7,133 bases in length. (D) Distribution of polypyrimidine (polyPy) tract lengths. A window of 250 nt upstream of the each primary *trans*-splicing site (8,981 total) was scanned to identify its corresponding polyPy tract, defined as the longest stretch of pyrimidine residues interrupted by no more than one purine. (E) Distribution of distances between each primary *trans*-splicing site and its corresponding polypyrimidine (polyPy) tract. (F) Distribution of distances between each polyPy tract and the polyadenylation site of the upstream gene. A total of 6,174 instances of a neighboring polyadenylation site and polyPy tract were identified. (G) Diagram of a "typical" *L. major* genic region. The median lengths of the gene structure components of *L. major* were used to construct the structure of a "typical" gene region (2 genes and the intergenic region). The median values of each component are shown. Colors correspond to the features depicted in panels A-F.

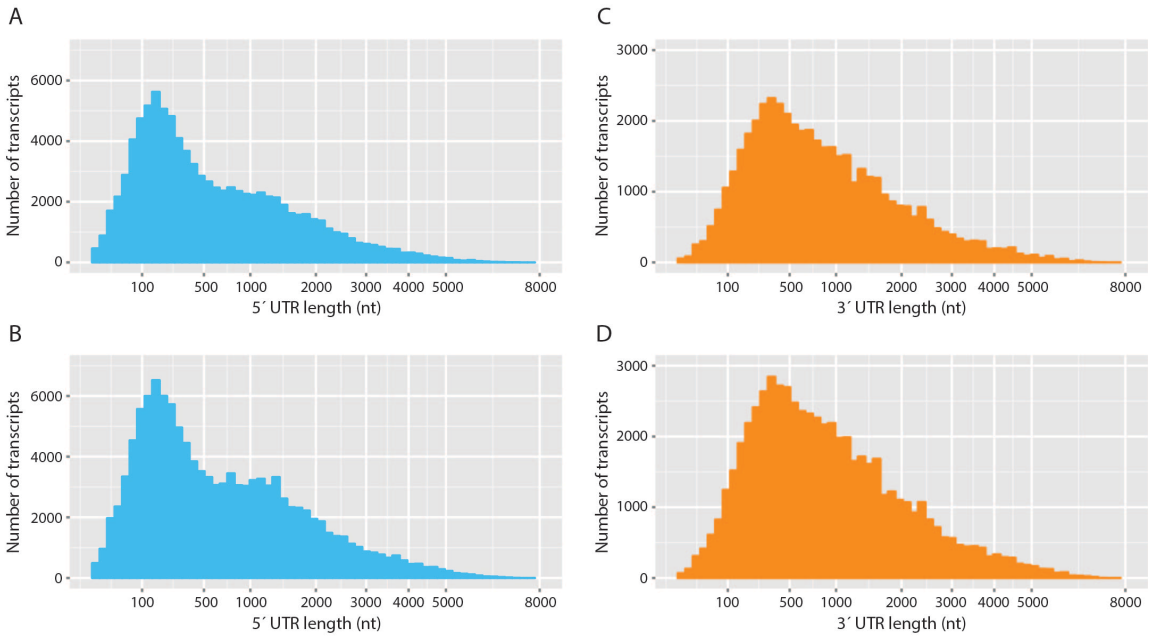


**Figure 4: UTR length distribution by developmental stage.**

*Trans*-splicing sites were identified for each developmental stage and used to determine the coordinates and lengths of 5´ UTRs for procyclic promastigotes (A) and metacyclic promastigotes (B). An analysis of polyadenylation sites in each developmental stage was performed to determine 3´ UTR coordinates and lengths for procyclic promastigotes (C) and metacyclic promastigotes (D).

downstream SL addition site was 64 nt (Figure 3E) and the median distance between the polyPy tract and the upstream polyadenylation site (if both were detected; 6,174 instances) was 558 nt (Figure 3F). When considering only intergenic regions that were bound by both a detectable upstream polyadenylation site and a detectable downstream SL addition site (6,152 instances), a median intergenic distance of 556 nt was observed.

The median values of each gene structure element were used to determine a representative gene structure for *L. major* genes, with a median mRNA length of 2,517 nt, of which the 5′ UTR, CDS, and 3′ UTR account for 22%, 49%, and 29%, respectively (corresponding to a 5′ UTR of 547 nt, a CDS of

1,241 nt, and a 3′ UTR of 729 nt) (Figure 3G). The median intergenic length was 643 nt. This observed gene structure indicates significantly longer 5´ and 3´ UTRs and longer intergenic distances than what has been reported in either *T. cruzi* and *T. brucei* (Kolev, 2010; Siegel, 2010) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision), and is consistent with previous observations regarding the relative compactness of the species' genomes (El-Sayed, 2005).

**Alternative RNA processing events**

The sequencing depth of our *L. major* transcriptomic profiling experiments allowed not only the identification of the SL-addition and polyadenylation sites at a single-base resolution, but also the quantification of alternative RNA processing events. Of the 8,981 genes with SL-addition sites detected, 8,777 (~98%) used more than one *trans*-splicing site in at least one developmental stage. We were able to detect alternative splicing in *L. major* with a greater sensitivity than has been previously reported (Rastrojo, 2013), presumably due to the deeper coverage of this dataset. Indeed, for genes with detectable *trans*-splicing events, alternative *trans*-splicing was pervasive with 88%, 56%, and 18% of genes using at least 5, 10, or 20 sites, respectively, in at least one developmental stage. This observation indicates that *L. major* exhibits a somewhat higher degree of alternative splicing than related species *T. cruzi* and *T. brucei* where <90% of genes were identified as alternatively spliced (Kolev, 2010) (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). This observation persisted even after accounting for differences in sequencing depth.

The distribution of the distances between the primary and minor *trans*-splicing sites revealed that almost half (~48%) of the alternative splice sites are located within 200 bases of the primary site in either direction. Even so, a significant percentage (18%) of minor sites were observed more than 1000 bp from the primary site, with most of these (78%) occurring upstream of the primary site.

An examination of the *trans*-splicing sites revealed a propensity for usage of the canonical acceptor sequence (AG) at both the primary (~97%) and minor (~43%) splicing sites (Table 1), consistent with previous findings in *T. cruzi* and *T. brucei* () (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). A sequence composition analysis of the region upstream of the SL-addition site allowed the visualization of the tail end of the polypyrimidine tract through the *trans*-splicing acceptor site (Figure 5D). As reported previously (Rastrojo, 2013; Requena, 2003), a C nucleotide was preferred prior to the AG acceptor sequence. When considering minor sites that are located within 1 kb of the primary site, a majority (64.2%) of minor sites that use the canonical AG acceptor are located downstream of the primary site (Figure 5A). This observation supports a model (based on a study of mammalian introns) that proposes that the 3´ splice site is located by a scanning process that recognizes the first AG downstream of the branch point in a sequence-specific context (Smith, 1989; Smith, 1993). When minor sites that do not use the canonical AG acceptor sequence were considered, this phenomenon was largely absent and the percentage of minor sites that are downstream of the primary site

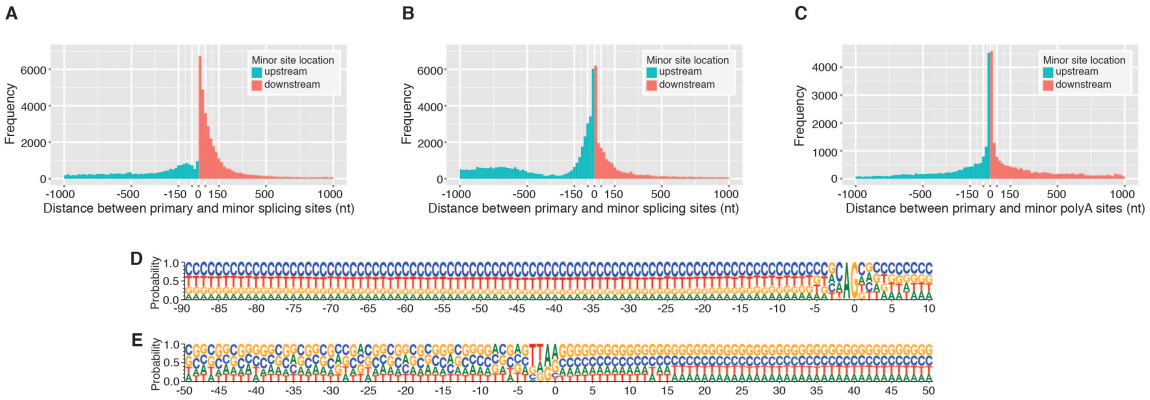


**Figure 5: Characterization of alternative RNA processing.**

Alternative RNA processing events were detected in both promastigote developmental stages and pooled for this analysis. The distribution of distances between primary and minor *trans*-splicing sites is shown for minor splice sites that use (A) the canonical AG acceptor sequence or (B) an acceptor sequence other than AG. Panel (C) depicts the distribution of distances between primary and minor polyadenylation sites. Only minor sites within 1000 nt of the primary site are plotted. The full ranges for A, B, and C, are -7234 to 5973, -7178 to 5,876, and -6429 to 5993, respectively. About 18% (11,201 of 62,098), 29% (24,303 of 82,947), and 10% (7,301 of 75,487) of values fell outside of the plotted range for A, B, and C, respectively. (D) Sequence composition of the region spanning from 90 nt upstream to 10 nt downstream of each primary *trans*-splicing site. (E) Sequence composition of the region spanning from 50 nt upstream to 50 nt downstream of each primary polyadenylation site.

| Acceptor Seqeunce | Primary | Minor |
|---|---|---|
| AG | 96.80% | 42.81% |
| TG | 0.68% | 11.27% |
| GG | 0.56% | 9.35% |
| CG | 0.47% | 7.79% |
| TT | 0.37% | 3.35% |
| GC | 0.21% | 3.41% |
| GT | 0.14% | 3.31% |
| TC | 0.14% | 2.16% |
| CA | 0.11% | 3.07% |
| AC | 0.11% | 2.35% |
| AT | 0.11% | 2.32% |
| GA | 0.08% | 2.11% |
| CT | 0.07% | 2.07% |
| CC | 0.06% | 1.89% |
| AA | 0.04% | 1.56% |
| TA | 0.04% | 1.19% |

**Table 1: Dinucleotide acceptor site usage frequency.**

The percentage of primary and minor *trans*-splicing sites that use each dinucleotide acceptor sequence.

drops to 36.3% (Figure 5B). This observation was maintained when procyclic and metacyclic promastigotes were considered separately.

Alternative polyadenylation sites were detected for 8,391 (~95%) of the 8,841 genes for which polyadenylation events were observed with 61%, 21%, and 5% of genes using at least 5, 10, or 20 polyadenylation sites. As observed for *trans*-splicing events, this observation indicates a higher degree of alternative polyadenylation in *L. major* than what has been reported in either *T. cruzi* or *T. brucei* where 63% and 92% of genes had detectable alternative polyadenylation, respectively (Li Y, Caradonna KL, Belew AT, Corrada Bravo H, Burleigh BA, El-Sayed NM, in revision). A sequence composition analysis was done to visualize the region surrounding the polyA-addition site. Even though no consensus motif was observed upstream of the polyadenylation site, such as the AAUAAA required for polyadenylation in higher eukaryotes, we did note an (A/G)(A/G) motif preceded by 1-2 thymines abutting the polyA addition site for both primary and minor polyadenylation sites (Figure 5E). Similar to what was found for SL addition sites above, the analysis of the distribution of the distances between primary and alternative sites revealed that ~49% of the minor polyadenylation sites were located within a 200 nt window of the primary site (Figure 5C).

Alternative *trans*-splicing or polyadenylation are suspected to play a role in the regulation of gene expression in *L. major*, but instances of regulation through alternative RNA processing between developmental stages have not been systematically identified. We sought to identify the subset of genes that change the use of their primary *trans*-splicing or polyadenylation sites between the

procyclic and metacyclic stages and to investigate possible correlations between these changes and differential expression. We were specifically interested in genes that showed a strong preferential usage for the primary site over other sites within a given stage (dominance), as determined using the ratio of reads that map to the primary site to those that map to the secondary site (P/S). Of the 8,797 genes that had at least one *trans*-splicing site identified in both stages, 523 showed preferential usage of different primary *trans*-splicing sites between the stages. We plotted the lengths of the UTRs for each gene, as determined by the primary *trans*-splicing site in each stage (Figure 6A). Each gene was represented by a single point with the color indicating the average P/S ratio for the two stages (thereby providing a measure of a primary site's dominance) and the size indicating the average number of reads mapping to that gene's primary sites (thereby indicating expression level and an indirectly providing confidence in the data). Data points along the diagonal represent genes that did not exhibit a change in the primary *trans*-splicing site between the stages. Largely, genes that had high dominance did not exhibit a change in primary site location between the stages, but instead used the same primary site in both procyclic and metacyclic promastigotes. Genes that did change primary site tended to have only a slight preference for each stage-specific primary site. A few interesting genes did not follow this trend and showed both a change in primary site (location away from the diagonal), high dominance (red), and high confidence/expression (large). Examples of the alternative usage of *trans*-splicing sites for a subset of these

interesting genes - LmjF.31.0710, LmjF.33.0310, and LmjF.36.3810 - are depicted in Figure 7.

We compared the expression profiles (see Chapter 3) for genes that displayed a change in *trans*-splicing site preference versus those that did not. The proportion of differentially expressed genes in both sets was nearly identical (39.4% for genes that changed primary site and 37.0% for genes that did not),

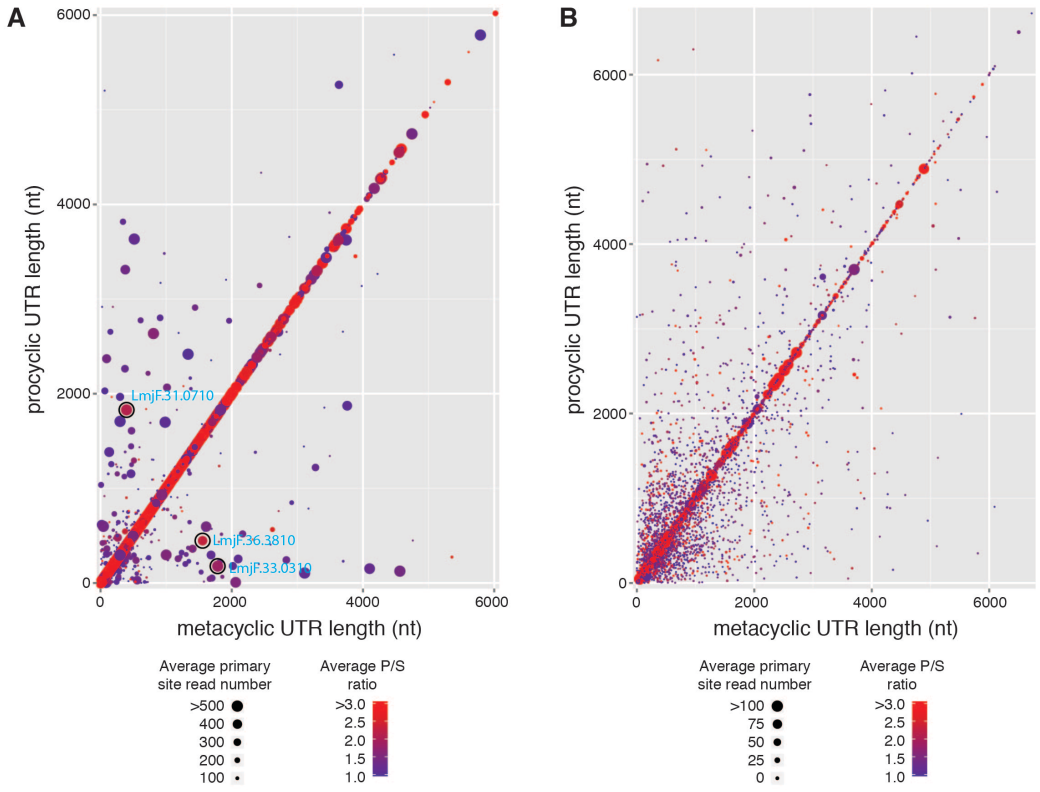


**Figure 6: Preferential usage of primary site across developmental stages.**

(A) For each gene, the length of the 5´ UTR determined by the primary trans-splicing site for the metacyclic stage was plotted against the length of the 5´ UTR determined by the primary *trans*-splicing site for the procyclic stage. Each point represents a single gene. Points that do not fall along the diagonal represent a change in the primary site between the stages. The color of each point represents the dominance of the primary site (preference for usage of the primary site over other sites within a stage), as determined by the ratio of primary site read counts to secondary site read counts (P/S), averaged for both stages. The size of each point depicts average read count for the primary sites from both stages, thereby showing expression levels and providing a measure of confidence in the data. Three genes highlighted in Figure 7 are circled with the gene identifier labeled in blue. (B) A similar plot was done for 3´ UTR length as determined by the primary polyadenylation site for each stage.

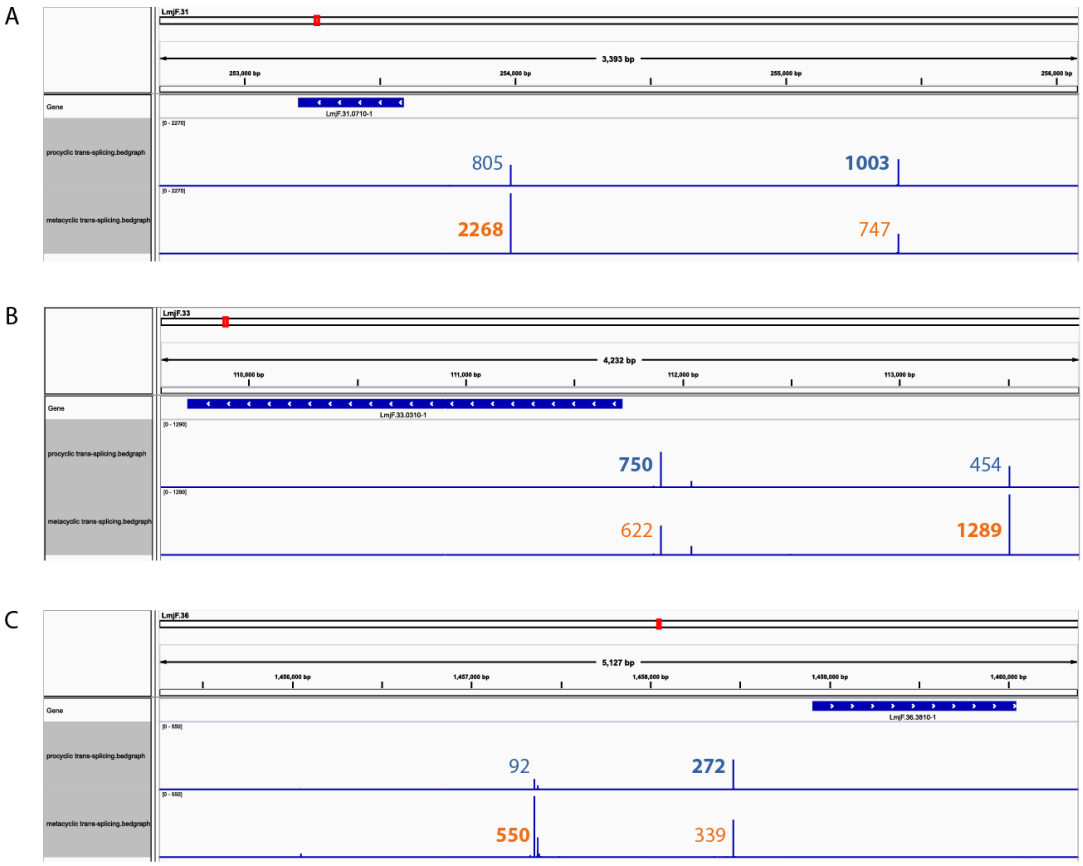


**Figure 7: Visualization of changes in primary *trans*-splicing sites across developmental stages.**

The Integrative Genomics Viewer (IGV) (Robinson, 2011) was used to visualize the number of reads that mapped to each *trans*-splicing site for three genes which showed a change in the preferred primary site between developmental stages and had a significant number of reads mapped to each primary site - LmjF.31.0710 (A), LmjF.33.0310 (B), and LmjF.36.3810 (C). The number of reads that mapped to each primary site (bold) and secondary site are shown for both procyclic promastigotes (blue text) and metacyclic promastigotes (orange text).

indicating that there was no association between changes in primary site used by individual genes and their expression levels ($\chi^2 = 0.28$).

A similar analysis was done to assess alternative polyadenylation between procyclic and metacyclic promastigotes (Figure 6B). Strikingly, this analysis revealed that most of the genes that showed high dominance (red) and high confidence (large) of their primary polyA sites did not exhibit a change in primary site location between the stages. For the large number of genes that showed a

change in the primary polyA site between developmental stages (4,377 of 8,337), very few reads were mapped to the primary sites, resulting in low confidence in these data points. The low numbers of mapped polyA-containing reads was likely due to the extensive heterogeneity of polyadenylation sites or the relative low coverage of polyA-containing reads mapping at unique sites. Of the 4,377 genes that changed primary polyadenylation sites between stages, the portion that was differentially expressed (37.2%) was remarkably similar to the portion that was not (38.7%; $\chi^2 = 0.15$). This lack of systematic association with differential expression also supports that conclusion that differences in expression levels in the samples examined by this study cannot be attributed to stage-regulated alternative RNA processing.

**Conclusion**

Since *Leishmania* and related trypanosomatids employ polycistronic transcription across their entire genomes, post-transcriptional RNA processing is thought to be a likely mechanism for regulating the levels of individual mRNAs. We were able to exploit signal sequences added during the processing of polycistronic pre-mRNAs to identify *trans*-splicing and polyadenylation sites in *Leishmania major* with unprecedented depth and reliability. In addition, we were able to evaluate how transcript structure compares to related species, to examine *trans*-splicing and polyadenylation events within and between *L. major* developmental stages, and to assess the possible relationship between alternative RNA processing and gene regulation in the context of the same biological samples.

The RNA-seq datasets generated in this work enabled us to precisely delineate the 5´ and 3´ UTR boundaries of *L. major* transcripts, providing a substantial resource for the *Leishmania* research community. While we were able to detect widespread alternative *trans*-splicing and polyadenylation for the large majority of genes, the observed heterogeneity of RNA processing sites was not systematically associated with the differential expression of the genes that showed the alternative RNA processing. Thus, in this analysis, alternative processing of pre-mRNA did not appear to be a driving force for determining the expression levels of individual genes. Even so, insights into *trans*-splicing and polyadenylation in trypanosomatids may shed light on gene regulation in others species that rely on similar mechanisms of post-transcriptional control, even if only for a subset of genes. Additionally, the precise definition of UTR boundaries opens up opportunities for regulatory motif analyses and comparative analyses of UTR usage across *Leishmania* species.

# Chapter 3

## Transcriptomic profiling of *Leishmania* metacyclogenesis

### Introduction

*Leishmania* parasites are taken up from the skin of an infected mammalian host upon blood feeding by a female phlebotomine sand fly. Within the sand fly midgut, these aflagellated amastigote-form parasites transform into flagellated procyclic promastigotes that replicate within the sand fly midgut. Further differentiation events occur and result in the generation of infective metacyclic promastigote parasites by the process of metacyclogenesis. These metacyclic promastigotes are able to infect a new mammalian host when deposited on the host's skin as the sand fly takes a subsequent blood meal (Bates, 2007).

Transcriptomic profiling by RNA-seq was used to identify global changes in gene expression as *Leishmania major* undergoes metacyclogenesis from the proliferative, non-infective procyclic promastigote form to the non-dividing, infective metacyclic promastigote form, a developmental progression that is well mimicked *in vitro* using axenic cultivation methods (Sacks, 1984). A differential expression analysis of relative RNA abundance between the promastigote forms was carried out to identify a robust set of markers for each developmental stage and to reveal genes and processes involved in the transition between stages as the parasite becomes capable of infecting mammalian host cells. This analysis resulted in the generation of reliable and substantially deep lists of differentially

expressed genes that include RNAs of low abundance which may have fallen outside of the limits of detection in past studies. Hundreds of genes of unknown function were also implicated in the developmental transition from the procyclic to metacyclic promastigote developmental forms, providing possible evidence for their function by guilt-through-association inferences.

**Experimental design**

RNA was isolated from cultured *L. major* grown to log phase (procyclic form) or enriched for metacyclic forms using: 1) a Ficoll gradient or 2) negative selection using peanut agglutinin (PNA). These two methods for metacyclic promastigote enrichment were used to test whether different methods for the procurement of metacyclic parasites could be responsible for different findings in previous studies (Akopyants, 2004; Depledge, 2009; Guerfali, 2008; Saxena, 2003). PolyA enriched cDNA libraries were generated using the Illumina TruSeq protocol and 100-bp paired end sequences were generated. A total of 6 procyclic promastigote biological replicates and 9 metacyclic promastigote biological replicates were collected (Table 11). Each procyclic replicate was matched to one or two metacyclic replicates from the same batch/expansion of cells. Phase contrast images, promastigote sample quantification, and an infectivity curve for the parasites in murine macrophages are provided in Figure 8.

A total of ~1.1 billion sequence reads were produced across the 15 samples, 91% of which mapped to the *L. major* reference genome (Table 11). For each sample, the number of reads mapping to existing gene annotations was

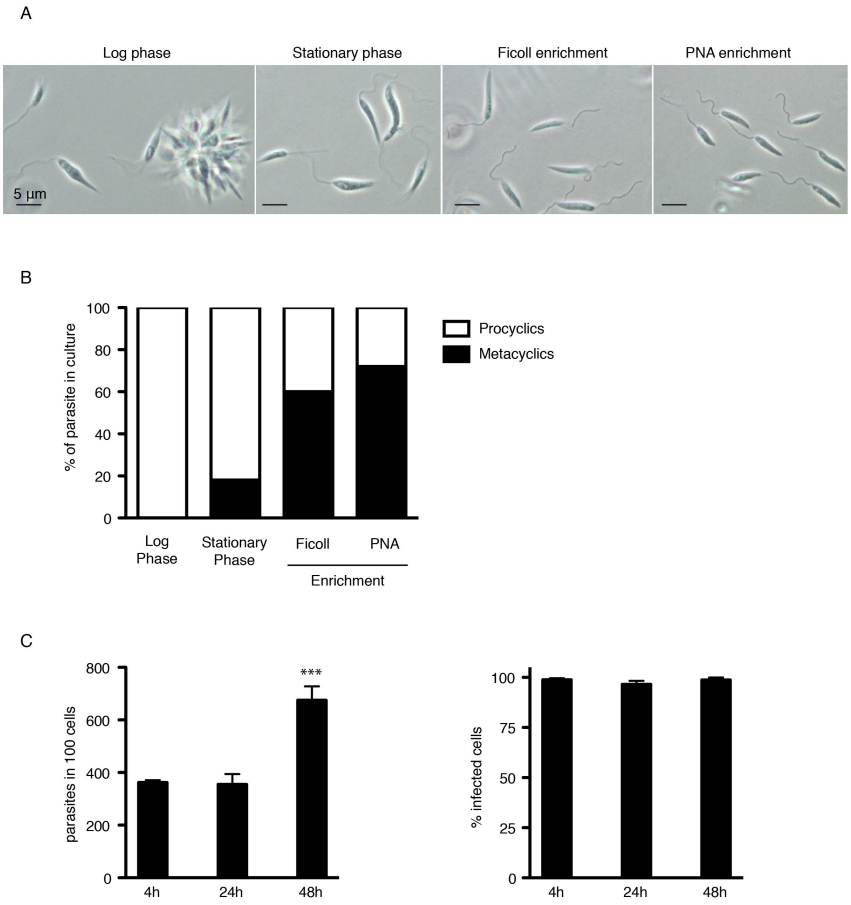


**Figure 8: Sample characterization.**

(A) Phase contrast images of log-phase procyclic promastigotes, stationary phase promastigotes prior to enrichment for metacyclics, promastigotes following negative selection by PNA, and Ficoll-purified promastigotes. The bar in each panel represents 5 µm. (B) Relative percentages of procyclic and metacyclic promastigotes in culture prior to and after the application of enrichment methods, as determined by counting 15 fields. (C) Infections were established in peritoneal macrophages isolated from C57BL/6 mice using Ficoll-purified metacyclic promastigotes at an MOI of 5:1 in the presence of C5-deficient serum from DBA mice. Plots show the number of parasites observed per macrophage and the percentage of infected macrophages observed over the first 48 hours of the infection from one representative experiment. Asterisks indicate a significant difference in the number of parasites per 100 cells at 48 hours.

determined. The resulting count table was restricted to the 8,486 protein-coding genes in the TriTrypDB *L. major* annotation v. 6.0.

## Statistical evaluation of biological replicates and batch effects

We used multiple robust statistical methods to evaluate the global characteristics of samples and to identify outlier samples that should be removed

prior to differential expression and gene structure analysis (Figure 9). Box plots were used to compare the distribution of per-gene read counts within each sample. All 15 samples showed a similar distribution of these counts with median steady-state expression levels of ~7.2 log2 counts per million and very few genes (5-10 per sample) expressed at levels of <4 counts per million. This observation is consistent with a lack of gene regulation at the level of transcription and may indicate that very few protein-coding transcripts are completely degraded following polycistronic transcription. A heatmap of Pearson correlations was used to visualize the relationship between each pair of samples. While all samples showed a pairwise correlation (r) of at least 0.85, samples prepared on one experimental date (batch A) were less correlated to samples from other batches, which largely showed r values of >0.95 when compared to one another. Median pairwise correlation was also computed to assess global correlation between samples and a standard outlier identification method was applied to establish a cut-off for the identification of outliers. Consistent with observations from the Pearson correlation heatmap analysis, this method identified the two samples from batch A as outliers. These 2 samples were excluded from further analyses.

The dataset used for differential expression analysis was further restricted to genes expressed at a level of at least 1 read per million in at least 5 of the 13 remaining samples. Of the 8,486 protein-coding genes analyzed, 8,475 met this threshold, consistent with observations described above that few genes were completely degraded after transcription. No statistical difference was found in protein coding gene expression between metacyclic promastigote samples

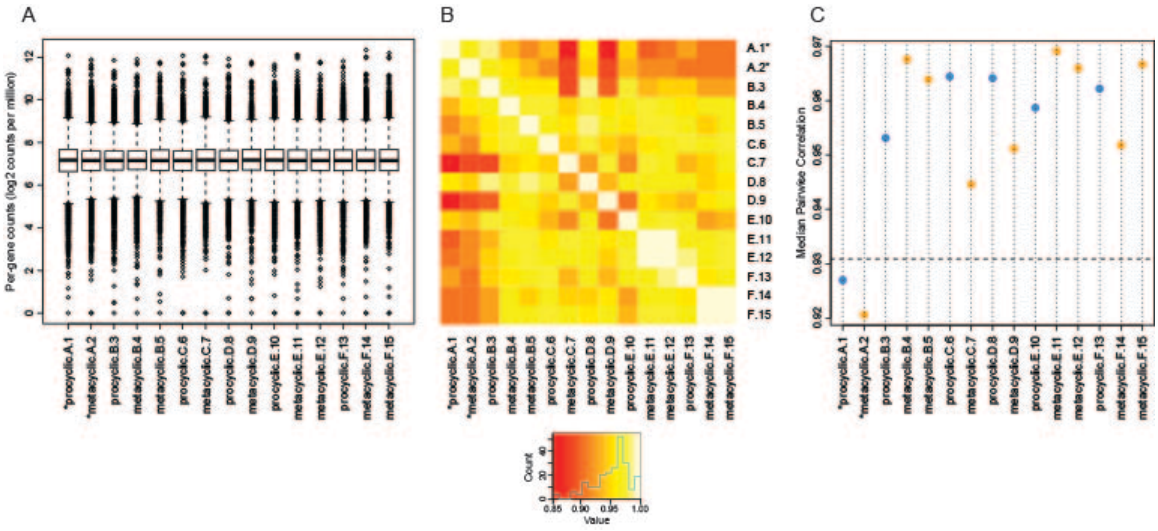


**Figure 9: Sample diagnostics to globally assess data similarities and identify outliers.**

RNA-seq was carried out using the Illumina platform on *L. major* procyclic and metacyclic promastigotes. Letters (A-F) in the sample name refer to experimental batch. Numbers are unique identifiers as shown in Table 11. Samples identified as outliers are indicated with an asterisk. (A) *Distribution of normalized gene counts by sample.* For each sample, counts were normalized for sequencing library size and a box plot was generated to compare the distribution of per-gene counts (log2 counts per million with an offset of 1). The ends of the whiskers represent the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Gene features with extremely high or low expression levels are shown as open circles above and below the whiskers, respectively. (B) *Heatmap of Pearson correlation between samples.* Raw count data were used to generate a heatmap to illustrate the Pearson correlation between samples. The color key and histogram for the frequency of correlation values (range of 0.85-1) is shown below the heatmap. (C) *Median pairwise correlation.* Raw count data were used to compute the median pairwise correlation between each sample and all other samples. The median pairwise correlation across all samples was used to establish a cutoff value to identify outlier samples (dotted line). Samples are colored according to stage (blue=procyclic, orange=metacyclic).

prepared using the Ficoll or PNA protocols. Consequently, all metacyclic promastigote samples were pooled together for the remainder of the analyses.

The large number of biological replicates used for the analysis necessitated the evaluation of the dataset for batch effects. A batch effect represents experimental variation caused by sub-groups of measurements that are independent of the underlying biology of the system being studied. They have been shown to introduce unwanted variability into biological studies and

confound the results, leading to erroneous conclusions. Previous analyses of high-throughput data, like those produced by RNA-seq, have indicated the need to assess and correct batch effects (Leek, 2010). In this study, we used experimental start date as a surrogate for batch when testing for differential expression between developmental stages of *L. major*.

Principal component analysis (PCA) and Euclidean distance heatmap analysis were used to visualize the relationship between samples both prior to and after accounting for batch effects (Figure 10). PCA reduces the dimensionality of a dataset while allowing variability to be represented to the greatest extent possible (Jolliffe, 2002). The PCA plots showed the first two principal components, which account for the greatest percent of variability in the data, on the X and Y axes, respectively, with each of the 13 samples represented as a single point. When batch was accounted for, a clear separation between procyclic promastigote and metacyclic promastigote samples was seen along the X axis of the PCA plot (Figure 10C). Separation between the stages was not as pronounced when batch was not considered (Figure 10A). Indeed, prior to accounting for batch effects, 25% of the variance represented by principal component (PC) 1 and 77% of the variance represented by PC2 were attributable to the batch of the samples.

Likewise, when Euclidean distance between samples was computed and used to create a heatmap color image and dendrogram depicting the closeness between samples, a clear separation between procyclic promastigote and metacyclic promastigote samples was observed after accounting for batch effects

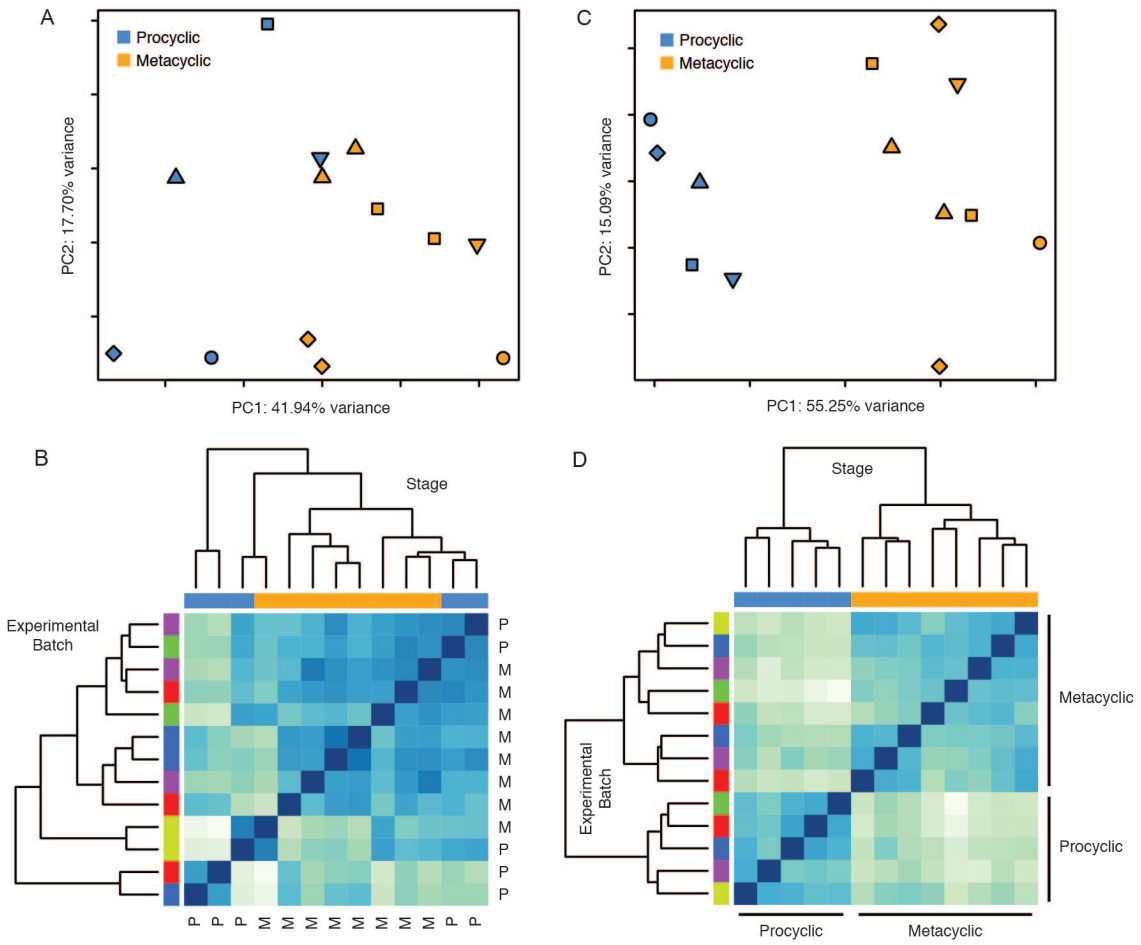


**Figure 10: Global gene expression profiles of the procyclic and metacyclic promastigote forms of *Leishmania major*.**

RNA-seq was carried out on *L. major* procyclic (log phase) promastigotes and metacyclic promastigotes isolated after enrichment using Ficoll or PNA. A principal component analysis (PCA) plot (A,C) and heatmap of a hierarchical clustering analysis using the Euclidean distance metric (B,D) are shown prior to (A-B) and after (C-D) accounting for batch effects in the statistical model. Analyses were performed using all *L. major* annotated genes after filtering for low counts and quantile normalization (8,475 genes). In the PCA plot, each point represents an experimental sample with point color indicating *L. major* developmental stage (blue = procyclic promastigote, orange = metacyclic promastigote) and point shape indicating batch/experimental date. Colors along the top of the heatmap indicate the developmental stage (blue = procyclic promastigote, orange = metacyclic promastigote) and colors along the left side of the heatmap indicate the batch/experimental date.

(Figure 10D) but not before (Figure 10B). As a result of these analyses, batch effects were controlled for in the subsequent differential expression analysis by including experimental batch as a covariate in the statistical model used by `limma`.

**Identification of genes differentially expressed between stages**

Differential expression (DE) analysis identified 3,138 genes that were expressed at significantly different levels between procyclic and metacyclic promastigotes at an adjusted *P* value cutoff of <0.05. Fold change differences ranged from 3.1-fold downregulated to 3.6-fold upregulated in metacyclic promastigotes. These genes were visualized using an MA plot showing the relationship between mean expression and fold change for each gene (Figure 11). Almost 60% of the DE genes (1,829 of 3,138) are annotated as hypothetical proteins. The remaining gene products have been characterized to different extents, albeit not always in the context of their possible role(s) in metacyclogenesis.

We extended our DE analysis to a set of 1,044 novel open reading frames (ORFs) of at least 90 nt in length identified based on evidence of translation in *L. major* by ribosome profiling data (Table 12). The addition of these ORFs lead to an increase in the list of differentially expressed genes by ~12% (a total of 368, from 3,138 to 3,506) with fold changes ranging from 5.7-fold downregulated to 3.6-fold upregulated in the metacyclic stage. The top 25 down- and up-regulated genes are shown in Table 2. Of these, 8 are novel ORFs, including the most downregulated gene.

The list of DE genes was used as input into gene ontology (GO) analysis to identify cellular functions and processes that are enriched during *L. major* metacyclogenesis. Genes downregulated in metacyclic promastigotes were considered separately from upregulated genes. Forty GO categories were

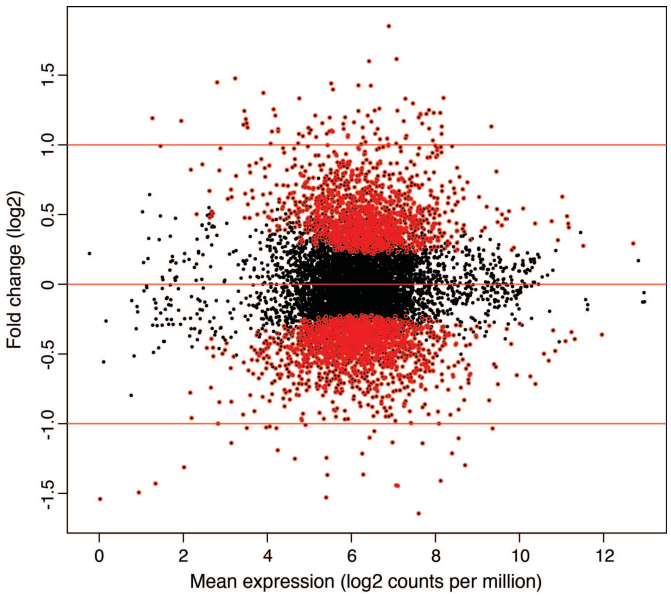


**Figure 11: MA plot of differentially expressed genes in the *L. major* procyclic to metacyclic promastigote transition.**

Differential expression analysis of *L. major* procyclic and metacyclic promastigotes was done using `limma` after `voom` transformation, taking experimental batch into account as part of the `limma` statistical model. The MA plot shows the relationship between mean expression (log2 counts per million with an offset of 0.5) and fold change. Each point represents one gene. Points colored in red represent 3115 genes expressed at significantly different levels between procyclic and metacyclic promastigotes at an adjusted *P* value of <0.05, with genes upregulated in the metacyclic stage relative to the procyclic stage exhibiting positive fold changes.

identified as being significantly enriched (*P* value cutoff of <0.05) for genes downregulated (33 categories) and upregulated (7 categories) in metacyclic promastigotes (Table 3).

## Examination of gene lists and gene ontology-based enrichment analyses

Many novel genes were identified among the most downregulated during metacyclogenesis, including multiple genes with unknown function. GO enrichment analysis of these genes reflected a clear reduction in a number of cellular processes including DNA replication and nucleosome assembly, translation-related activities (initiation and elongation), protein metabolism, and

| Metacyclic promastigotes, downregulated | | |
|---|---|---|
| **ID** | **Product Description** | **Fold Change** |
| LmjF.23_3931 | novel ORF, LmjF.23, 477090-477218 (-) | 5.65 |
| LmjF.31.3070 | iron/zinc transporter protein-like protein (LIT1) | 3.14 |
| LmjF.35.1310 | histone H4 | 2.90 |
| LmjF.36.0020 | histone H4 | 2.74 |
| LmjF.35.2160 | adenine aminohydrolase (AAH) | 2.72 |
| LmjF.31.3180 | histone H4 | 2.68 |
| LmjF.33.1760 | hypothetical protein, unknown function | 2.66 |
| LmjF.14.0470 | hypothetical protein, conserved | 2.65 |
| LmjF.21.0740 | ATPase subunit 9, putative | 2.57 |
| LmjF.35.2130 | hypothetical protein, unknown function | 2.56 |
| LmjF.33.3240 | h1 histone-like protein | 2.56 |
| LmjF.25.2450 | histone H4 | 2.48 |
| LmjF.32_7004 | novel ORF, LmjF.32, 1161019-1161147 (-) | 2.39 |
| LmjF.36.5845 | kinetoplast-associated protein, putative | 2.37 |
| LmjF.36.3080 | lipoate protein ligase, putative | 2.35 |
| LmjF.35.4760 | hypothetical protein, conserved | 2.34 |
| LmjF.02.0020 | histone H4 | 2.34 |
| LmjF.32.2940 | hypothetical protein, conserved | 2.33 |
| LmjF.23.0200 | endoribonuclease L-PSP (Pb5) | 2.22 |
| LmjF.35_8354 | novel ORF, LmjF.35, 877847-877972 (+) | 2.20 |
| LmjF.25.1470 | cyclin (CYCA) | 2.20 |
| LmjF.20.0030 | histone-lysine N-methyltransferase, putative (DOT1) | 2.20 |
| LmjF.19_3054 | novel ORF, LmjF.19, 382655-382816 (+) | 2.16 |
| LmjF.13_1846 | novel ORF, LmjF.13, 171578-171685 (+) | 2.16 |
| LmjF.36.3910 | S-adenosylhomocysteine hydrolase | 2.16 |
| **Metacyclic promastigotes, upregulated** | | |
| **ID** | **Product Description** | **Fold Change** |
| LmjF.34.0070 | ascorbate peroxidase (APX) | 3.61 |
| LmjF.19_3059 | novel ORF, LmjF.19, 395719-395889 (+) | 3.31 |
| LmjF.02.0460 | voltage-dependent anion-selective channel, putative | 3.04 |
| LmjF.17.0890 | META domain containing protein (META1) | 3.03 |
| LmjF.23.0730 | RNA-binding protein, putative | 2.97 |
| LmjF.12.0480 | hypothetical protein, unknown function | 2.95 |
| LmjF.28.0980 | P27 protein, putative (P27) | 2.77 |
| LmjF.23.0780 | hypothetical protein, conserved | 2.68 |
| LmjF.29.1350 | RNA binding protein, putative | 2.68 |
| LmjF.16.0500 | hypothetical protein, unknown function | 2.68 |
| LmjF.22.0250 | phosphoinositide phosphatase | 2.63 |
| LmjF.29.1360 | RNA binding protein, putative | 2.63 |
| LmjF.36.2290 | serine/threonine protein kinase | 2.61 |
| LmjF.34.1940 | amastin-like surface protein, putative | 2.54 |
| LmjF.17_2659 | novel ORF, LmjF.17, 423627-423884 (+) | 2.53 |
| LmjF.04.0350 | hypothetical protein, conserved | 2.52 |
| LmjF.16.1050 | hypothetical protein, conserved | 2.52 |
| LmjF.34.2500 | protein phosphatase 2C-like protein | 2.49 |
| LmjF.35.5000 | hypothetical protein, conserved | 2.46 |
| LmjF.04.1210 | casein kinase I, putative | 2.45 |
| LmjF.34.1820 | amastin-like surface protein, putative | 2.44 |
| LmjF.12.0460 | hypothetical protein, unknown function | 2.43 |
| LmjF.09_1121 | novel ORF, LmjF.09, 124955-125314 (+) | 2.41 |
| LmjF.34.1800 | amastin-like surface protein, putative | 2.40 |

**Table 2: Top differentially expressed genes in the *L. major* procyclic to metacyclic promastigote transition.**

A total of 3,506 previously annotated genes and novel ORFs were differentially expressed (DE) between procyclic and metacyclic promastigotes. The top 25 down- and up-regulated genes/novel ORFs are shown.

| Metacyclic promastigotes, downregulated | | |
|---|---|---|
| **GO ID** | **GO term** | ***P* Value** |
| GO:0015986 | ATP synthesis coupled proton transport | 7.36E-11 |
| GO:0005737 | cytoplasm | 5.33E-09 |
| GO:0015991 | ATP hydrolysis coupled proton transport | 2.00E-08 |
| GO:0046961 | proton-transporting ATPase activity, rotational mechanism | 2.79E-08 |
| GO:0004812 | aminoacyl-tRNA ligase activity | 6.96E-08 |
| GO:0006418 | tRNA aminoacylation for protein translation | 6.96E-08 |
| GO:0046933 | proton-transporting ATP synthase activity, rotational mechanism | 7.26E-08 |
| GO:0044267 | cellular protein metabolic process | 1.41E-07 |
| GO:0003746 | translation elongation factor activity | 1.55E-07 |
| GO:0005634 | nucleus | 3.38E-07 |
| GO:0006260 | DNA replication | 1.25E-06 |
| GO:0005525 | GTP binding | 1.42E-06 |
| GO:0003677 | DNA binding | 1.94E-06 |
| GO:0003924 | GTPase activity | 2.64E-06 |
| GO:0043234 | protein complex | 1.64E-05 |
| GO:0051258 | protein polymerization | 1.64E-05 |
| GO:0003743 | translation initiation factor activity | 1.74E-05 |
| GO:0045261 | proton-transporting ATP synthase complex, catalytic core F(1) | 5.23E-05 |
| GO:0006457 | protein folding | 7.78E-05 |
| GO:0051082 | unfolded protein binding | 8.59E-05 |
| GO:0005874 | microtubule | 1.07E-04 |
| GO:0004298 | threonine-type endopeptidase activity | 1.37E-04 |
| GO:0005839 | proteasome core complex | 1.37E-04 |
| GO:0051603 | proteolysis involved in cellular protein catabolic process | 1.37E-04 |
| GO:0006334 | nucleosome assembly | 1.47E-04 |
| GO:0006413 | translational initiation | 1.93E-04 |
| GO:0046982 | protein heterodimerization activity | 2.60E-04 |
| GO:0003887 | DNA-directed DNA polymerase activity | 3.45E-04 |
| GO:0006414 | translational elongation | 6.65E-04 |
| GO:0005198 | structural molecule activity | 9.43E-04 |
| GO:0004175 | endopeptidase activity | 9.86E-04 |
| GO:0016272 | prefoldin complex | 1.07E-03 |
| GO:0050660 | flavin adenine dinucleotide binding | 1.09E-03 |
| **Metacyclic promastigotes, upregulated** | | |
| **GO ID** | **GO term** | ***P* Value** |
| GO:0004674 | protein serine threonine kinase activity | 4.61E-22 |
| GO:0006468 | protein phosphorylation | 1.91E-21 |
| GO:0004672 | protein kinase activity | 8.67E-20 |
| GO:0004713 | protein tyrosine kinase activity | 1.28E-18 |
| GO:0005524 | ATP binding | 3.82E-11 |
| GO:0006950 | response to stress | 2.77E-10 |
| GO:0016791 | phosphatase activity | 1.63E-04 |

**Table 3: Gene ontology (GO) categories enriched during the *L. major* procyclic to metacyclic promastigote transition.**

`GOseq` (Young, 2010) was used to perform gene ontology analysis using differentially expressed genes identified as the parasite undergoes metacyclogenesis. Using a *P* value cut off of <0.05, a total of 33 GO categories were enriched among genes that were downregulated in metacyclic promastigotes and a total of 7 GO categories were enriched among genes that were upregulated in metacyclic promastigotes.

energy metabolism (i.e., ATP synthesis) while enriched GO categories for genes upregulated in metacyclic promastigotes indicated an increase in cell signaling and stress response (Table 3).

A close examination of differentially expressed genes and genome ontology enrichments confirmed earlier findings and, more importantly, revealed new insights into the parasite's transformation at a critical stage of its life cycle. The top downregulated gene in metacyclic promastigotes, LIT1 (LmjF.31.3070) is an iron transporter previously reported to be upregulated by the parasite upon iron depletion (Huynh, 2006). Its downregulation in metacyclics is consistent with the low metabolic rate and low demand for ATP in this developmental stage of the parasite. Interestingly, its paralogous copy (LmjF.31.3060) was regulated to a lesser extent (downregulated only ~1.5-fold in metacyclics). Multiple histones (H2A, H2B, H4, and H1 histone-like protein) previously identified as downregulated during metacyclogenesis (Genske, 1991; Soto, 2004) were also identified as such in this analysis with H4 mRNA levels particularly depleted. The decrease in histone transcripts as the parasite enters the non-dividing stationary phase suggests a mode of regulation that is dependent on the cell cycle and is consistent with observations in higher eukaryotes that histone gene expression decreases in differentiated cells (Gerbaulet, 1992; Stein, 1996). Also consistent with previous findings, multiple β-tubulin family members were identified as downregulated ~1.4-fold as the parasite becomes infective (Coulson, 1996). The downregulation of β-tubulin as the parasite undergoes metacyclogenesis correlates with morphological changes of the parasite as it prepares to enter host

cells. Additionally, the steady-state RNA level for adenine aminohydrolase (AAH), a purine metabolism protein that converts adenine to hypoxanthine and lacks homologs in humans as well as *T. cruzi* and *T. brucei* (Boitz, 2013), was found to be reduced in *L. major* metacyclic promastigotes in our study, as were cyclin A and DOT1, which are both involved in cell cycle progression (Hochegger, 2008; Kim, 2014).

The top upregulated gene, ascorbate peroxidase (APX), is protective against both endogenous and exogenous $H_2O_2$ and appears to play a role in differentiation to the metacyclic form as well as in protecting the cell against oxidative stress-induced apoptosis (Pal, 2010). Other genes that were upregulated in metacyclic promastigotes include casein kinase 1, a serine/threonine protein kinase that exists in multiple isoforms and has been identified as playing a role in *Leishmania* infectivity (Allocco, 2006), and *meta1,* which encodes a protein that localizes in the region of the flagellar pocket of stationary phase promastigotes and is thought to play a role in virulence, potentially through altering secretory processes (Nourbakhsh, 1996; Puri, 2011). The *p27* gene, which encodes a mitochondrial membrane protein that is an important component of the cytochrome oxidase complex, was also more abundantly expressed in metacyclic promastigotes. This result is consistent with previous findings reporting its upregulation in both metacyclics and intracellular amastigotes and its role in promoting parasite survival and virulence in the host (Dey, 2010) and supports the hypothesis that metacyclic promastigotes are pre-adapted to survival within the mammalian host (Sacks, 1989). Finally, two known

differentiation markers of metacyclic promastigotes, SHERP and HASPB (Flinn, 1994; Knuepfer, 2001; Sádlová, 2010), were also identified in this analysis, with SHERP upregulated ~1.9-fold and HASPB upregulated ~2.3-fold in metacyclic promastigotes.

The results of the differential expression analysis were compared to the list of differentially expressed genes identified in an earlier study by Saxena *et al. (Saxena, 2003)* that used microarrays of PCR-amplified fragments from genomic survey sequence (GSS) clones. Only GSS clones whose 5´ and 3´ sequences could be mapped to the same gene in the *L. major* Friedlin genome sequence (31 in total) were considered in our comparison and 19 of the corresponding genes showed a similar differential expression trend, albeit to varying degrees and levels of significance. Given the disparate platforms, the level of agreement was reasonable.

**Conclusion**

This work represents a comprehensive characterization of the global transcriptional changes that define the transition of the human pathogen *Leishmania major* from its non-infective to human-infective forms. Through the exploitation of massively parallel sequencing to detect subtle changes in steady-state levels of mRNA, the use multiple biological replicates to derive robust statistical analyses, and the careful consideration of batch effects that often confound and mask true biological signal, we observed changes in expression profiles between the procyclic promastigote and metacyclic promastigote forms of *L. major*. Indeed, since the individual fold changes observed were relatively

modest (3.1-fold downregulated to 3.6-fold upregulated in metacyclics), they may have been missed except for the sensitivity afforded by the RNA-seq technology and the statistical power provided by the use of 5 biological replicates. The extension of the differential expression analysis to novel ORFs identified from ribosome profiling data pointed to some genes that were not included in the initial annotation of *L. major* which may have functionally important roles in the differentiation of the parasite. These genes should be added to the list of candidates to be included in future analyses.

GO analysis using the lists of differentially expressed genes revealed how they are collectively involved in a number of cellular processes as the parasite transforms into its infective form. Specifically, as the parasites become infective, processes including cell signaling and stress responses increased while translation, replication, and metabolism decreased. The transcriptome profiles reported here set the stage for the construction of co-expression networks, which are useful for identifying driver mechanisms underlying co-regulation and for tentatively annotating unknown genes through guilt-by-association inferences.

# Chapter 4

## Transcriptomic profiling of *Leishmania*-infected murine macrophages

### Introduction

In order to survive and replicate, intracellular pathogens must survive the innate and adaptive immune responses of the host. A number of human pathogens - including *Leishmania* sp., *Trypanosoma cruzi*, *Mycobacterium tuberculosis*, *Toxoplasma gondii*, *Francisella tularensis*, *Legionella pneumophila*, and *Ehrlichia* - have evolved mechanisms not only to evade the host immune system, but to infect the very immune cells that are recruited to clear an infection. *Leishmania* is used here as a model of intracellular infection of immune cells to study transcriptomic changes that take place in both the host and the pathogen over the course of an infection.

*Leishmania* enter mammalian macrophages by receptor-mediated phagocytosis and are thought to do so in a quiescent manner, failing to produce a significant oxidative burst and to activate the innate immune system (Bennett, 2001; Laskay, 2003; Locksley, 1988; Zhang, 2010). It has previously been shown that a Th1 response by the host leads to parasite killing while a Th2 response leads to parasite growth (Etges, 1998; Reiner, 1995; Scharton-Kersten, 1995). *Leishmania* thus attempt to prevent their killing by altering cytokine expression (thereby influencing T cell responses), impeding antigen display by MHC class II molecules, and hindering nitric oxide production (reviewed in (Kaye, 2011;

Olivier, 2005; Sacks, 2002)). Less is known about the mechanisms used by the parasite to influence the host response and the global changes that take place at the transcriptomic level in both the parasite and host over the course of an infection.

Previous studies using microarrays or SAGE tags have started to elucidate changes that occur within *Leishmania* or within the mammalian host as infection occurs, but have been hindered by technical limitations that have prevented the interrogation of both systems simultaneously. In this study, we performed transcriptomic profiling of *L. major*-infected murine macrophage cells to identify genes that were differentially expressed by both the parasite and host cells as *Leishmania* entered and persisted within macrophages during the first 72 hours of an infection.

**Infection dynamics and global transcription patterns**

Transcriptomic profiling by RNA-seq was used to identify global changes in gene expression over the course of the first 72 hours of an infection of murine macrophages with *Leishmania major*. Peritoneal macrophages from C57BL/6 mice were infected with *L. major* metacyclic promastigotes and samples collected at 4, 24, 48, and 72 hours post-infection (hpi). The dynamics of the infection were monitored by counting the number of parasites per 100 macrophages (Figure 12). RNA sequencing was carried out for each sample and for matched uninfected controls, as well as for the *L. major* metacyclic promastigotes used for the infection. Over 2.4 billion sequence reads were collected across three

independent experiments (biological replicates) and labeled as batches A-C (Table 13).

Since there is little sequence conservation between mouse and *L. major*, we were able to unambiguously map reads from mouse and parasite RNAs from the mixed sample to their respective genomes. For uninfected samples, the percentage of reads mapping to the mouse genome ranged from 88.9% to 95.9%. Greater heterogeneity was observed for the percentage of reads mapping to *L. major* in the infected samples, providing clues about the transcriptional activity of both parasite and host (Figure 12). For each batch, the percentage of reads mapping to the parasite decreased over the course of the infection. Interestingly, this decrease in the relative portion of parasite reads did not match the parasite load of the infected cells, which decreased sharply between 4 and 24 hours for each batch, but slowly rose again from 24 through 72 hours (Figure 12). Possible explanations for the mismatch between the increasing number of parasites and the decreasing percentage of parasite reads is that the individual parasites are less transcriptionally active as the infection progresses or that the mouse cells became more transcriptionally active over the course of the infection, thereby diluting the proportion of reads attributable to the parasite.

Samples from host and from parasite showed a similar distribution of per-gene read counts per sample, as visualized by box plots (Figure 13), but striking differences were observed when comparing the two organisms. The median steady-state expression level was elevated in *L. major* compared to mouse (6.0 v vs. 3.6 log2 counts per million) and *L. major* genes showed a much more narrow

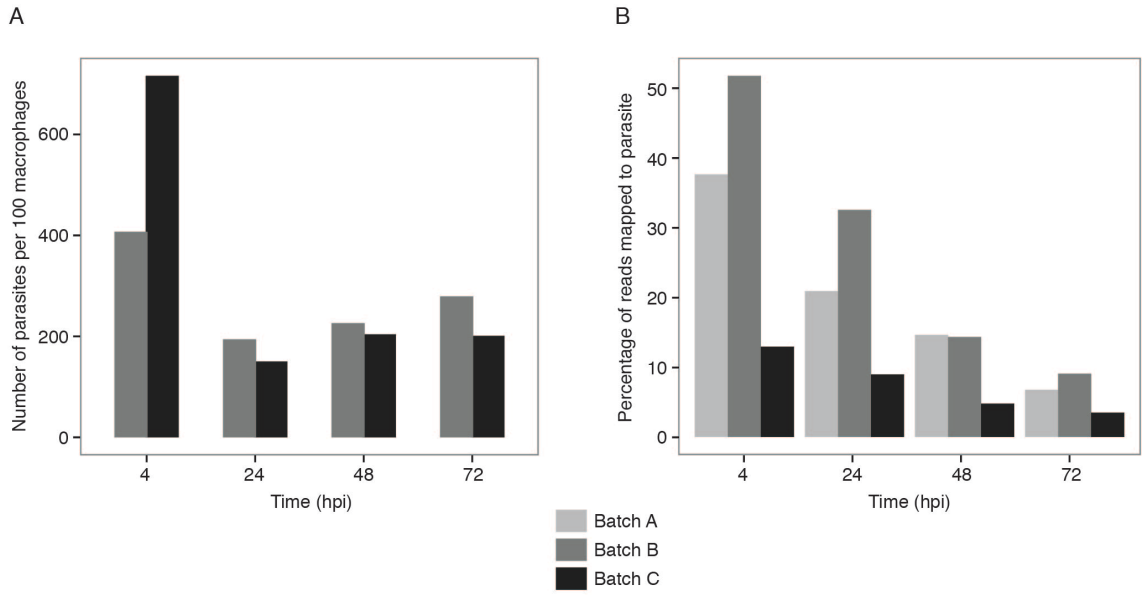


**Figure 12: Characterization of L. major intracellular growth and proportion of RNA from the parasite.**

Mouse macrophages infected with *L. major* were collected at 4, 24, 48, and 72 hpi and subjected to transcriptional profiling by RNA-seq. Bar plots are used to illustrate (A) the number of parasites observed per 100 host cells (available for 2 of 3 replicates) and (B) the percentage of trimmed RNA-seq reads that map to the *L. major* genome. Data from each experimental batch is shown using a different shade of gray.

distribution (interquartile ranges of 5.6 to 6.5 for *L. major* and 0 to 6.8 for mouse). Additionally, when non- and lowly-expressed genes were removed from the datasets prior to differential expression analysis, this filter led to the removal of 10,548 genes from the mouse dataset, but only 7 genes from the parasite one.

The differences in the distribution of genes between mouse and *L. major* are consistent with differences in how each organism controls gene expression - mouse employs both transcriptional and post-transcriptional mechanisms to control the expression levels of individual genes, resulting in many genes that are not expressed and a wide dynamic range for those that are. This differs from *Leishmania*, which employs polycistronic transcription by RNA polymerase II at roughly the same rate across its genome (Ivens, 2005; Martinez-Calvillo, 2003).

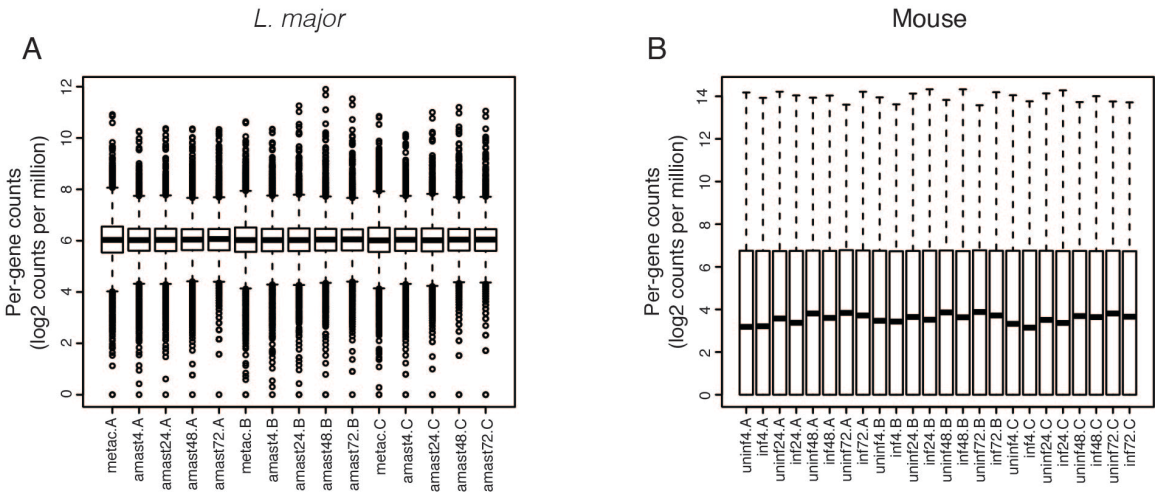


**Figure 13: Global assessment of data distribution.**

For each sample, the counts of reads mapping to *L. major* (A) and mouse (B) were normalized for sequencing library size and a box plot was generated to compare the distribution of per-gene counts (log2 counts per million with an offset of 1). The ends of the whiskers represent the lowest datum still within 1.5 interquartile range (IQR) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Gene features with extremely high or low expression levels are shown as open circles above and below the whiskers, respectively. Samples are named according to sample type, timepoint, and experimental batch.

The resulting polycistronic mRNAs are split into their component mature mRNAs by coupled *trans*-splicing and polyadenylation events (LeBowitz, 1993; Matthews, 1994; Ullu, 1993). Parasite mRNAs levels are thus not determined by the amount of transcription, but by post-transcriptional processes such as RNA processing and degradation (reviewed in (Clayton, 2007)). The dearth of lowly expressed *L. major* genes may indicate that very few mRNAs are completely degraded following polycistronic transcription.

## Statistical assessment of biological replicates and batch effects

The use of multiple biological replicates necessitated the evaluation of the data to assess reproducibility and account for batch effects, i.e. experimental variation caused by sub-groups of measurements that are not related to the

underlying biology of the system being studied. Previous analyses of high-throughput data, like those produced by RNA-seq, have indicated the need to assess and correct batch effects in order to prevent misinterpretation of results (Dillon, 2015; Leek, 2010). In this study, experimental start date was used as a surrogate for batch when testing for differential expression.

We used principal component analysis (PCA) and Euclidean distance heatmap analysis to visualize the relationship between experimental datasets both prior to and after adjusting for batch effects (Figure 14 and Figure 15, respectively). The PCA plots revealed that samples from the same experimental condition (hpi and infection status) grouped together for the parasite and mouse macrophages demonstrating a high level of reproducibility between replicates. The dendrograms associated with the Euclidean distance heatmaps further illustrated this point, with like samples clustering most closely with one another. The grouping of samples by experimental condition was partially obscured when batch effects were not considered, illustrating the benefits of this approach. Batch effects were therefore controlled for in the subsequent differential expression analyses by including experimental batch in the statistical models used.

The PCA and clustering analysis also suggested interesting biological relationships between the samples. The global profile of *L. major* gene expression changed over the time course of the experiment, moving from left to right across principal component (PC) 1 (Figure 15A). The clear separation between promastigote and amastigote samples along PC1 highlights the differences between the two developmental stages. While a time progression can

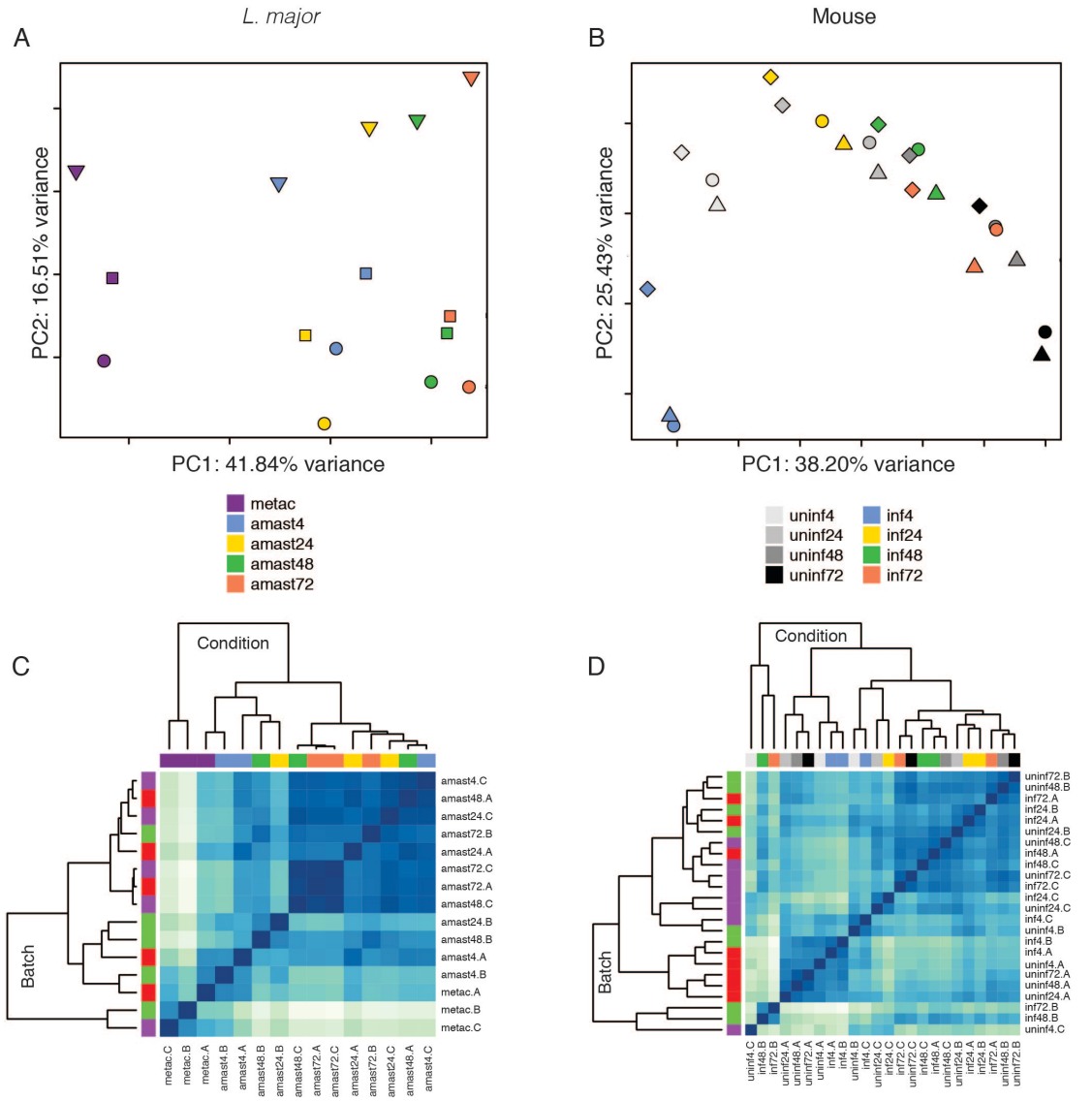


**Figure 14: Principal Component Analysis (PCA) and hierarchical clustering analysis prior to accounting for batch effects.**

RNA-seq was carried out on mouse macrophages infected with *L. major* at 4, 24, 48, and 72 hpi as well as on the metacyclic promastigotes used for the infection. Principal component analysis (PCA) plots and heatmaps of hierarchical clustering analyses using Euclidean distance are shown for the *L. major* (A, C) and mouse (B, D) transcriptomes over the course of the experiment. The analyses were performed using all annotated protein-coding genes following filtering for low counts and quantile normalization prior to accounting for batch effects in the statistical model (8,479 genes for *L. major* and 12,552 genes for mouse). In the PCA plots, the first two principal components are shown on the X and Y axes, respectively, with the proportion of total variance attributable to that PC indicated. Each experimental sample is represented as a single point with color indicating sample type/timepoint and shape indicating experimental batch. Colors along the tops of the heatmaps indicate the sample type/timepoint and colors along the left sides of the heatmaps indicate the experimental batch. Samples are named according to sample type, timepoint, and experimental batch.

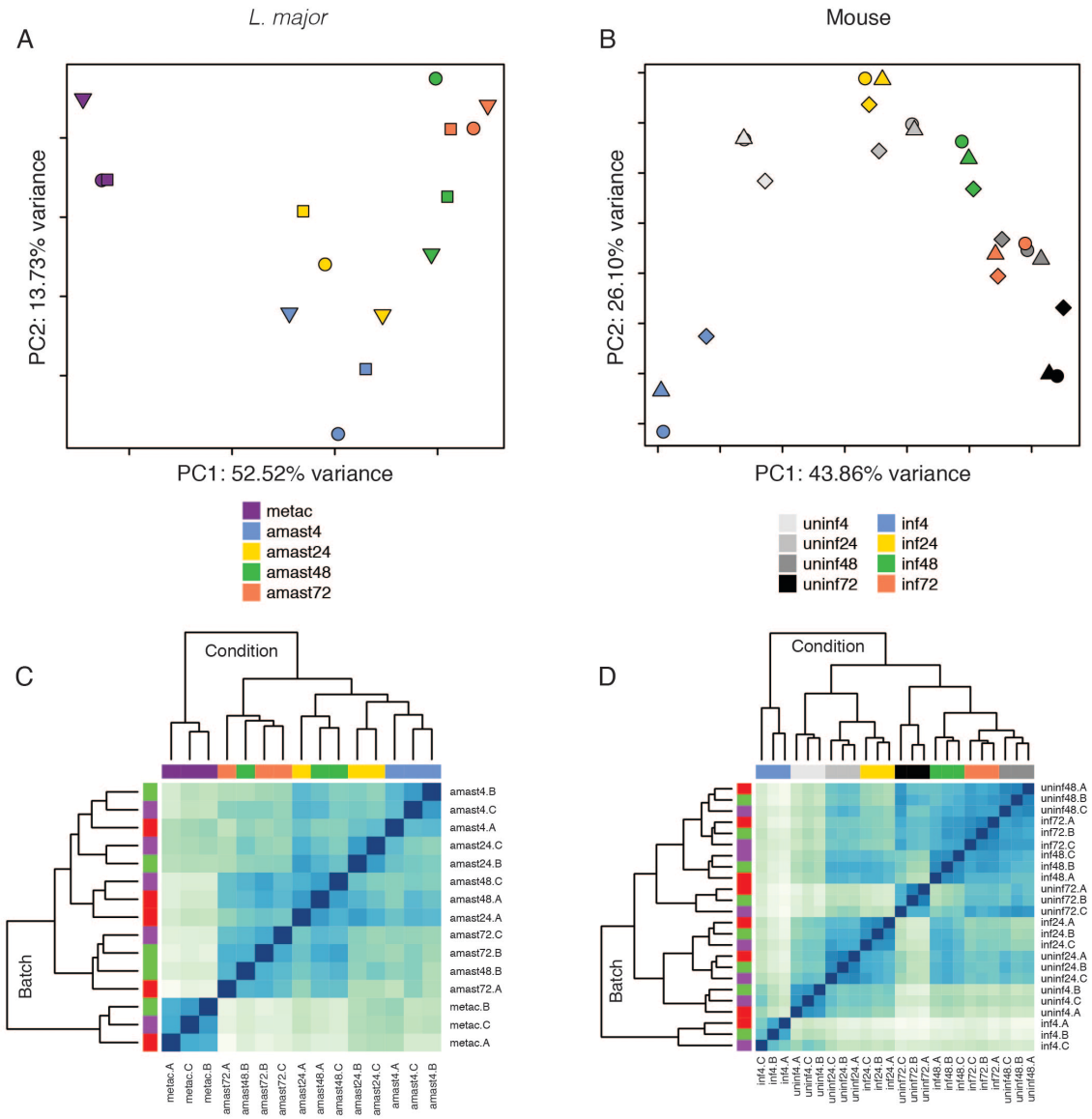


**Figure 15: Principal Component Analysis (PCA) and hierarchical clustering analysis after accounting for batch effects.**

RNA-seq was carried out on mouse macrophages infected with *L. major* at 4, 24, 48, and 72 hpi as well as on the metacyclic promastigotes used for the infection. Principal component analysis (PCA) plots and heatmaps of hierarchical clustering analyses using Euclidean distance are shown for the *L. major* (A, C) and mouse (B, D) transcriptomes over the course of the experiment. The analyses were performed using all annotated protein-coding genes following filtering for low counts and quantile normalization after accounting for batch effects in the statistical model (8,479 genes for *L. major* and 12,552 genes for mouse). In the PCA plots, the first two principal components are shown on the X and Y axes, respectively, with the proportion of total variance attributable to that PC indicated. Each experimental sample is represented as a single point with color indicating sample type/timepoint and shape indicating experimental batch. Colors along the tops of the heatmaps indicate the sample type/timepoint and colors along the left sides of the heatmaps indicate the experimental batch. Samples are named according to sample type, timepoint, and experimental batch.

be seen within the intracellular amastigote samples, some intermixing of samples from different timepoints was observed beyond 4 hpi (Figure 15C), suggesting an overlap in the gene expression profiles for these samples.

A time progression was also seen in the mouse macrophage data for both infected and uninfected samples (Figure 15B and Figure 15D). This observation underscored the necessity of collecting uninfected controls from each timepoint studied rather than relying on a control from a single timepoint. In addition, all infected macrophages from 24-72 hpi clustered more closely with the uninfected macrophages while the 4 hpi infected macrophages clustered apart. This is especially apparent in the heatmap dendrogram where the 4 hpi infected macrophages appear as a major outgroup.

**Differential expression and pathway enrichment analyses in murine macrophages**

While clustering analyses provided a high level overview of the behavior of the murine macrophage and parasite transcriptomes during the infection, further analyses were needed to evaluate changes in the expression levels of individual genes. A differential expression analysis was carried out to closely dissect the host murine macrophage response to infection by *Leishmania major* at 4, 24, 48, and 72 hours after infection. Pairwise analyses were conducted within and across individual timepoints with infected macrophages evaluated against matched uninfected macrophages for each timepoint.

The most substantial response to infection by the macrophage was observed at 4 hpi, with 6897 genes differentially expressed (DE) between

uninfected and infected cells with fold changes (FC) ranging from 29-fold downregulated to 56-fold upregulated. The response is reduced in later timepoints as reflected in smaller numbers of DE genes - 931, 1813, and 1460 genes at 24, 48, and 72 hpi, respectively - and in reduced fold changes with no downregulation beyond 12-fold or upregulation beyond 18-fold for these timepoints. The breakdown of up- and down-regulated genes for each timepoint comparison is illustrated in Figure 16 and Figure 17 (top).

A direct comparison of the overlap in DE genes at each timepoint revealed that the macrophage response to infection at 4 hpi was vastly different from the response at the later timepoints (Figure 17B and Figure 18). Of the 6897 DE genes at 4 hpi, 93% (6418 genes) were unique to 4 hpi. This is in contrast to 40% of genes at 24 hpi (377), 33% of genes at 48 hpi (598), and 33% of genes at 72 hpi (484) that were DE only at that specific timepoint. Indeed, only 47 genes were consistently up- or down-regulated at all 4 timepoints (38 genes upregulated at all timepoints and 9 genes downregulated at all timepoints). These genes do not appear to be functionally related, although the list does include the heavy metal transporters metallothionein 1 and 2. The latter has previously been associated with *Leishmania* resistance to treatment with antimonial drugs (Gómez, 2014). Interestingly, two genes associated with *Bcl2* (*Bnip3* and *Bcl2a1b*), an inhibitor of apoptosis, are also on this list, suggesting that infection by *L. major* may induce changes in host gene expression to prevent macrophages from undergoing cell death.

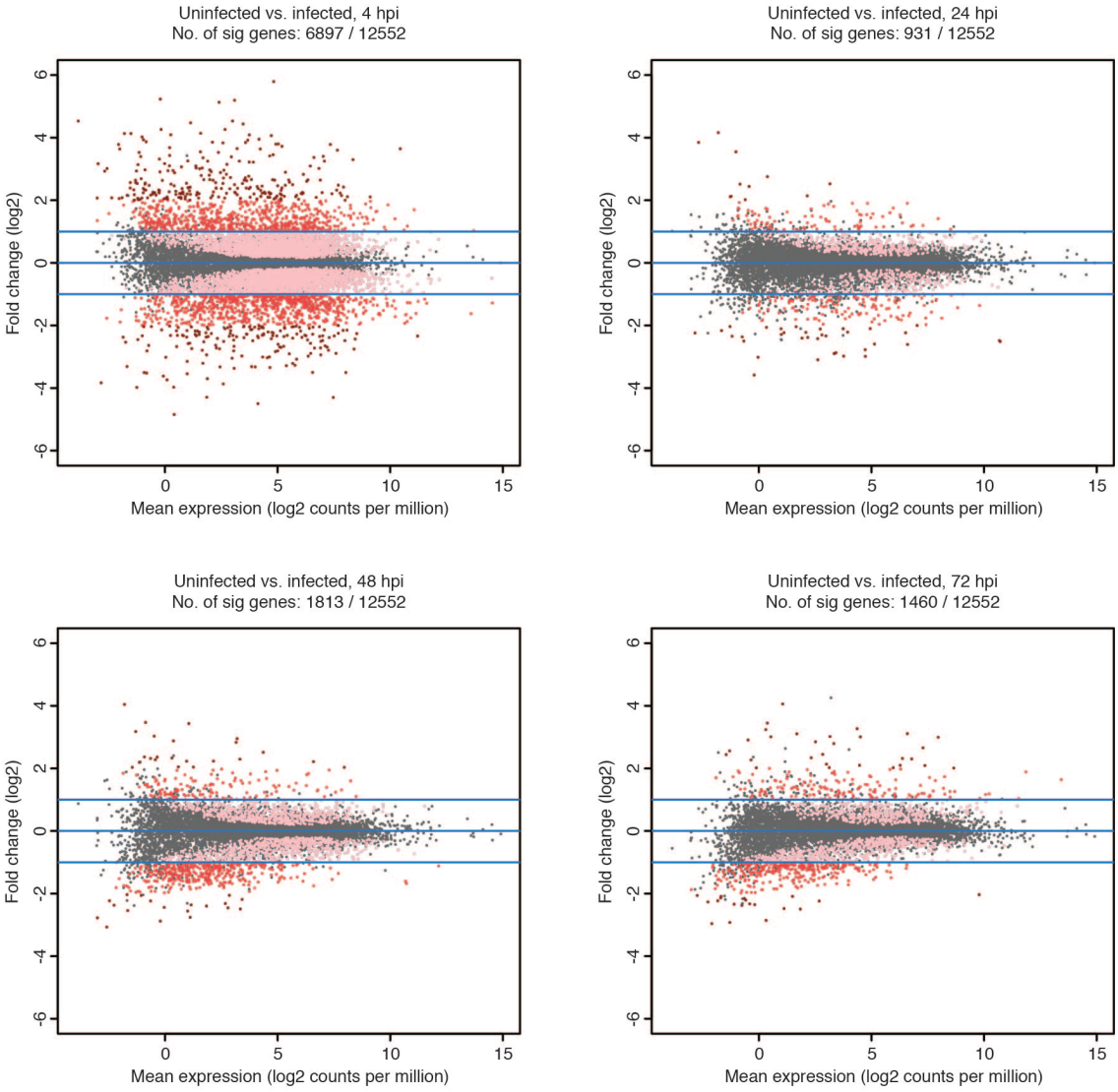


**Figure 16: MA plot of differentially expressed genes in murine macrophages upon infection by *L. major*.**

Differential expression analysis was done to compare infected mouse macrophages and uninfected controls at 4, 24, 48, and 72 hpi using `limma` after `voom` transformation, taking experimental batch into account as part of the `limma` statistical model. The MA plots show the relationship between mean expression (log2 counts per million with an offset of 0.5) and fold change (log2) for each timepoint. Each point represents one gene. Genes upregulated in infected samples relative to uninfected samples exhibit positive fold changes, while downregulated genes exhibit negative fold changes. Points colored in gray represent genes that were not significantly different between uninfected and infected macrophages (*P* value < 0.05) while points colored in shades of red represent significant genes, with those showing a < 2-fold difference (logFC < 1) colored in pink, between 2 to 4-fold difference colored in red, and > 4-fold difference colored in burgundy.

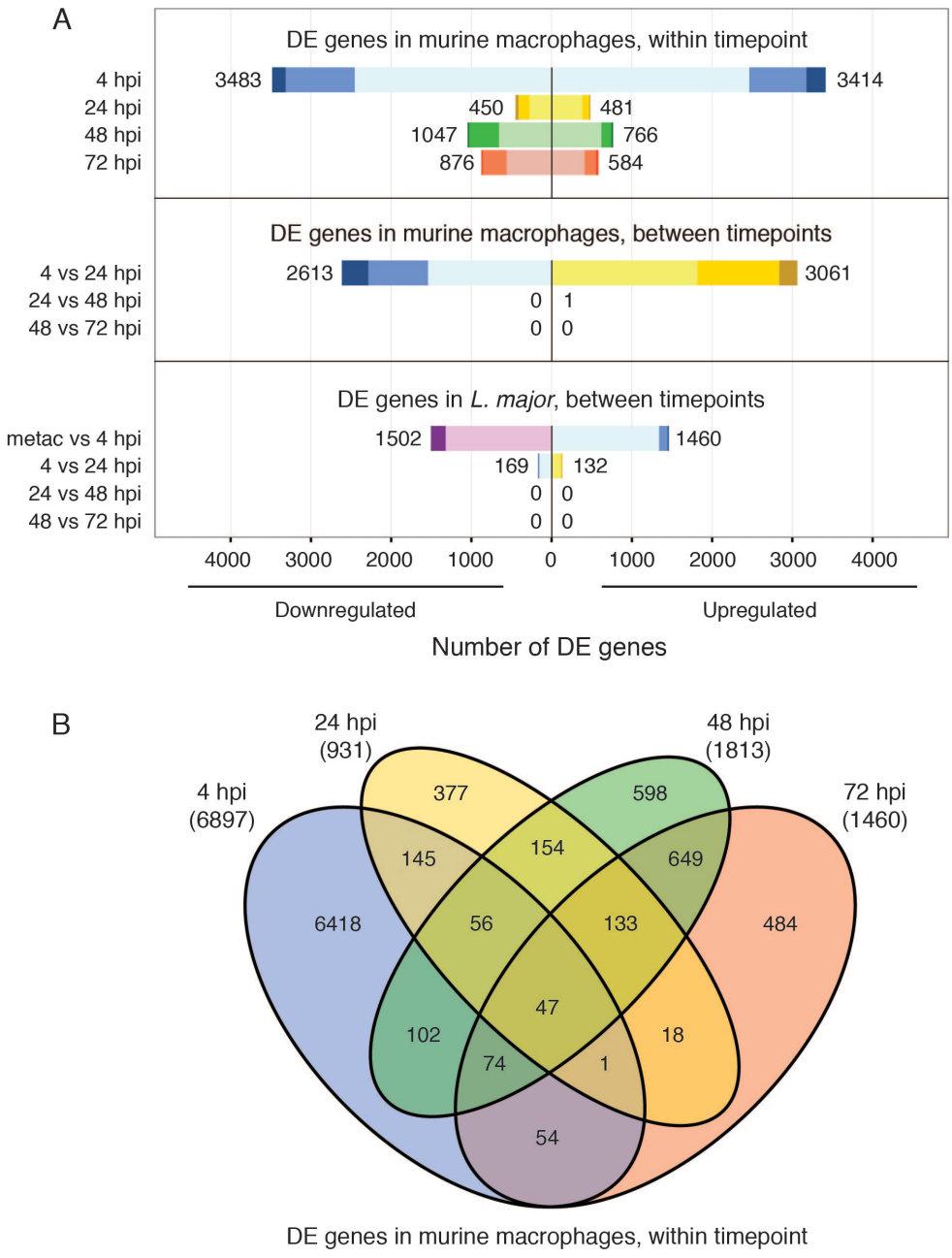


**Figure 17: Differentially expressed genes in L. major parasites and their murine macrophage host cells.**

Pairwise comparisons were done to determine differentially expressed (DE) genes from uninfected vs. infected mouse samples at each timepoint (A, top) and between timepoints (A, middle), and for *L. major* parasite samples between timepoints (A, bottom). Box length depicts the number of DE genes either downregulated (left) or upregulated (right) at an adjusted *P* value of < 0.05 with the total number of down- and up-regulated genes shown. Color hue indicates sample type/timepoint as defined in Figure 14 and Figure 15 and color shade indicates the proportion of genes with > 4-fold differential expression (dark), between 2- and 4-fold differential expression (medium), or 2-fold differential expression (light). The DE gene lists for uninfected vs. infected mouse cells at each timepoint were compared and the overlap shown as a Venn diagram in (B).

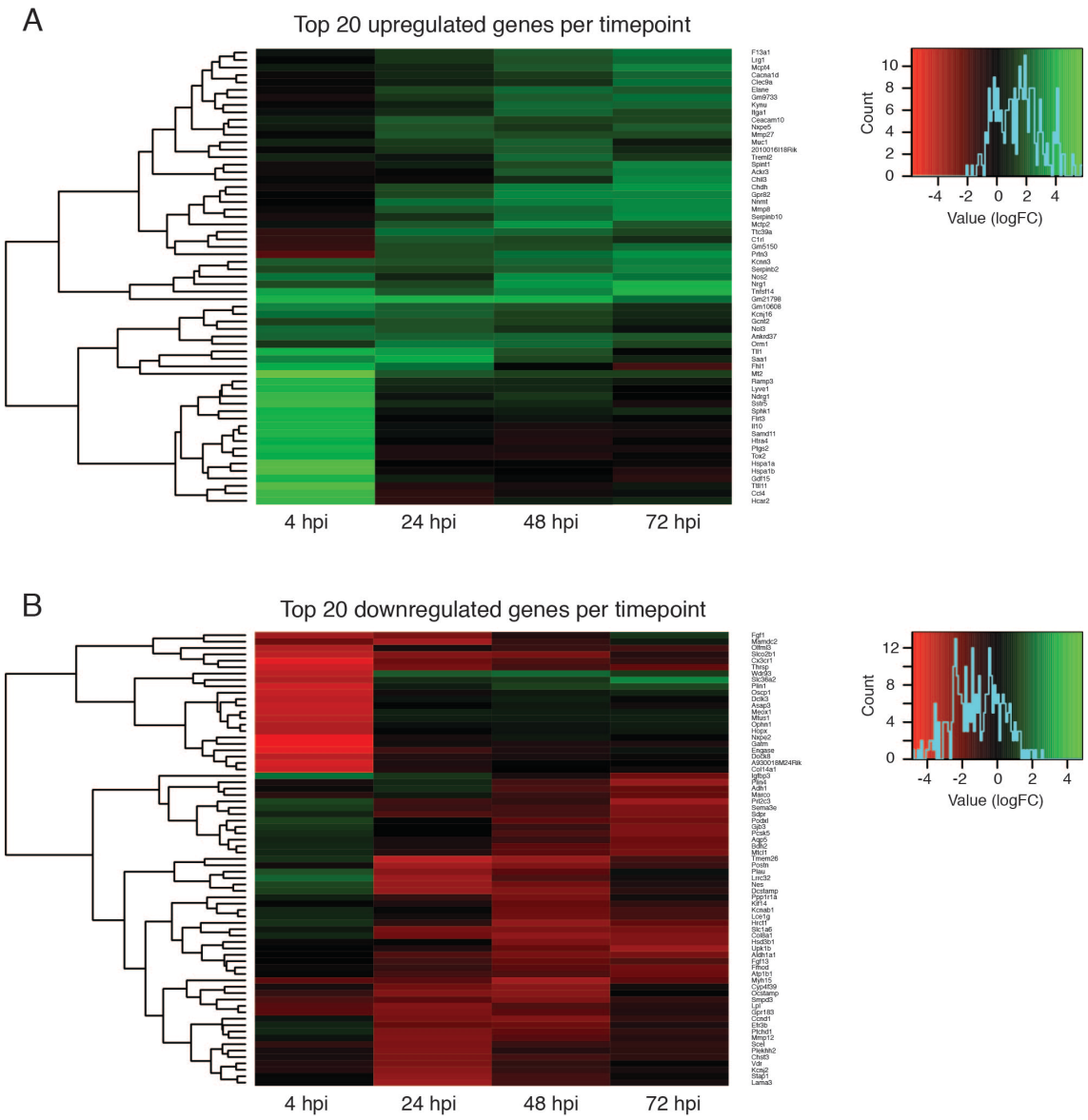


**Figure 18: Expression patterns for the top up- and down-regulated genes in *L. major*-infected murine macrophages.**

A heatmap is used to illustrate the pattern of changes in gene expression over time for the top differentially expressed genes in uninfected vs. *L. major*-infected murine macrophages. The top 20 significantly up-regulated (A) and down-regulated (B) genes at each timepoint (*P* value < 0.05) were selected for inclusion. A color key and histogram for the frequency of fold change values is shown with each panel.

In order to detect statistically significant differences in gene expression over time, we conducted differential expression analysis across timepoints. Contrasts between successive timepoints revealed a large number of DE genes

during the 4 to 24 hpi transition (5674 DE genes), but almost no statistically significant genes between 24 and 48 hpi (1 gene) or between 48 and 72 hpi (0 genes) (Figure 17A middle and Figure 19). This suggests that the large initial response of the murine macrophages to *L. major* infection reached a plateau by 24 hpi, which persisted for the remainder of the time course as the infected cells mostly maintain their expression profile relative to matched uninfected cells.

The cellular processes most affected at each timepoint were characterized by KEGG pathway enrichment analysis using ConsensusPathDB (Kamburov, 2011). Genes that were DE >2-fold were used as input with upregulated and downregulated genes considered separately. The results of this analysis are reported in Table 4, Table 5, and Table 6.

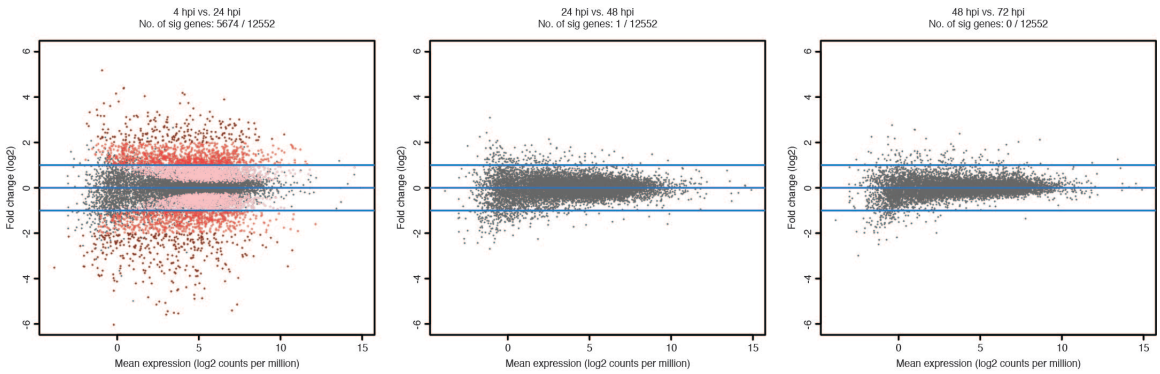


**Figure 19: MA plot of differentially expressed genes in murine macrophages over time upon infection by *L. major*.**

Differential expression analysis was done to compare changes in *L. major*-infected mouse macrophages relative to uninfected controls over the time course of the experiment (4, 24, 48, and 72 hpi). Comparisons were done using `limma` after `voom` transformation, taking experimental batch into account as part of the `limma` statistical model. The MA plots show the relationship between mean expression (log2 counts per million with an offset of 0.5) and fold change (log2). Each point represents one gene. Genes upregulated in the second of the specified timepoints relative to the first are shown as exhibiting positive fold changes, while downregulated genes exhibit negative fold changes. Points colored in gray represent genes that were not significantly different between the tested conditions (*P* value < 0.05) while points colored in shades of red represent significant genes, with those showing a < 2-fold difference (logFC < 1) colored in pink, 2 to 4-fold difference colored in red, and > 4-fold difference colored in burgundy.

The enriched KEGG pathways were considered alongside the DE gene lists to gain insights into the cellular response to infection and how it differs between early and later infection. Early in the infection (4 hpi), many of the KEGG pathways that are most highly induced in infected macrophages are related to immune responses, specifically cytokine-cytokine receptor interactions and arginine and proline metabolism, glycolysis, and multiple signaling pathways including those for TNF, HIF-1, NF-kappa B, Jak-STAT, PI3K-Akt, and MAPK. A closer look at the DE genes in the immune response-related pathways reveals an interesting picture of the murine macrophage response to infection at 4 hpi. Infected macrophages produce a set of transcripts with paradoxical functions involved in both activating immune responses and promoting tissue regeneration and wound healing. Many of the DE genes associated with enriched pathways are anti-inflammatory in character or are involved in tissue growth and repair, including *Csf1*, *Csf3*, *Il10*, *Il11r*, *Il1rn*, *Socs3*, *Hmox1*, *Egfr*, *Vegf*, and fos-induced growth factor (*Figf*). The product of each of these genes has either been associated with reducing inflammation or promoting cell survival or differentiation. Therefore, macrophages infected with *L. major* appear to assume an immunoregulatory phenotype that resembles previously described macrophages treated with LPS and immune complex (Fleming, 2015). However, not all of the differentially induced genes in infected macrophages at 4 hpi were anti-inflammatory. There was also an upregulation of transcripts encoding well-described inflammatory cytokines and their receptors including *Il1*, *Il6*, *Tnf*, and *Nos2*, as well as *Il1rap* and *Il18r1*. The pro-inflammatory NOD-like receptor

| Upregulated genes | | | |
|---|---|---|---|
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Cytokine-cytokine receptor interaction | 43 | 266 | 3.49E-12 |
| TNF signaling pathway | 24 | 111 | 3.05E-09 |
| Transcriptional misregulation in cancer | 24 | 180 | 5.72E-09 |
| HIF-1 signaling pathway | 20 | 113 | 4.57E-07 |
| Hematopoietic cell lineage | 16 | 87 | 5.84E-07 |
| NF-kappa B signaling pathway | 19 | 102 | 1.55E-06 |
| Jak-STAT signaling pathway | 22 | 155 | 2.24E-06 |
| Legionellosis | 13 | 59 | 1.05E-05 |
| PI3K-Akt signaling pathway | 37 | 353 | 6.29E-05 |
| MAPK signaling pathway | 28 | 257 | 2.57E-04 |
| Salmonella infection | 12 | 79 | 9.24E-04 |
| Malaria | 9 | 49 | 1.00E-03 |
| NOD-like receptor signaling pathway | 10 | 59 | 1.03E-03 |
| Rheumatoid arthritis | 12 | 84 | 1.60E-03 |
| Glycolysis / Gluconeogenesis | 9 | 65 | 2.19E-03 |
| Mineral absorption | 8 | 46 | 2.72E-03 |
| ECM-receptor interaction | 11 | 87 | 6.42E-03 |
| Tuberculosis | 18 | 177 | 6.45E-03 |
| Gap junction | 11 | 88 | 7.00E-03 |
| Axon guidance | 14 | 129 | 8.88E-03 |
| Arginine and proline metabolism | 8 | 56 | 9.30E-03 |
| Prion diseases | 6 | 35 | 9.77E-03 |
| **Downregulated genes** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Osteoclast differentiation | 21 | 126 | 2.99E-06 |
| Terpenoid backbone biosynthesis | 7 | 21 | 7.30E-05 |
| Staphylococcus aureus infection | 11 | 52 | 7.61E-05 |
| Steroid biosynthesis | 6 | 17 | 1.71E-04 |
| NOD-like receptor signaling pathway | 10 | 59 | 2.51E-04 |
| Peroxisome | 12 | 81 | 1.18E-03 |
| PPAR signaling pathway | 12 | 81 | 1.18E-03 |
| Chagas disease (American trypanosomiasis) | 14 | 104 | 1.24E-03 |
| Biosynthesis of unsaturated fatty acids | 6 | 25 | 1.69E-03 |
| Leishmaniasis | 10 | 66 | 2.50E-03 |
| ABC transporters | 8 | 46 | 2.75E-03 |
| Axon guidance | 15 | 129 | 3.69E-03 |
| Leukocyte transendothelial migration | 14 | 121 | 5.16E-03 |
| Fc gamma R-mediated phagocytosis | 11 | 89 | 7.70E-03 |
| Pancreatic cancer | 9 | 67 | 9.02E-03 |

**Table 4: KEGG pathways enriched in *L. major*-infected murine macrophages at 4 hpi.**

KEGG pathway analysis using ConsensusPathDB (Kamburov, 2011) identified signaling and metabolic pathways that were over-represented in *L. major*-infected mouse macrophages at 4 hpi (*P* value < 0.01) relative to uninfected controls. Genes that were differentially expressed (DE) by more than 2-fold were used as input with up- and down-regulated genes considered separately. For each enriched KEGG pathway, the number of DE genes assigned to that pathway, the total number of genes in the pathway, and the *P* value for the enrichment are reported.

signaling pathway is enriched among both up-regulated and down-regulated genes at 4 hpi with some members of the pathway upregulated (*Tnf*, *Il6*, *Il1β*, *Cxcl2*) and others downregulated (NLRP-family members, *Pycard*, *Card9*, and *Il18*). It is unclear to what extent the observed expression patterns are being driven by the parasite as it is taken into the host cell, or are being promoted by the host as it attempts to limit infection or to protect itself from the negative effects of a robust immune response.

Pathway analysis also showed an upregulation of glycolysis/ gluconeogenesis at 4 hpi. Upregulated genes encode glycolytic enzymes, including hexokinases, enolase, phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase, and lactate dehydrogenase A. An increase in anaerobic glycolysis has been associated with the stimulation of inflammatory responses in macrophages following toll-like receptor (TLR) ligation (Tannahill, 2013). Glycolytic pathways were also enriched among upregulated genes at 24 hpi, but this effect was not seen at 48 and 72 hpi. Thus, following phagocytosis of *L. major*, macrophages undergo transient metabolic alterations that result in an increase in glycolysis and the generation of ATP.

KEGG pathways downregulated at 4 hpi are mostly related to lipid metabolism and biogenesis and include osteoclast differentiation, terpenoid backbone biosynthesis, steroid biosynthesis, PPAR signaling, and biosynthesis of unsaturated fatty acids. Decreases in the receptors and signaling molecules involved in the process of phagocytosis itself (“Fc gamma R-mediated phagocytosis” KEGG pathway) were also observed at 4 hpi. Transcripts

encoding all four Fcγ receptors (*Fcgr1*, *Fcgr2b*, *Fcgr3*, and *Fcgr4*), complement component 1 q subcomponent polypeptides (*C1q* family), and the C5a receptors (*C5ar1* and *C5ar2*) were downregulated following *L. major* phagocytosis. It has been previously shown that phagocytosis of *Leishmania* and other pathogens via Fc gamma receptors leads to parasite killing while phagocytosis via pathways mediated by complement receptor 3 (CR3) leads to parasite survival (Babadjanova, 2015; Da Silva, 1989; Fadul, 1995; Mosser, 1987; Woelbing, 2006). The downregulation of Fc gamma receptors may be a mechanism by which the parasite induces changes in host cells to promote alternative entry mechanisms that will support its survival.

An examination of the enriched KEGG pathways at 24, 48, and 72 hpi reveals different enrichment patterns from those observed at 4 hpi (Table 5 and Table 6). There are a greater number of pathways downregulated than upregulated during each of these later timepoints, but we see no clear picture that otherwise emerges.

While a significant number of disease-specific KEGG pathways are enriched among up-and down-regulated genes reported at various timepoints in this study (i.e., leishmaniasis, Chagas disease, tuberculosis, malaria, legionellosis, amoebiasis, *Salmonella* infection, *Staphylococcus aureus* infection, Fanconi anemia, prion diseases, and cancer subtypes; see Table 4, Table 5, and Table 6), a linkage between these disease and our dataset may not be particularly meaningful. This is because the KEGG pathways for these individual diseases appear to be constructed using a combination of heterogeneous

| Upregulated genes, 24 hpi | | | |
|---|---|---|---|
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Glycolysis / Gluconeogenesis | 4 | 65 | 2.88E-06 |
| Chemical carcinogenesis | 6 | 95 | 2.62E-05 |
| Metabolism of xenobiotics by cytochrome P450 | 6 | 96 | 2.78E-05 |
| Drug metabolism - cytochrome P450 | 6 | 97 | 2.95E-05 |
| Starch and sucrose metabolism | 4 | 54 | 3.44E-04 |
| Glutathione metabolism | 4 | 55 | 3.70E-04 |
| Staphylococcus aureus infection | 3 | 52 | 4.11E-03 |
| HIF-1 signaling pathway | 2 | 113 | 5.33E-03 |
| Natural killer cell mediated cytotoxicity | 2 | 119 | 6.40E-03 |
| Osteoclast differentiation | 4 | 126 | 7.80E-03 |
| **Upregulated genes, 48 hpi** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Osteoclast differentiation | 8 | 126 | 9.58E-07 |
| **Upregulated genes, 72 hpi** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Osteoclast differentiation | 10 | 126 | 4.94E-08 |
| Staphylococcus aureus infection | 5 | 52 | 1.68E-04 |
| Calcium signaling pathway | 8 | 183 | 5.06E-04 |
| Chemokine signaling pathway | 8 | 199 | 8.78E-04 |
| Hematopoietic cell lineage | 5 | 87 | 1.81E-03 |
| Nicotinate and nicotinamide metabolism | 3 | 32 | 3.99E-03 |
| Leishmaniasis | 4 | 66 | 4.34E-03 |
| Amoebiasis | 5 | 120 | 7.19E-03 |
| Salivary secretion | 4 | 77 | 7.49E-03 |
| Complement and coagulation cascades | 4 | 77 | 7.49E-03 |
| Tuberculosis | 6 | 177 | 8.87E-03 |

**Table 5: KEGG pathways enriched among genes upregulated in *L. major*-infected murine macrophages at 24, 48, and 72 hpi.**

KEGG pathway analysis using ConsensusPathDB identified signaling and metabolic pathways that were over-represented among genes upregulated > 2-fold in *L. major*-infected mouse macrophages at 24, 48, and 72 hpi (*P* value < 0.01) relative to uninfected controls. For each enriched KEGG pathway, the number of differentially expressed (DE) genes belonging to that pathway, the total number of genes in the pathway, and the *P* value for the enrichment are reported.

observations from multiple studies of individual genes using data from different experiments, hosts, cell types, and timepoints, rather than using gene lists derived from a single or set of genome-wide studies, and as a result paint an incomplete picture of the global changes in gene expression. Therefore, while it is interesting to note small overlaps between the comprehensive DE profiles in

| Downregulated genes, 24 hpi | | | |
|---|---|---|---|
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Transcriptional misregulation in cancer | 10 | 180 | 2.10E-05 |
| Pathways in cancer | 13 | 326 | 4.09E-05 |
| Focal adhesion | 10 | 206 | 6.63E-05 |
| Melanoma | 6 | 73 | 1.21E-04 |
| PI3K-Akt signaling pathway | 12 | 353 | 3.64E-04 |
| Proteoglycans in cancer | 9 | 228 | 7.14E-04 |
| Glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulf | 3 | 20 | 1.17E-03 |
| Small cell lung cancer | 5 | 86 | 2.20E-03 |
| Thyroid cancer | 3 | 29 | 3.51E-35 |
| Wnt signaling pathway | 6 | 143 | 4.18E-03 |
| Colorectal cancer | 4 | 64 | 4.74E-03 |
| Hippo signaling pathway | 6 | 156 | 6.36E-03 |
| Amoebiasis | 5 | 120 | 9.04E-03 |
| **Downregulated genes, 48 hpi** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Cell cycle | 20 | 128 | 7.91E-14 |
| Progesterone-mediated oocyte maturation | 9 | 88 | 4.20E-06 |
| p53 signaling pathway | 8 | 69 | 2.25E-05 |
| Oocyte meiosis | 10 | 112 | 4.28E-05 |
| ECM-receptor interaction | 8 | 87 | 7.47E-04 |
| Focal adhesion | 13 | 206 | 7.82E-04 |
| Pathways in cancer | 17 | 326 | 1.13E-03 |
| Hippo signaling pathway | 10 | 156 | 2.78E-03 |
| Small cell lung cancer | 7 | 86 | 3.21E-03 |
| Homologous recombination | 4 | 28 | 3.37E-03 |
| HTLV-I infection | 13 | 279 | 4.30E-03 |
| Amoebiasis | 8 | 120 | 5.71E-03 |
| Fanconi anemia pathway | 5 | 52 | 6.12E-03 |
| **Downregulated genes, 72 hpi** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| ECM-receptor interaction | 15 | 87 | 1.08E-10 |
| Focal adhesion | 20 | 206 | 3.31E-09 |
| PI3K-Akt signaling pathway | 21 | 353 | 5.47E-06 |
| Pathways in cancer | 20 | 326 | 5.86E-06 |
| Protein digestion and absorption | 10 | 88 | 7.90E-06 |
| p53 signaling pathway | 7 | 69 | 5.63E-05 |
| Cell cycle | 9 | 128 | 2.02E-04 |
| Small cell lung cancer | 8 | 86 | 2.69E-04 |
| Amoebiasis | 9 | 120 | 5.66E-04 |
| Oocyte meiosis | 8 | 112 | 1.57E-03 |
| Progesterone-mediated oocyte maturation | 6 | 88 | 1.65E-03 |
| Hippo signaling pathway | 9 | 156 | 3.53E-03 |
| Tyrosine metabolism | 4 | 43 | 9.68E-03 |

**Table 6: KEGG pathways enriched among genes downregulated in *L. major*-infected murine macrophages at 24, 48, and 72 hpi.**

KEGG pathway analysis using ConsensusPathDB identified signaling and metabolic pathways that were over-represented among genes downregulated > 2-fold in *L. major*-infected mouse macrophages at 24, 48, and 72 hpi (*P* value < 0.01) relative to uninfected controls. For each enriched KEGG pathway, the number of differentially expressed (DE) genes belonging to that pathway, the total number of genes in the pathway, and the *P* value for the enrichment are reported.

this study and genes included in these pathways, the differences in scope between the lists limit the usefulness of comparisons to disease KEGG pathways as currently defined.

We compared the results of our differential expression analysis to previously published reports on the murine macrophage response to *Leishmania* infection. Since global transcriptome studies of *L. major* infection of macrophages isolated from C57BL/6 mice are not available, we compared our results to studies that used either the same mouse strain or the same *Leishmania* species. In a study detailing the infection of C57BL/6 peritoneal macrophages with *L. amazonensis* (Probst, 2012), 105 genes were identified as DE (FC > 1.5) between infected and uninfected macrophages at 18 hpi. This number compares to 3972 and 546 DE genes (FC > 1.5) at 4 and 24 hpi, respectively, in our dataset with an overlap of only 42 genes (of the 105) differentially expressed in the same direction at 4 or 24 hpi. Similarly, we compared the infection of Balb/c bone marrow-derived macrophages with *L. major* as reported in two recent studies. Of the 769 genes identified when comparing the transcriptomes of *L. major*-infected Balb/c macrophages to macrophages that had ingested latex beads (24 hpi timepoint), only 53 were in common with our 24 hpi timepoint (Gregory, 2008). However, 446 genes were in common with our 4 hpi timepoint, indicating that the time progression observed between the two systems may be somewhat offset. Of the 40 genes identified as DE by more than 1.5-fold during the 24 hr timecourse experiment reported by Rabhi *et al*. (Rabhi, 2012), 29 genes were also found to be differentially expressed in the same direction in our

dataset at either 4 or 24 hpi with the greatest number of genes found in common with our 4 hpi dataset (26 genes). The limited agreement between the previously reported DE genes and our dataset is likely a reflection of the differences in study design (e.g., combination of mouse strain and parasite species, macrophage source, parasite opsonization), batch effects between laboratories, technical platforms, and data analysis methods.

**Differential expression and gene ontology enrichment analyses in *L. major***

While past studies have used microarrays and other methods to study changes in *Leishmania* gene expression between promastigotes and amastigotes (Akopyants, 2004; Almeida, 2004; Depledge, 2009; Holzer, 2006; Leifso, 2007; McNicoll, 2006; Quijada, 2005; Rochette, 2008; Saxena, 2007; Walker, 2006), this study provided a unique opportunity to additionally characterize gene expression patterns within amastigotes over the course of an intracellular infection. Differential expression analysis revealed a large number of genes (2962) to be DE between metacyclic promastigotes and 4-hr amastigotes at an adjusted *P* value cutoff of <0.05, reflecting the significant differentiation event that occurs as the parasite enters host cells. Significantly fewer changes in expression were observed in amastigotes across timepoints, with 301 DE genes found between 4- and 24-hr amastigotes and no DE genes between either 24- and 48-hr amastigotes or between 48- and 72-hr amastigotes (Figure 17 and Figure 20). This pattern of differential gene expression is consistent with observations of gene expression over time in the mouse macrophages and

suggests that the initial reprogramming that occurs upon infection had largely been stabilized by 24 hpi.

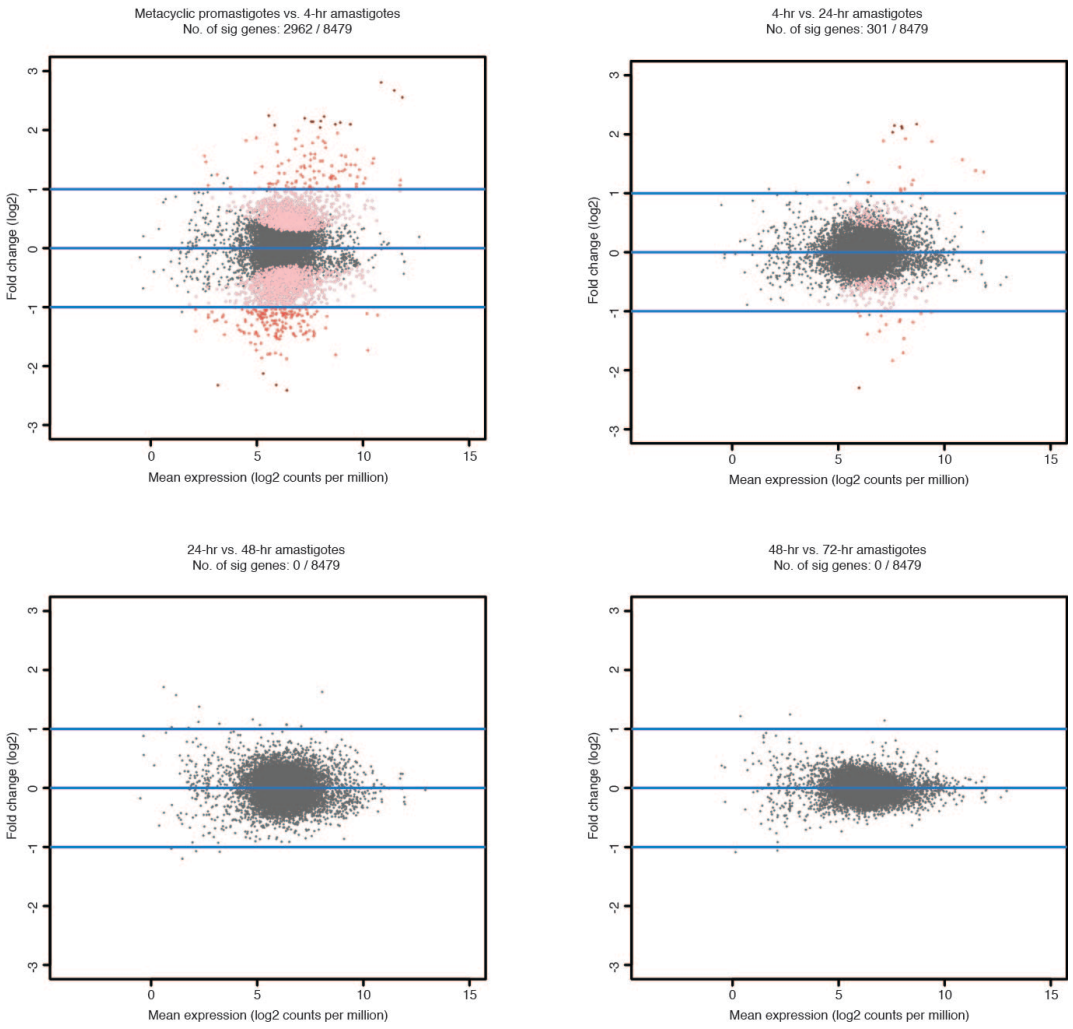


**Figure 20: MA plot of differentially expressed genes in *L. major* across developmental stages and intracellular timepoints.**

Differential expression analysis was done to compare in *L. major* metacyclic promastigotes and intracellular amastigote gene expression over time during a murine macrophage infection (4, 24, 48, and 72 hpi). Comparisons were done using `limma` after `voom` transformation, taking experimental batch into account as part of the `limma` statistical model. The MA plots show the relationship between mean expression (log2 counts per million with an offset of 0.5) and fold change (log2). Each point represents one gene. Genes upregulated in the second of the specified timepoints relative to the first are shown as exhibiting positive fold changes, while downregulated genes exhibit negative fold changes. Points colored in gray represent genes that were not significantly different between the tested conditions (*P* value < 0.05) while points colored in shades of red represent significant genes, with those showing a < 2-fold difference (logFC < 1) colored in pink, 2 to 4-fold difference colored in red, and > 4-fold difference colored in burgundy.

Gene ontology (GO) analysis was used to identify cellular functions and processes that characterize the entry and survival of *L. major* in the murine macrophage system. These results were considered alongside the lists of DE genes to draw insights into how the parasite adapts to its new environment. Efforts were made to compare the results of this study to previous work, though our ability to conduct meaningful systematic comparisons was constrained due to the use of different parasite systems (e.g., *Leishmania* species, source of amastigotes, developmental stages and timepoints collected), host systems (e.g., mice strains, cell lines), experimental platforms (e.g., microarrays, number of genes interrogated), and methods for assessing and reporting differential expression (fold change vs. statistical cut-offs; details of reported gene lists).

GO analysis of the up- and down-regulated genes identified during the metacyclic promastigote to 4-hr amastigote transition revealed a total of 20 enriched GO categories. Enriched GO terms for genes upregulated during this transition were primarily related to mitigating the effects of an oxidative response by the host immune system and the regulation of proteins (Table 7). Heat shock proteins, especially HSP83, multiple tryparedoxin peroxidase family members, and multiple cyclophilins were upregulated upon entry of metacyclic promastigotes into host cells and contributed strongly to the GO enrichment results for the upregulated genes. HSP83, the cytoplasmic form of HSP90, is known to stabilize transcription factors and kinases and is thus largely involved in the regulation of signaling networks involved in differentiation (Hombach, 2014; Requena, 2015), and tryparedoxin peroxidase is known to contribute to

| GO ID | GO term | *P* value |
|---|---|---|
| **Metacyclics to 4-hr amastigotes, upregulated** | | |
| GO:0006950 | response to stress | 5.50E-13 |
| GO:0006486 | protein glycosylation | 9.95E-09 |
| GO:0006457 | protein folding | 5.70E-06 |
| GO:0051920 | peroxiredoxin activity | 2.72E-05 |
| GO:0051082 | unfolded protein binding | 9.44E-05 |
| GO:0016209 | antioxidant activity | 1.04E-04 |
| GO:0004386 | helicase activity | 2.06E-04 |
| **Metacyclics to 4-hr amastigotes, downregulated** | | |
| GO:0004713 | protein tyrosine kinase activity | 6.68E-12 |
| GO:0004674 | protein serine/threonine kinase activity | 2.15E-11 |
| GO:0004672 | protein kinase activity | 3.40E-10 |
| GO:0006468 | protein phosphorylation | 3.43E-10 |
| GO:0005516 | calmodulin binding | 2.76E-08 |
| GO:0009434 | microtubule-based flagellum | 1.93E-07 |
| GO:0005840 | ribosome | 8.85E-07 |
| GO:0006412 | translation | 1.42E-06 |
| GO:0003735 | structural constituent of ribosome | 1.81E-06 |
| GO:0006633 | fatty acid biosynthetic process | 2.72E-05 |
| GO:0009190 | cyclic nucleotide biosynthetic process | 8.29E-05 |
| GO:0016849 | phosphorus-oxygen lyase activity | 8.29E-05 |
| GO:0007165 | signal transduction | 2.66E-04 |
| **4-hr to 24-hr amastigotes, upregulated** | | |
| GO:0005874 | microtubule | 3.25E-09 |
| GO:0043234 | protein complex | 5.80E-09 |
| GO:0051258 | protein polymerization | 5.80E-09 |
| GO:0005198 | structural molecule activity | 3.32E-08 |
| GO:0006184 | GTP catabolic process | 3.91E-07 |
| GO:0007018 | microtubule-based movement | 9.88E-06 |
| GO:0003924 | GTPase activity | 1.81E-05 |
| **4-hr to 24-hr amastigotes, downregulated** | | |
| GO:0055114 | oxidation-reduction process | 8.81E-08 |
| GO:0006096 | glycolytic process | 1.24E-07 |
| GO:0020015 | glycosome | 4.22E-07 |
| GO:0004743 | pyruvate kinase activity | 1.50E-05 |
| GO:0050661 | NADP binding | 2.36E-05 |
| GO:0004365 | GAPDH (NAD+) (phosphorylating) activity | 5.91E-05 |
| GO:0006006 | glucose metabolic process | 1.45E-04 |

**Table 7: Gene ontology categories enriched across *L. major* stages/timepoints.**

`GOseq` (Young, 2010) was used to identify enriched GO categories for the transition between metacyclic promastigotes and 4-hr amastigotes and between 4-hr amastigotes and 24-hr amastigotes at a *P* value cutoff of < 0.05. For each transition, up- and down-regulated genes were considered separately.

*Leishmania* virulence and resistance to antimonial drugs (Iyer, 2008). Less is known about the role of cyclophilins in *Leishmania*, though these peptidyl-prolyl cis/trans isomerases have been shown to reactivate cellular proteins by promoting their disaggregation (Chakraborty, 2002). The upregulation of cyclophilin in *L. major* amastigotes has been previously observed (compared to procyclic promastigotes) (Leifso, 2007). Many types of surface antigens were highly upregulated, as was the zinc-metalloprotease GP63 (a known virulence factor that subverts macrophage signaling) (Olivier, 2012), and phosphoglycan beta 1,3 galactosyltransferase (SCG) family members, which are responsible for modifications to the *Leishmania* lipophosphoglycan (LPG) surface molecule side chains (Dobson, 2003) and have been previously observed to be upregulated in *Leishmania* amastigotes (Rochette, 2008). That there are significant changes taking place on the surface of the parasite is not surprising given the transformation that takes place as differentiation progresses from the promastigote to amastigote forms. Other genes upregulated during the metacyclic promastigote to 4-hr amastigote transition included *meta1,* which is thought to play a role in virulence and was previously found to be upregulated in infective metacyclic promastigotes relative to non-infective procyclic promastigotes (Dillon, 2015; Nourbakhsh, 1996; Puri, 2011), and inositol-3-phosphate synthase (*ino1*), which synthesizes *myo*-inositol, a precursor molecule for the backbone of the GPI anchors used by many *Leishmania* surface proteins, including multiple virulence factors (Ilg, 2002). Interestingly, the gene encoding the transporter for *myo*-inositol (*mit1*) (Drew, 1995) was among the most down-

regulated during the transition, perhaps indicating that the parasite favors synthesizing *de novo* rather than importing this important substrate once it enters mammalian cells.

GO terms that were enriched among downregulated genes during the metacyclic promastigote to 4-hr amastigote transition were related to translation, cell signaling, fatty acid biosynthesis, and flagellum structure (Table 7). A number of genes were responsible for driving the GO results, including ribosomal proteins, casein kinase, receptor-type adenylate cyclase, fatty acid elongase, sphingolipid delta desaturase, and paraflagellar rod protein. The downregulation of ribosomal proteins during the *Leishmania* promastigote to amastigote transition is consistent with previous reports (Almeida, 2004) and may reflect a reduction in translation taking place within the cell. Casein kinase is thought to play a role in *Leishmania* promastigote growth and parasite virulence via interactions with host proteins, though its exact function is unclear (Allocco, 2006; Dan-Goor, 2013). Likewise, adenylate cyclase is a suspected regulator of cAMP signaling during *Leishmania* differentiation (Biswas, 2011; Sanchez, 1995), but the exactly signaling pathways through which this is accomplished are unknown. Fatty acids have many potential roles within the cell related to membrane structure and composition, signaling, and energy. Fatty acid elongase family members are involved in fatty acid synthesis and may be involved in GPI anchoring (Ramakrishnan, 2013), while sphingolipid delta 4 desaturase is involved in the degradation of sphingolipids, a process that has been linked to enabling *Leishmania* to survive the acidic environment of the phagolysosome

once it is taken into host cells (Xu, 2011). Genes that influence microtubule length and dynamics were systematically downregulated, including those that encode paraflagellar rod proteins, calmodulin-related proteins (Ginger, 2013), multiple kinesins (Blaineau, 2007; Vicente, 2015), NIMA-related kinases (Cao, 2009), and mitogen-activated protein (MAP) kinase kinases (Bengs, 2005; Erdmann, 2006; Wiese, 2003). These genes have been implicated in the regulation of flagellum length and mitotic spindle formation and may reflect changes in morphology and cell division that take place as the flagellated promastigote stage parasites transform into the aflagellated amastigotes. Consistent with previous microarray-based studies (Akopyants, 2004; Almeida, 2004), we also detected the downregulation of the known metacyclic markers SHERP and HASP as the parasite transformed from promastigotes to amastigotes.

Of the 14 GO categories enriched during the transition from 4-hr to 24-hr amastigotes, those related to microtubules and protein complexes were specific to upregulated genes with many copies of β-tubulin, the primary component of microtubules, almost exclusively driving the GO analysis results. The upregulation of β-tubulin in early amastigotes compared to metacyclic promastigotes is noteworthy considering its downregulation in metacyclic promastigotes compared to procyclic promastigotes (Coulson, 1996; Dillon, 2015). Besides β-tubulin, many other top upregulated genes encoded surface antigens, including the developmentally-regulated amastigote-enriched protein amastin (Akopyants, 2004; Holzer, 2006; Rochette, 2008; Wu, 2000). GO

categories related to glucose metabolism and oxidation-reduction processes were enriched among downregulated genes with many of the downregulated genes contributing to these categories belonging to the glycolysis pathway (GAPDH, pyruvate kinase, ENOL, triosephosphate isomerase, ALD, and glucose-6-phosphate isomerase) or related to cholesterol/sterol metabolism (NSDHL, HMGR, FPPS, CYP51) or purine metabolism (XRPT). This may indicate a change in the metabolic requirements or preferences of *Leishmania* once the parasites have survived their initial entry into the mammalian cell. Other highly downregulated genes included heat shock proteins, biopterin transporters, and *meta1* (Dillon, 2015; Nourbakhsh, 1996; Puri, 2011).

A very significant dataset resulting from this study includes the large proportion of the DE genes that have no known function and are annotated as hypothetical proteins. Those make up 58% (1724 of 2962) of the DE genes in the metacyclic to 4-hr amastigote transition and 49% (147 of 301) of the DE genes in the 4- to 24-hr amastigote transition. While hypothetical genes have largely been overlooked to date, they almost certainly constitute an integral part of the transcriptomic signature of the parasites as members of co-expressed gene networks which are involved in common functional pathways or regulated by shared regulatory mechanisms.

**Conclusion**

In this study, we performed transcriptomic profiling to identify *L. major* and mouse genes that were differentially expressed as the parasite entered and persisted within murine macrophages. The generation of RNA-seq data followed

by the unambiguous mapping of reads from infected samples to the genomes of both the host and the parasite enabled us to identify genes that changed over the first 72 hours of an infection in the real-time context of a dynamic dual biological system, and to a much greater depth and sensitivity than has been previously reported. Collection of data from multiple biological replicates, the use of matched host control samples, careful statistical analysis of variation, and removal of batch effects provided us with a unique ability to detect biological differences between samples and timepoints with high confidence and sensitivity.

General patterns of transcription in both the mouse and the parasite revealed robust transcriptional responses early in infection but suggested a stabilization in the transcriptional response by around 24 hpi, with few genes showing statistically significant changes in expression levels after that point. At 4 hpi, host pathways related to pro- and anti-inflammatory immune signaling and interactions and metabolism were upregulated while pathways related to lipid metabolism and biogenesis were downregulated. The parasite showed upregulation of genes involved in antioxidant responses and the regulation of proteins and downregulation of genes related to translation, cell signaling, fatty acid biosynthesis, and flagellum structure. In addition to providing robust sets of markers for multiple developmental stages of *L. major* parasites and *Leishmania*-infected macrophages over several timepoints, this work contributes to a growing body of literature on the broader field of host-pathogen interactions.

# Chapter 5

## Transcriptomic profiling of *Leishmania*-infected human macrophages

### Introduction

*Leishmania* infection causes a wide range of clinical outcomes, with some species causing cutaneous, mucocutaneous, or diffuse cutaneous leishmaniasis where symptoms remain localized to the skin or mucosal surfaces, and other *Leishmania* species causing visceral leishmaniasis after migration to internal organs such as the liver, spleen, and bone marrow. Propensity for rapid self-cure, dissemination, persistence, latency, and reactivation include factors related to the species of the parasite, and also to the host's acquired and innate immunities (Herwaldt, 1999; Murray, 2005). Despite the striking variability in pathogenicity and tissue tropism of different *Leishmania* species, their genomes are remarkably similar, displaying a high degree of conservation in gene content and synteny (Ivens, 2005; Peacock, 2007; Real, 2013; Rogers, 2011).

Macrophages had been widely regarded as primary host cells of *Leishmania* parasites, although recent studies have established that neutrophils (Laskay, 2003; Peters, 2008) and dendritic cells (Charmoy, 2010; León, 2007; Ng, 2008) can also be infected, acting therefore as potentially important players in the host-pathogen interplay. Nonetheless, it is within the mononuclear phagocytes - a hostile environment that is lethal to other microbes - that there is

the most documented evidence for the parasite's replication and long-term survival (Bogdan, 1996).

A limited number of studies have been performed to determine transcriptional changes that occur within macrophages harboring different pathogens including *Mycobacterium tuberculosis, Mycobacterium bovis*, *Mycobacterium avium*, *Yersinia enterocolitica, Escherichia coli*, *Legionella pneumophila* and *Leptospira interrogans* (Bent, 2015; Blumenthal, 2005; Kabara, 2010; Magee, 2012; Mavromatis, 2015; Nalpas, 2013; Price, 2014; Subbian, 2015; Xue, 2013). While an inflammatory response is reported in all cases, the availability of partial gene lists of differentially expressed genes hinders the comparison of global profiles and the identification of specific response(s) to these pathogens. These studies stand in contrast to a number of studies on *Leishmania*-infected macrophages which report that macrophages are impaired in their response to INF-γ and attribute chronic infections to macrophage deactivation (Forget, 2001; Kwan, 1992; Nandan, 1995; Nandan, 1999; Olivier, 1992; Reiner, 1994). Overall results from recent transcriptome analyses of human or murine macrophages infected by different species of *Leishmania* and generated mostly using microarray platforms indicate that infected cells down-modulate the expression of many pro-inflammatory genes and up-regulate the expression of several genes implicated in anti-inflammatory responses (Chaussabel, 2003; Ettinger, 2008; Gregory, 2008; Osorio y Fortéa, 2009; Ovalle-Bracho, 2015; Rabhi, 2012; Ramírez, 2012; Rodriguez, 2004). Whether a general suppression of gene expression in infected human macrophages occurs

has yet to be elucidated and the current reports remain contradictory (Buates, 2001; Chaussabel, 2003; Ettinger, 2008; Gregory, 2008; Ramírez, 2012; Rodriguez, 2004). Since data collected to date represent a miscellany of experiments carried out using different macrophage types, host backgrounds, parasite species, and timepoints, and because the interpretations of most results have been focused on a limited set of markers, it is not possible to integrate these findings to describe the comprehensive state of *Leishmania*-infected macrophages.

Here we carried out a study on human macrophages infected with one of two species of *Leishmania* that cause cutaneous leishmaniasis with potentially different clinical manifestations: *L. major* (cutaneous self-healing) and *L. amazonensis* (cutaneous self-healing/cutaneous diffuse). We utilized a well-defined time course to simultaneously and comprehensively interrogate the transcriptome profiles of both the host and the parasites over time. The concurrent collection of samples from macrophages fed latex beads allowed us to discern the pathogen-specific response from a more general response that can be attributed to large particle phagocytosis. By analyzing data from multiple biological replicates and employing careful statistical analyses to account for batch effects, we were able to generate robust datasets that represent the comprehensive signatures of *Leishmania* infection in a human host.

**Study design**

In order to capture the global transcriptional response during the initiation and maintenance of intramacrophage infection by 2 distinct species of

*Leishmania*, the transcriptomes of the parasite and the infected human macrophage were simultaneously profiled using RNA-seq at 4, 24, 48, and 72 hours post-infection (hpi). CD14+ monocytes were obtained from 4 donors and differentiated into macrophages after an 8-day incubation period in CSF-1 (Figure 21). In accordance with the recently proposed standards for macrophage-activation nomenclature (Murray, 2014), we consider these cells to be CSF-1 monocyte-derived macrophages and we refer to them simply as "macrophages" throughput the manuscript. These macrophages were neither polarized nor

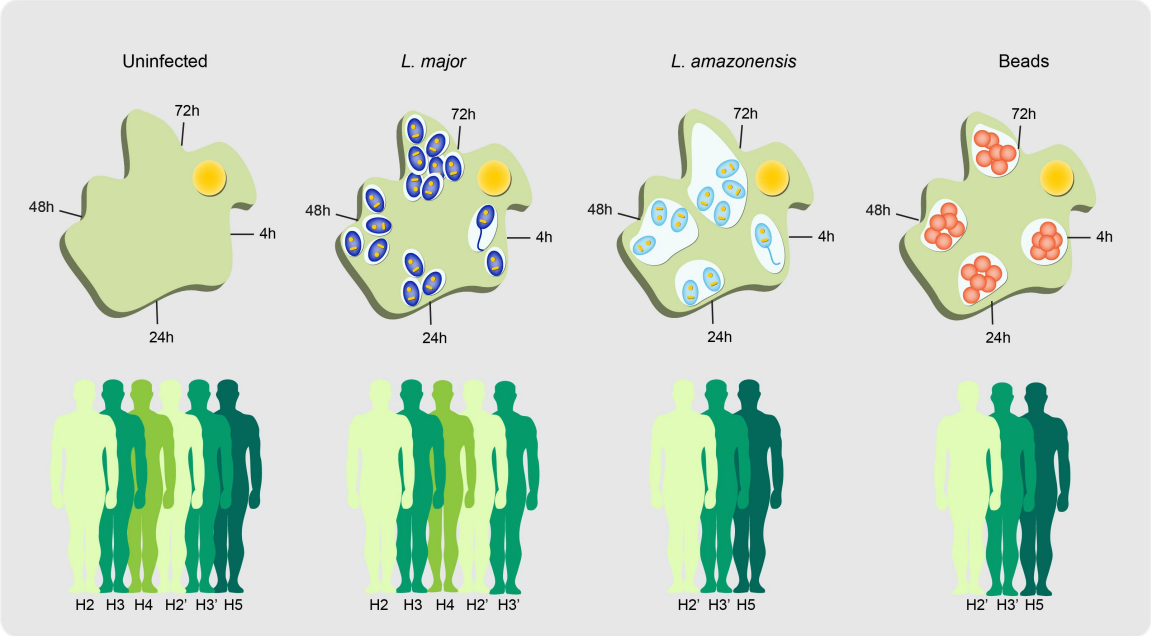


**Figure 21: Study design, donors, and timepoints.**

Monocyte-derived human macrophages were infected with the metacyclic promastigote form of *L. major* (dark blue) or *L. amazonensis* (light blue) parasites. Following phagocytosis, metacyclic promastigotes transform into the amastigote form (3-10 hpi) and reside in membrane bound compartments called phagosomes. As amastigotes divide starting at ~24 hpi, each phagosome matures into a membrane bound singular (*L. major*) or communal (*L. amazonensis*) vacuole inside the macrophage. Samples were collected from infected macrophages, macrophages that were allowed to phagocytose latex beads (red), and uninfected controls over 4 timepoints spanning from 4 to 72 hpi, as pictured. Total RNA was isolated from each sample and analyzed by RNA-seq. A total of 4 biological replicates were collected from 4 different human donors (H2-H5), and are represented in differing shades of green. A second collection from two of the donors was used as a technical replicate within the *L. major*/uninfected dataset (H2' and H3') and for a *L. amazonensis* and latex bead experiment that constituted a later addition to the study design. The experimental design is further detailed in Table 14.

activated with additional cytokine stimulation and were chosen precisely due to their potential for a broad range of responses.

Macrophages from each donor were infected with purified metacyclic promastigotes from *Leishmania major* or *Leishmania amazonensis* in the presence of human serum. In parallel, macrophages were allowed to phagocytose opsonized latex beads to evaluate the effect of inert particle phagocytosis on global gene expression (hereafter referred to as the "phagocytosis effect"). Uninfected macrophages were used as controls. The dynamics of the infection were monitored at each timepoint by determining the number of parasites per 100 macrophages and the percentage of infected cells (Figure 22 and Table 14). Both *Leishmania* species infected macrophages at similar percentages (~90% for *L. major*, ~80% for *L. amazonensis*; Figure 22).

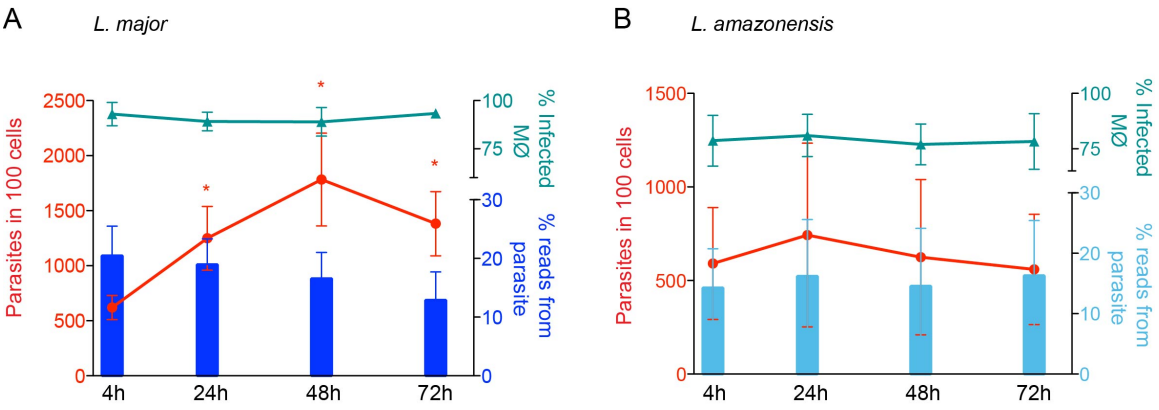


**Figure 22: Dynamics and characterization of CSF1-induced human macrophage infection with *Leishmania*.**

CSF1-induced human macrophages were infected with either *L. major* (A) or *L. amazonensis* (B) for 4 h, washed and further incubated until 24 hpi, 48 hpi and 72 hpi. Samples collected at 4, 24, 48, and 72 hpi were subjected to transcriptional profiling by RNA-seq. The number of internalized parasites and percentage of infected cells were determined microscopically. Bars indicate the percentage of trimmed RNA-seq reads that map to the respective parasite reference genome. Red lines indicate the number of parasites per 100 macrophages and cyan lines indicate % of infected macrophages. The graphs incorporate data from all experiments +/- SD. A star (*) represents a *P* value < 0.05 using a Student's t test to compare each timepoint to 4 hpi.

The overall parasite loads (parasite per infected cell) were slightly higher for *L. major*-infected macrophages (ranging from 6.3 to 18) than for *L. amazonensis*-infected cells (ranging from 6.1 to 8). The intracellular growth rate of each species was evaluated by determining the number of parasites per 100 cells. *L. major* showed a significant increase in growth between 4 hpi and the remaining timepoints and *L. amazonensis* growth remaining somewhat constant. It is not clear whether the different growth behavior of the two parasites can be attributed to an intrinsically different behavior in human cells or simply reflects a lower preference of *L. amazonensis* for the medium used to maintain the human macrophages.

**Global transcriptome profiles of *Leishmania* and its human macrophage host cell**

The global transcriptome profiles of the parasites and host cells were characterized using RNA-seq. PolyA-enriched cDNA libraries were generated for each macrophage sample (uninfected, infected with *L. major* or *L. amazonensis*, or following ingestion of latex beads) at all timepoints as well as for the promastigotes from both species that were used for the infection. Paired-end sequence reads of 100 nucleotides were generated for a total of 74 samples, representing an average of 4 independent biological replicates for each condition and yielding a total of 6.3 billion high-quality reads (see Table 14 for details). Each sample from infected cells consisted of a pool of mixed RNAs from *Leishmania* and human macrophages. To resolve these, the RNA-seq reads generated from these samples were mapped against the corresponding

reference genomes. The fraction of reads mapping to the parasite or human reference genomes yielded an estimate of the proportion of RNA molecules from each source. While the proportion of *L. major* reads in parasite-infected macrophages did not vary significantly across timepoints, the number of intracellular parasites increased over the course of infection (Figure 22A). This may reflect an increase in global transcriptional activity in the human macrophages or a similar decrease in the parasites. The proportion of parasite-attributable reads also remained constant during *L. amazonensis* infection, consistent with the attenuated growth rate for the species.

To investigate general trends in the data, principal component analyses (PCA) were carried out after accounting for batch effects (Figure 23). A high degree of similarity between biological replicates was evident in the PCA plots for both *Leishmania* species and macrophages with similar samples clustering together by infection status/timepoint. We note that the spread of samples along the first principal component (PC1; X-axis) seems to reflect the global transcriptional changes in the cells over time (Figure 23A). Uninfected control macrophages showed notable variance, denoting a drift in the transcriptome as the cells were maintained in culture and thus validating the use of a matched control for each timepoint. Interestingly, cells that had ingested latex beads clustered tightly with uninfected cells at the 4 hour timepoint, revealing that macrophages are equipped with the cellular components necessary for phagocytosis and experience no marked perturbation in their steady-state transcriptome upon ingestion of particles. This contrasts with macrophages

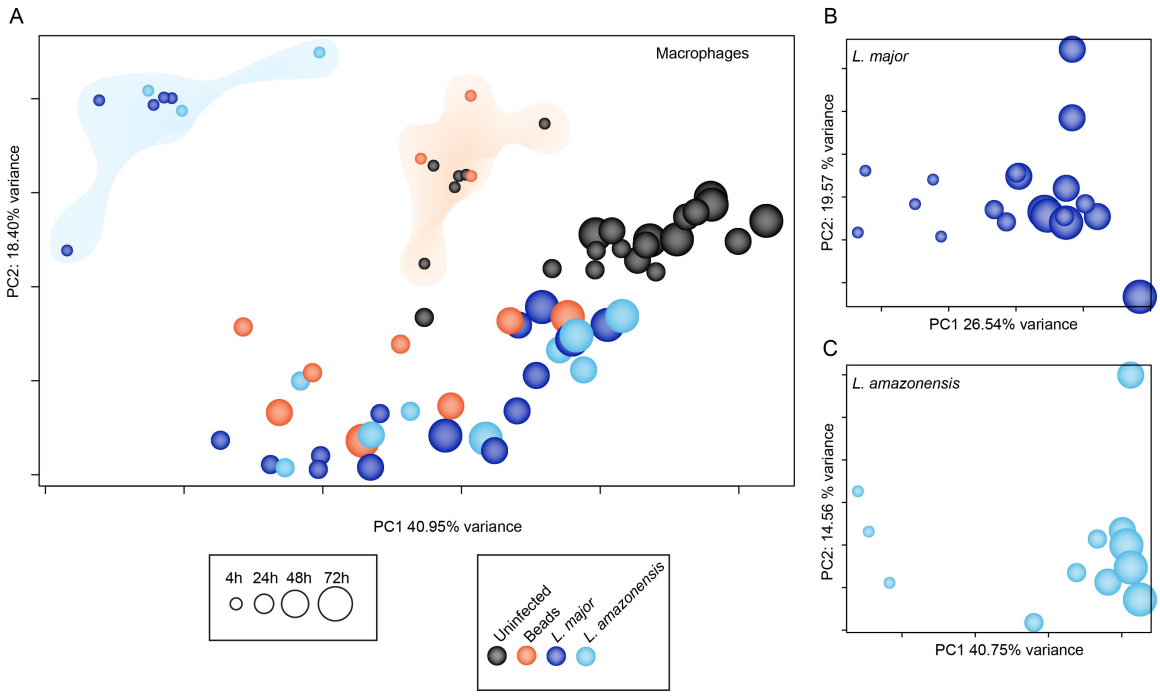


**Figure 23: Global expression profile of human macrophages and *Leishmania* parasites during infection.**

A comprehensive principal component analysis (PCA) was conducted to evaluate the relationships between samples across timepoints and to visualize sample-sample distances for (A) human macrophages (uninfected controls, *Leishmania*-infected cells, and bead-containing cells) and (B-C) the intracellular forms of *Leishmania*. Each sphere represents an experimental sample with increasing size indicating the progression of timepoints from 4 to 72 hpi. Sphere color indicates sample type (dark blue for *L. major*-infected cells/parasites, light blue for *L. amazonensis*-infected cells/parasites, orange for bead-containing cells, and black for uninfected cells). In panel A, the clouds highlight the separation of the *Leishmania*-infected cluster (blue) from the uninfected/bead-containing cluster (orange). All analyses were performed after filtering out non- and lowly-expressed genes, quantile normalization, and including experimental batch as a covariate in the statistical model.

infected with either *L. major or L. amazonensis*, which clustered together and distinctly away from other samples, pointing to an apparently unique and *Leishmania*-specific human macrophage response early in the infection (4 hpi). As the infection progressed, the separation between *Leishmania*-infected and bead-containing macrophages became less prominent, and by 24 hpi, macrophages with persisting parasites resembled those harboring inert particles. In addition, both populations displayed a clear trending towards uninfected

macrophages in the later timepoints (48 and 72 hpi). The global transcriptional patterns for intracellular *Leishmania* parasites of both species were similar with 4 hpi samples partitioning away from those observed at later timepoints (Figure 23B-C). This is consistent with the pronounced and distinct host response to the parasite, which also displayed a clear separation between responses at early and later infection.

**Discriminating between parasite- and phagocytosis-driven changes in human macrophages**

We examined the differential expression of individual human macrophage genes following *Leishmania* infection. In macrophages infected with *L. major*, the largest host response was observed at 4 hpi, with 5713 differentially expressed (DE) genes between uninfected and infected macrophages at an adjusted *P* value cutoff of <0.05. Later timepoints showed decreasing numbers of DE genes – 4846, 4188, 2142 genes at 24, 48 and 72 hpi, respectively (Figure 24A, top). A similar trend was observed in the *L. amazonensis* infection of the macrophages (Figure 24B, top) with smaller numbers of DE genes detected, which reflects a slightly reduced statistical power associated with fewer biological replicates. Most of the genes that were DE during *L. amazonensis* infection were also contained within the *L. major*-infected macrophage datasets (see comparative DE analysis below). It is also worthwhile noting that among the DE genes in macrophages infected with either parasite, the average proportion of up- and down-regulated genes was ~40%/60% at 4 hpi and ~50%/50% at the remaining timepoints. While this points to a slightly higher fraction of downregulated genes, we did not

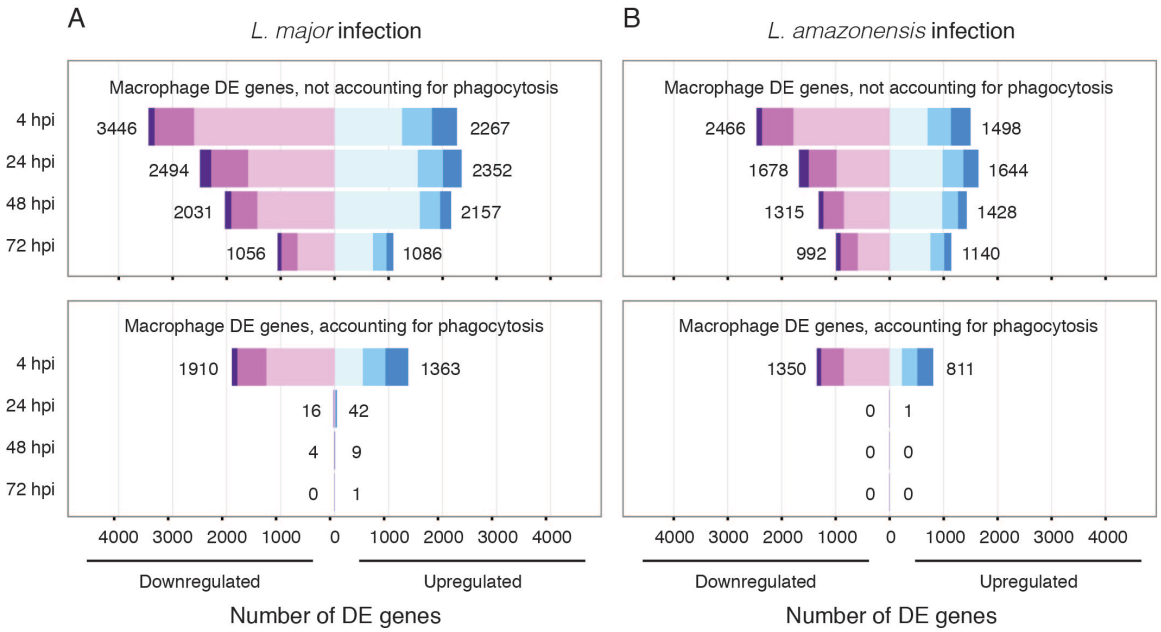


**Figure 24: Differentially expressed genes in macrophages infected with *Leishmania* species.**

The numbers of differentially expressed (DE) genes in (A) *L. major*-infected macrophages and (B) *L. amazonensis*-infected macrophages relative to uninfected controls both before and after accounting for the phagocytosis effect (top panels and bottom panels, respectively) are depicted as horizontal bar plots for 4, 24, 48, and 72 hpi. Box width depicts the number of DE genes downregulated (left; purple) and upregulated (right; blue) at an adjusted *P* value of < 0.05 with the total number of down- and up-regulated genes shown. Color shade indicates the proportion of genes with > 4-fold differential expression (dark), between 2- and 4-fold differential expression (medium), or 2-fold differential expression (light).

observe a general suppression of gene expression in murine macrophage as reported in earlier studies (Buates, 2001; Gregory, 2008). Rather, our results are consistent with studies that report similar number of genes up- and down-regulated upon *Leishmania* infection (Chaussabel, 2003; Ettinger, 2008; Ramírez, 2012; Rodriguez, 2004) and challenge the notion that phagocytosis of *Leishmania* by macrophages induces an overall state of dormancy after uptake.

In order to evaluate the extent of the phagocytosis effect, differential expression analysis was carried out to compare uninfected macrophages to bead-containing macrophages at each timepoint. This analysis revealed no DE

genes at 4 hpi, consistent with the apparent lack of transcriptional response at 4 hpi following phagocytosis of beads observed in Figure 23A. A well pronounced phagocytosis effect was observed at later timepoints with 3787, 6045, and 2659 DE genes at 24, 48, and 72 hpi, respectively. In a previous study of the response of murine bone marrow-derived macrophages to *L. major* and *L. donovani* as assessed using microarrays, Gregory *et al*. (2008) found that macrophages that had ingested beads were nearly identical to control macrophages at the one timepoint used in the study (24 hpi), and thus did not use the bead-containing samples to further account for the effects of phagocytosis. This finding more closely matches our results at 4 hpi and may indicate a time offset in the murine system relative to the human system. Additionally, the observed differences may be reflective of differences in macrophage type and incubation conditions prior to phagocytosis.

Subsequent analyses were aimed at distinguishing between changes that constitute a specific response to infection and those attributable to the macrophage response to phagocytosis. Series of pairwise analyses were conducted at each timepoint (4, 24, 48, and 72 hpi) for macrophages infected with either *L. major* or *L. amazonensis* evaluated against both matched uninfected macrophages and macrophages that had ingested latex beads. Using a novel approach based on a dual statistical test to identify genes that were differentially regulated not only in infected macrophages relative to uninfected cells, but also relative to macrophages that have ingested inert particles, we were able to filter out genes that are differentially regulated due to the phagocytosis

effect and thereby select genes that were specific to the response of the macrophage to *Leishmania* infection. This resulted in significantly reduced numbers of DE genes for each timepoint (Figure 24A-B, bottom panels). Remarkably, the parasite-specific response was most pronounced in macrophages at 4 hpi (3273 and 2161 DE genes in *L. major*- and *L. amazonensis*-infected macrophages, respectively) and greatly attenuated in later timepoints due to the fact that most of the response was attributable to phagocytosis.

### *Leishmania*-induced remodeling of gene expression in human macrophages

In examining the DE genes after accounting for the phagocytosis effect, we detected previously identified elements of the macrophage response to *Leishmania* at 4 hpi. Most recognizable was the upregulation of genes encoding inflammatory cytokines including interleukin 1B (IL1B), tumor necrosis factor (TNF), TNF superfamily members, and interleukin 6 (IL6), as well as a number of inflammatory mediators such as prostaglandin-endoperoxide synthase 2 (PTGS2), granulocyte-macrophage colony stimulating factor 2 (CSF2), and superoxide dismutase 2 (SOD2, *L. major* only). Some of these gene products have been previously implicated during infection by *Leishmania major* (Chaussabel, 2003) or other species of *Leishmania* in both human or murine macrophage systems (Ettinger, 2008; Ramírez, 2012; Rodriguez, 2004).

Among the top upregulated genes (up to 136-fold in *L. major* infection and 196-fold in *L. amazonensis* infection) at 4 hpi are multiple metallothionein 1

family members (G, M, H, E, and A). Metallothioneins are known to have an immunomodulatory role (Lynes, 1993) and to be induced by a wide range of conditions including exposure to reactive oxygen species (Ghoshal, 2001), and were recently shown to affect the host response to *Listeria* infection (Emeny, 2015). Although metallothioneins have previously been shown to be highly upregulated in macrophages infected with *Leishmania* (Chaussabel, 2003; Ettinger, 2008), their potential role in the establishment of infection is poorly understood.

*L. major* and *L. amazonensis* show different behavioral characteristics, most notably varying clinical outcomes and distinct organization of the intracellular amastigotes within the macrophage (one vacuole per parasite for *L. major* and communal vacuoles that house multiple parasites for *L. amazonensis*) (Kaye, 2011; Real, 2014). We explored whether the two species elicited significantly different responses by the macrophage across various timepoints of infection using a direct statistical comparison of the gene-level macrophage response to each parasite. This revealed that they did not trigger a significantly different response in the human macrophage with only 4 genes surfacing as differentially expressed at 4 hpi and none at any of the subsequent timepoints.

Two of the genes that are DE between *L. major*-infected and *L. amazonensis*-infected macrophages encode synaptotagmin family members (*SYT2* and *SYT8*). Synaptotagmins are membrane proteins that regulate vesicle docking and fusion in processes such as exocytosis (Arango Duque, 2013; Baram, 1999) and phagocytosis (Arango Duque, 2013; Czibener, 2006; Vinet,

2008). While some synaptotagmin family members (SYT5 and SYT11) have been implicated in *Leishmania* infection (Arango Duque, 2014; Vinet, 2009; Vinet, 2011), the involvement of SYT2 and SYT8 have not yet been investigated. Given the general role of synaptotagmins as regulators of membrane trafficking and fusion, it is possible that the higher levels of SYT2 and SYT8 observed during *L. major* infection compared to *L. amazonensis* infection may be linked to differences in the parasitophorous vacuole maintenance throughout the infection - *L. major* divide in membrane-bound compartments with each parasite division maintaining singular parasites in a vacuole; conversely, *L. amazonensis* may possibly require fewer fission events to maintain its communal vacuoles. Additionally, synaptotagmins are also known to play a role in SNARE activity regulation by mediating membrane fusion in a $Ca^{2+}$-dependent manner (Andrews, 2005; Südhof, 2009; Tucker, 2002). Since *Leishmania* have been shown to target SNARES (VAMP8 in particular) in order to modulate antigen cross-presentation (Matheoud, 2013), it is possible to speculate that transcriptional upregulation of synaptotagmins upon infection by *Leishmania* may be related to their role in the regulation of SNAREs. *CMIP*, which encodes c-Maf inducing protein, and *GABRE*, which encodes gamma-aminobutyric acid (GABA) A receptor epsilon, are also expressed at higher levels in *L. major*-infected macrophages than in *L. amazonensis*-infected macrophages, though the mechanisms by which these genes may differentially interact with or be influenced by each parasite species is unclear.

**Towards defining the signature of mammalian macrophage infection by *Leishmania***

An important aim of this study was to characterize the human macrophage response to infection by *Leishmania* parasites with the fundamental goal of defining shared features regarding how intracellular pathogens modulate their host environments. We took an initial step towards uncovering a common signature in mammalian macrophages infected with *Leishmania* by integrating our findings with data we collected from the same timepoints in *L. major*-infected murine macrophages (see Chapter 4).

Since the differential expression dataset collected from murine macrophages did not account for the phagocytosis effect, we sought to restrict this analysis to orthologous genes that were DE in both human and murine macrophages as the result of infection and not phagocytosis. To do this, we identified the mouse orthologs of the human genes that constituted the *L. major*-specific response and then compared the expression profiles for these (human) genes to corresponding DE gene profiles generated for the murine macrophages. Of the 3273 genes that were DE in *L. major*-infected human macrophages at 4 hpi after accounting for the phagocytosis effect, we identified mouse orthologs for 3017. The relationship between genes in each host system was best visualized using a scatterplot showing the magnitude and direction of differential expression for each gene and its ortholog (Figure 25). Of the 1735 genes that were DE in both host systems, 1340 were differentially expressed in the same direction in both systems and thus constitute a unique signature of *L. major* infection of

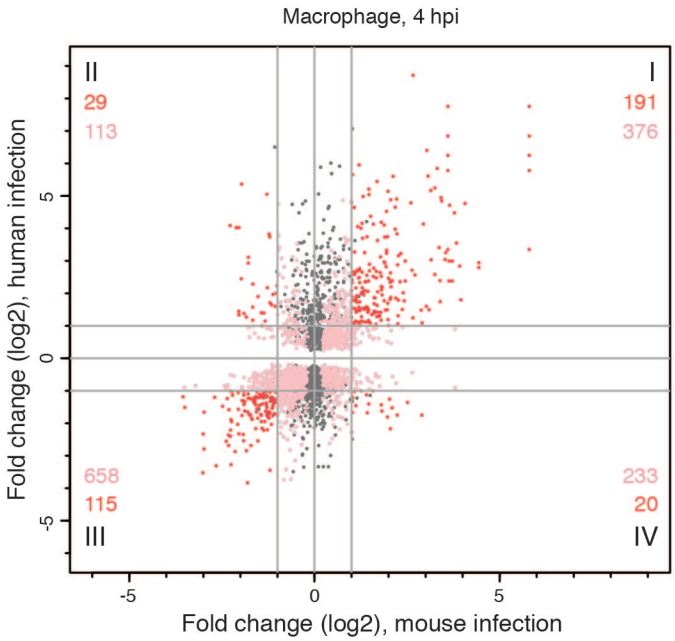


**Figure 25: Response of human and murine macrophages to *L. major* infection at 4 hpi after accounting for the phagocytosis effect.**

The differential expression profile of *L. major*-infected human macrophages was compared to that of *L. major*-infected murine macrophages (Chapter 4). Orthology mapping to mouse was done for the 3,273 human genes identified as differentially expressed at 4 hpi after accounting for the phagocytosis effect, and the results compared to the murine expression dataset. A scatterplot showing the relationship between fold changes (log2) in mouse (X-axis) and human (Y-axis) is shown with each human-mouse gene pair represented by a single point. Some genes duplicates were introduced by the orthology mapping process. Points colored in gray represent genes in the human dataset with an ortholog that was not significantly DE in the murine dataset. Points colored in shades of red represent genes that were significantly DE in both datasets with those showing < 2-fold difference (logFC < 1) in either/both host system(s) colored in pink and those with > 2-fold difference (logFC > 1) in both host systems colored in red. The numbers of unique genes represented by the red and pink points are indicated for each quadrant.

mammalian macrophages. Those genes localize in quadrants I (567 genes upregulated in both systems) and III (773 genes downregulated in both systems) and contain the most highly differentially expressed genes.

In order to identify known cellular processes within this signature, we used a KEGG pathway enrichment analysis of genes commonly up- and down-regulated in both host systems (Table 8). Many of the KEGG pathways that were most enriched in the up-regulated set of genes are related to immune activation

| Upregulated genes | | | |
|---|---|---|---|
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| TNF signaling pathway | 24 | 109 | 2.07E-13 |
| NF-kappa B signaling pathway | 14 | 88 | 1.62E-06 |
| NOD-like receptor signaling pathway | 11 | 57 | 3.07E-06 |
| Epstein-Barr virus infection | 21 | 198 | 4.43E-06 |
| Hematopoietic cell lineage | 12 | 82 | 2.17E-05 |
| Shigellosis | 10 | 61 | 3.88E-05 |
| Chagas disease (American trypanosomiasis) | 13 | 103 | 4.98E-05 |
| Proteoglycans in cancer | 19 | 198 | 5.17E-05 |
| Ribosome biogenesis in eukaryotes | 11 | 81 | 9.64E-05 |
| ErbB signaling pathway | 11 | 85 | 1.50E-04 |
| HIF-1 signaling pathway | 12 | 104 | 2.30E-04 |
| Toll-like receptor signaling pathway | 12 | 105 | 2.52E-04 |
| Osteoclast differentiation | 13 | 121 | 2.61E-04 |
| Legionellosis | 8 | 54 | 4.75E-04 |
| Cytokine-cytokine receptor interaction | 19 | 248 | 9.33E-04 |
| Pertussis | 9 | 75 | 1.04E-03 |
| VEGF signaling pathway | 8 | 61 | 1.09E-03 |
| PI3K-Akt signaling pathway | 23 | 335 | 1.27E-03 |
| Herpes simplex infection | 15 | 179 | 1.30E-03 |
| FoxO signaling pathway | 12 | 127 | 1.41E-03 |
| Mineral absorption | 7 | 50 | 1.51E-03 |
| Hepatitis C | 12 | 129 | 1.61E-03 |
| Influenza A | 14 | 169 | 2.08E-03 |
| Small cell lung cancer | 9 | 84 | 2.32E-03 |
| MAPK signaling pathway | 18 | 249 | 2.41E-03 |
| Rheumatoid arthritis | 9 | 87 | 2.95E-03 |
| Hepatitis B | 12 | 143 | 3.82E-03 |
| Jak-STAT signaling pathway | 12 | 147 | 4.77E-03 |
| Measles | 11 | 130 | 5.17E-03 |
| TGF-beta signaling pathway | 8 | 79 | 5.66E-03 |
| Pathways in cancer | 23 | 381 | 6.36E-03 |
| HTLV-I infection | 17 | 254 | 6.79E-03 |
| Central carbon metabolism in cancer | 7 | 66 | 7.35E-03 |
| Prion diseases | 5 | 36 | 7.36E-03 |
| Tight junction | 11 | 137 | 7.62E-03 |
| T cell receptor signaling pathway | 9 | 101 | 7.91E-03 |
| Fc epsilon RI signaling pathway | 7 | 68 | 8.63E-03 |
| Salmonella infection | 8 | 85 | 8.74E-03 |
| RIG-I-like receptor signaling pathway | 7 | 69 | 9.33E-03 |
| **Downregulated genes** | | | |
| **KEGG pathway** | **Number DE genes** | **Pathway Size** | ***P* Value** |
| Valine, leucine and isoleucine degradation | 10 | 46 | 2.52E-05 |
| Lysine degradation | 9 | 51 | 3.46E-04 |
| Progesterone-mediated oocyte maturation | 12 | 89 | 5.19E-04 |
| Other glycan degradation | 5 | 18 | 8.77E-04 |
| Osteoclast differentiation | 14 | 121 | 8.88E-04 |
| FoxO signaling pathway | 14 | 127 | 1.43E-03 |
| Lysosome | 13 | 120 | 2.44E-03 |
| Glycosphingolipid biosynthesis - globo series | 4 | 14 | 2.65E-03 |
| Fatty acid elongation | 5 | 23 | 2.87E-03 |
| Apoptosis | 10 | 81 | 2.91E-03 |
| Chronic myeloid leukemia | 9 | 72 | 4.27E-03 |
| Platelet activation | 13 | 128 | 4.29E-03 |
| Chagas disease (American trypanosomiasis) | 11 | 103 | 5.70E-03 |
| Fc gamma R-mediated phagocytosis | 10 | 90 | 6.24E-03 |
| Propanoate metabolism | 5 | 28 | 6.99E-03 |
| Glycosaminoglycan degradation | 4 | 18 | 7.04E-03 |
| Estrogen signaling pathway | 10 | 95 | 9.07E-03 |

**Table 8: KEGG pathways enriched in genes common to human and murine macrophages infected with *L. major* at 4 hpi.**

KEGG pathway analysis was carried out to identify signaling and metabolic pathways over-represented in genes that constitute the mammalian response to *L. major* infection at 4 hpi. Genes that were commonly regulated in both systems were used as input. For each enriched KEGG pathway, the number of DE genes assigned to that pathway, the total number of genes in the pathway, and the *P* value for the enrichment are reported.

and signaling. Signaling pathways related to the recognition of pathogen-associated molecular patterns (PAMPs) (NOD-like receptor, RIG-I-like receptor, and TLR signaling) were implicated as were many other immune system signaling pathways (cytokine-cytokine receptor interaction, Fc epsilon RI, T cell receptor, Jak-STAT, MAPK, NF-kappa B, TNF, VEGF, HIF-1, ErbB, FoxO, PI3K-Akt, and TGF-beta signaling). Given the effects of many of these pathways on cell growth and metabolism, it is not surprising that many have been previously implicated in cancer and that a number of cancer-related pathways appear in the enriched KEGG pathway lists. KEGG pathways enriched among genes that were downregulated in both the human and murine systems are associated with energy metabolism (amino acid and glycan degradation), lysosome structure and processes, and apoptosis. Interestingly, the FoxO signaling pathway, which is involved in the regulation of cell growth, gluconeogenesis, and adipogenesis, was enriched among both up- and down-regulated genes, with 12 members of the pathway upregulated and 14 others downregulated.

It is important to note that there is a significant amount of redundancy within the gene candidates that drive the enrichment of known pathways. An analysis of the overlap between the DE genes contained in different pathways revealed that the KEGG enrichment results can be solely attributed to 142 up-regulated and 85 down-regulated genes. This is explained by the nature of KEGG pathways (and similar databases) which reflect finite classifications that largely represent areas of research emphasis. Genes driving the KEGG results among the 39 upregulated pathways include *TNF* (22 pathways), *PIK3CB* (22

pathways), *IL6* (21 pathways), *MAPK8* (20 pathways), *NFKBIA* (19 pathways), *MAPK11* (18 pathways), *IL1B* (17 pathways), *MAP2K1* (17 pathways), *MYC* (12 pathways), *IL1A* (10 pathways), *TRAF2* (10 pathways), *CDKN1A* (10 pathways), *NFKB1B* (8 pathways), *TICAM1* (8 pathways), and *VEGFA* (7 pathways) while those driving results among the 17 downregulated pathways include *AKT1* (9 pathways), *PIK3CG* (9 pathways), *PIK3R2* (9 pathways), *MAPK14* (5 pathways), *GNAI1* (4 pathways), *HADHA* (4 pathways), *HEXA* (4 pathways), *HEXB* (4 pathways), *PRKACA* (4 pathways), *PRKACB* (4 pathways), *TGFBR1* (4 pathways), and *TFGBR2* (4 pathways).

Disease-specific KEGG pathways appeared prominently in the enrichment analysis lists (including those for Chagas disease, Epstein-Barr virus infection, hepatitis, herpes simplex infection, HTLV-I infection, influenza, legionellosis, measles, pertussis, salmonella infection, and shigellosis). These enrichments were largely driven by redundant gene entries within those pathways as described above, and KEGG pathways for individual diseases appear to represent a comprehensive assemblage of observations from multiple studies of individual genes (rather than a consensus) from different experiments, hosts, cell types, and timepoints.

To further extend the interpretation of our *L. major* and *L. amazonensis* infection datasets, we attempted to identify some of the common elements in the human macrophage response to infection by comparing the datasets reported here to previous reports on the infection of human macrophages infected with *Leptospira interrogans* (Xue, 2013), *Legionella pneumophila* (Price, 2014), and

*Mycobacterium avium* (Blumenthal, 2005). Even though we were unable to conduct a comprehensive comparison since expression values were reported only for subsets of genes in the published datasets, we were able to make some useful observations. A number of cytokines, chemokines, and inflammatory mediators upregulated in the *L. major* and *L. amazonensis* infections were also upregulated in these other systems and therefore constitute a shared response to infection. Specifically, *IL6* and *TNF* were upregulated in infections by the other 3 pathogens, *IL1B* was also upregulated in *M. avium* and *L. interrogans* infections, *TNFSF9* was also upregulated in *M. avium* and *L. pneumophila infections, CCL20* and *CXCL3* were also upregulated in *L. pneumophila* and *L. interrogans* infections, *CXCL1* and *CXCL5* were also upregulated during *L. interrogans* infection, and *PTGS2* was also upregulated during *M. avium* infection. All of these genes except *CCL20* and *CXCL5* were also upregulated in the *L. major*-infected murine macrophage dataset described in Chapter 4. Most of these proteins are proinflammatory in nature (IL6 has both pro- and anti-inflammatory properties) and play a role in the recruitment and stimulation of a range of immune cells. The involvement of this broad repertoire of gene products reflects the complex and at times opposing immune responses that occur as intracellular pathogens attempt to establish an infection and survive intracellularly.

**Identification of differentially expressed genes in *Leishmania* parasites**

This study provided an opportunity to characterize not only the gene expression patterns of human macrophages as they responded to infection by

*Leishmania* parasites, but to simultaneously study changes in the parasites' gene expression programs as an intracellular infection was established and progressed. Differential expression analysis was carried out for parasite genes to determine how *Leishmania* species modify the expression of their individual genes during the transition from metacyclic promastigotes to intracellular amastigotes and over the course of an intracellular infection. The largest numbers of DE genes were observed as metacyclic promastigotes infected human macrophages with 3224 *L. major* genes and 3896 *L. amazonensis* genes implicated at an adjusted *P* value cutoff of <0.05 (Figure 26). Significantly fewer changes in expression were observed in intracellular parasites across time, with 356 DE genes and 577 DE genes in *L. major* and *L. amazonensis*, respectively, during the 4 hpi to 24 hpi transition. Only 3 DE genes were observed for each species during the 24 hpi to 48 hpi transition and no genes were DE during the

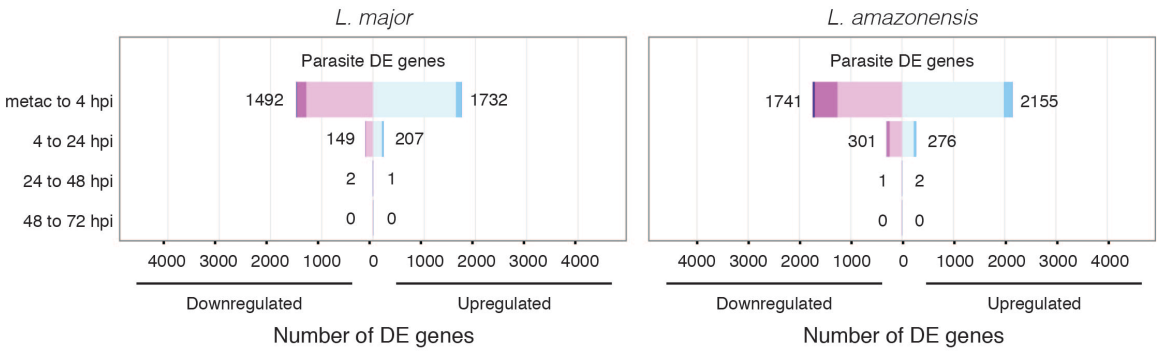


**Figure 26: Differentially expressed genes between *Leishmania* parasite developmental stages.**

The numbers of differentially expressed (DE) genes between *Leishmania* developmental stages/timepoints were determined. Box width depicts the number of DE genes downregulated (left; purple) and upregulated (right; blue) at an adjusted *P* value of < 0.05 with the total number of down- and up-regulated genes shown. Color shade indicates the proportion of genes with > 4-fold differential expression (dark), between 2- and 4-fold differential expression (medium), or 2-fold differential expression (light).

48 hpi to 72 hpi transitions. This pattern of expression - large numbers of DE genes during the metacyclic to early amastigote transition and decreasing numbers of DE genes over time - suggests that the large gene expression reprogramming that occurs as extracellular promastigote parasites transform into intracellular amastigote parasites generally stabilizes by around 24 hpi. This mirrors the changes in host transcriptomes where large changes were observed during early infection and many fewer changes were observed at later timepoints once the phagocytosis effect was considered (see above and Figure 24A-B lower panels). The parallel expression patterns between parasite and host suggest that host-pathogen interactions are extensive upon *Leishmania* entering and establishing infection within the cells but are virtually nonexistent by around 24 hours after infection.

In an attempt to directly compare gene expression programs between the two *Leishmania* species as they transition from the metacyclic form into intracellular forms, we focused our analysis on a set of *L. major/L. amazonensis* (*L. mexicana*) orthologous gene clusters that were precomputed in EuPathDB and include ~98% of *L. major* genes. A graphical representation of the results is shown in Figure 27, where dots correspond to the differential expression levels for orthologous genes. The orthology mapping between the two species included several one-many relationships because of paralogous gene families and this is manifested in the scatterplot as rows and columns of perfectly aligned dots representing amastin and SHERP gene products, among few others. We

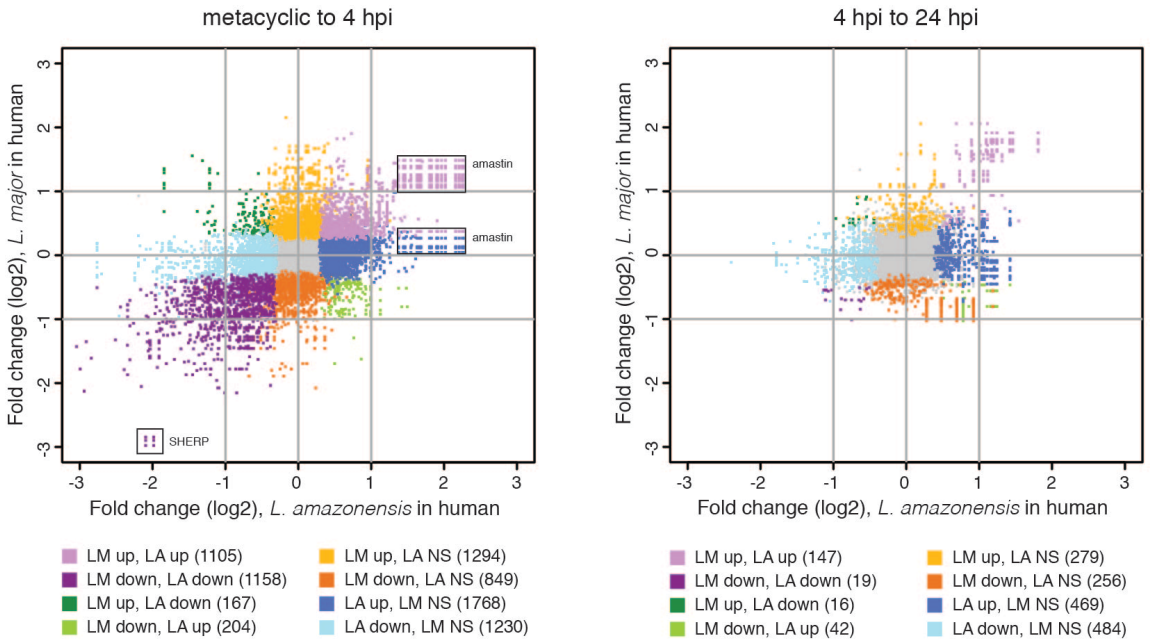


**Figure 27: Comparison of *Leishmania* transcriptomes between species.**

Orthology mapping was done to enable the comparison of expression profiles for *L. major* and *L. amazonensis* genes that were DE upon the infection of human macrophages. Scatterplots showing the relationship between parasite gene fold changes (log2) for *L. amazonensis* parasites (X-axis) and *L. major* parasites (Y-axis) are shown for the metacyclic promastigote to 4 hpi intracellular transition (left panel) and the 4 hpi to 24 hpi intracellular transition (right panel). Genes upregulated in the second stage/timepoint relative to the first exhibit positive fold changes, while downregulated genes exhibit negative fold changes. Genes highlighted in the text are noted. Points colored in gray represent genes that were not significantly DE in either system, points colored in shades of purple represent genes that were DE in the same direction in both systems, points colored in shades of green represent genes that were DE in different directions between the systems, and points colored in shades of orange and blue represent genes significant only in one system. The numbers of DE genes corresponding to each color are included in the legends. NS = not significant

observed a set of 1,558 unique gene pairs (737 upregulated and 811 downregulated) that were significantly differentially expressed in the same direction in both species during the metacyclic promastigote to 4 hpi intracellular transition and representing gene products that underlie a response common to both species as they adapt to the human macrophage environment (Figure 27, left panel). Two other subsets included transcripts that appeared significantly DE in one species but not in the other. Those included 3,947 unique transcripts in

total with only a small proportion (10%) that were differentially expressed greater than 2-fold in one species or the other. A closer inspection revealed that many of the genes that make up these subsets are member of multi-gene families that have members belonging to the common expression set (e.g. amastin, GP63, kinesin, flagellar attachment zone protein, AAT family members, dynein, cysteine peptidase B), and/or were DE in the other *Leishmania* species during the 4 hpi to 24 hpi transition, indicating a time offset in what can be considered a common response (e.g. cathepsin L-like protease, ATG8). A similar phenomenon was apparent during the 4 hpi to 24 hpi intracellular transition (Figure 27, right panel). While there were only 60 unique gene pairs that were commonly expressed in the same direction, most of the transcripts that appeared to be distinctly DE in one of the species were members of multigene families that were part of the common response or simply displayed a time-shifted expression pattern.

The general picture that emerged from the comparison of the two species of *Leishmania* is that the transcriptomes of their orthologous gene sets display minimal to non-existent differences within the human macrophage. This finding closely aligns with the lack of observed differences in the host response to infection by the two *Leishmania* species (described above). Of the genes that were not included in the above analysis because they lacked orthologs in the EuPathDB database, one gene emerged that is both highly differentially expressed and has supporting evidence of its species-specificity. This gene, the virulence factor A2, is present in *L. amazonensis* but exists only as a pseudogene in *L. major* (Zhang, 2003). It was among the most upregulated

genes in *L. amazonensis* during the metacyclic to 4 hpi transition and was further upregulated through 24 hpi. A2 is required for visceral infection by *L. donovani* (Zhang, 2001) and has been implicated in the parasite's resistance to heat shock (McCall, 2012). While *L. amazonensis* generally causes cutaneous disease, it has also been associated with visceral infections (Almeida, 1996). The strong upregulation of the A2 gene during early infection detected here is consistent with that observation.

**Towards defining the signatures of *Leishmania* differentiation and intracellular survival**

In order to characterize the overall signature of *L. major* differentiation from the metacyclic promastigote developmental stage to intracellular amastigotes and during intracellular infection, we compared the patterns of *L. major* gene expression identified in human (this study) to those observed for the same timepoints in *L. major*-infected murine macrophages (Chapter 4). In both host systems, large numbers of DE genes were detected in the parasite during the metacyclic to early amastigote transition and decreasing numbers of DE genes were seen between intracellular timepoints.

When individual genes were compared between the human and mouse analysis, a complex picture emerged (Figure 28). Among the 4,708 genes that are DE in *L. major* during the metacyclic promastigote to 4 hpi intracellular form transition in at least one of the hosts, 28% (1,336) represent a common parasite expression pattern in both human and mouse infections. A gene ontology (GO) enrichment analysis was done to glean insights into the functional overlaps in *L.*

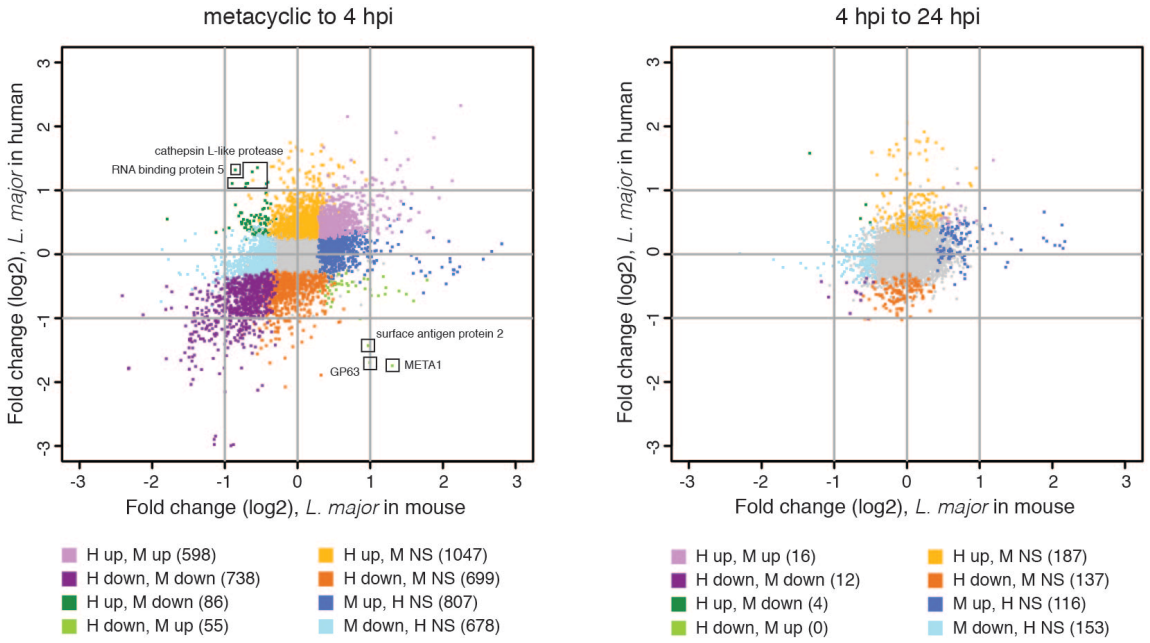


**Figure 28: Comparison of the *L. major* transcriptome upon infecting human and murine macrophages.**

Scatterplots were used to show a comparison of *L. major* parasite expression patterns upon infecting human vs. murine macrophage (Chapter 4). The relationship between parasite gene fold changes (log2) in the mouse system (X-axis) and the human system (Y-axis) are shown for the metacyclic promastigote to 4 hpi intracellular transition (left panel) and the 4 hpi to 24 hpi intracellular transition (right panel) with each of the 8,479 genes that passed low-count filtering in both datasets represented by a single point. Genes upregulated in the second stage/timepoint relative to the first exhibit positive fold changes, while downregulated genes exhibit negative fold changes. Genes highlighted in the text are noted. Points colored in gray represent genes that were not significantly DE in either system, points colored in shades of purple represent genes that were DE in the same direction in both systems, points colored in shades of green represent genes that were DE in different directions between the systems, and points colored in shades of orange and blue represent genes significant only in one system. The numbers of DE genes corresponding to each color are included in the legends. NS = not significant

*major* genes that were differentially expressed as metacyclic promastigotes infect mammalian macrophages over the first 4 hours. Two enriched GO categories emerged - peroxiredoxin activity and antioxidant activity - among the 598 *L. major* genes that were commonly upregulated during both human and mouse infections (Table 9). Tellingly, only 7 genes - peroxidoxin and 6 tryparedoxin peroxidase family members - drove these results. A greater number and range of GO categories were enriched among the 738 *L. major* genes downregulated in

both host systems. These 13 GO categories were largely related to signaling, microtubule dynamics, and fatty acid biosynthesis. Many of the genes driving the GO results for the genes consistently regulated in both systems have been previously implicated in *Leishmania* virulence (i.e. tryparedoxin peroxidase family members (Iyer, 2008), casein kinase (Dan-Goor, 2013), adenylate cyclase (Biswas, 2011)) or regulation of flagellar dynamics (i.e. calmodulin (Ginger, 2013), dynein heavy chain (Harder, 2010), kinesin (Blaineau, 2007; Vicente, 2015), paraflagellar rod protein (Lander, 2015), multiple mitogen-activated protein (MAP) and NIMA kinases (Bengs, 2005; Cao, 2009; Erdmann, 2006; Wiese, 2003)).

Patterns of parasite gene expression that were unique to the human and mouse hosts were also clearly detectable and likely reflect the biological

| GO ID | GO term | *P* value |
|---|---|---|
| **Upregulated, human and mouse** | | |
| GO:0051920 | peroxiredoxin activity | 3.93E-08 |
| GO:0016209 | antioxidant activity | 1.51E-07 |
| **Downregulated, human and mouse** | | |
| GO:0005516 | calmodulin binding | 0.00E+00 |
| GO:0004674 | protein serine/threonine kinase activity | 9.30E-14 |
| GO:0004672 | protein kinase activity | 3.53E-13 |
| GO:0006468 | protein phosphorylation | 8.03E-13 |
| GO:0004713 | protein tyrosine kinase activity | 8.85E-13 |
| GO:0009434 | motile cilium | 2.91E-11 |
| GO:0009190 | cyclic nucleotide biosynthetic process | 1.45E-07 |
| GO:0016849 | phosphorus-oxygen lyase activity | 1.45E-07 |
| GO:0005524 | ATP binding | 2.35E-06 |
| GO:0003777 | microtubule motor activity | 2.36E-06 |
| GO:0035556 | intracellular signal transduction | 4.01E-06 |
| GO:0006633 | fatty acid biosynthetic process | 3.51E-05 |
| GO:0030286 | dynein complex | 8.08E-05 |

**Table 9: GO categories enriched among *L. major* genes in common to human and mouse infections.**

`GOseq` was used to identify GO categories enriched among *L. major* genes that were DE during the metacyclic promastigote to 4 hpi intracellular transition (P value < 0.05). Parasite genes that were expressed in the same direction during infection of both human and mouse host systems were evaluated with up- and down-regulated genes considered separately.

differences inherent to each system. A total of 1,887 *L. major* genes were uniquely regulated in human macrophages (1133 up, 754 down) and 1,626 respond specifically to the infection of murine macrophages (862 up, 764 down). Among the genes that stood out as the most divergent in their behavior in the two host systems were META1, GP63, and surface antigen protein 2, which were downregulated in human and upregulated in mouse, and RNA-binding protein 5 and cathepsin L-like proteases, which were upregulated in human and downregulated in mouse. Both GP63 and META1 are known to play a role in virulence. GP63 is a zinc-dependent metalloprotease found on the surface of the parasite that cleaves C3b to iC3b, thus helping the parasite avoid complement-mediated lysis and promoting parasite entry into the cell via complement receptors 1 and 3 (Brittingham, 1995; Mosser, 1985; Olivier, 2012). Entry via these receptors has been found to contribute to parasite survival (Da Silva, 1989; Mosser, 1987). Additionally, GP63 hydrolyzes host kinase substrates (Corradin, 1999; Olivier, 2012), and may thus alter the signaling processes of the host cell. META1 is known to localize to the flagellar pocket of the procyclic form of the parasite and has been implicated in secretory processes (Nourbakhsh, 1996; Puri, 2011) and cathepsin-L like proteins have been shown to be key modulators of the host immune response (Alexander, 1998; Mottram, 2004) and are regarded as possible molecular targets against leishmaniasis (de Sousa, 2015; Desai, 2006). The differential expression of these proteins upon infection of human and mouse macrophages may reflect differences in the intricacies of

host-parasite immune-response/adaptation and/or the complex signaling pathways triggered by the parasite in the context of each host system.

A GO analysis yielded some additional insights into different parasite behavior in the two hosts (Table 10). Several categories related to proton transport were enriched among genes that *Leishmania* upregulated in human system, but not in mouse, reflecting an upregulation of the expression of multiple vacuolar ATP synthase genes. This difference in the parasite response to the host system could be indicative of the relative acidity of the phagosomes in each. A number of categories related to ribosomes and translation were also enriched among genes that parasites upregulated in human but downregulated in mouse. The reason for this difference is unclear.

As highlighted earlier, the 4 to 24 hpi intracellular form transition of the parasite was characterized by considerably smaller numbers of genes that were DE and by lower changes of magnitude. Of the 625 *L. major* genes that were DE during that transition in at least one of the hosts, only 28 represent a common parasite expression pattern in both human and mouse infections while 328 and 273 genes were unique to the human and mouse systems, respectively. Despite the large number of differences, almost all are less than 2 fold DE and therefore represent small changes in the context of the host system.

Even though our interpretations of the analyses above have focused on previously characterized gene products, it is worth noting that significant proportions of the DE genes across parasite timepoints encode hypothetical/ unspecified proteins - 60% and 67% of genes in the metacyclic to 4 hpi transition

| | | |
|---|---|---|
| **Upregulated, human only** | | |
| GO:0000786 | nucleosome | 1.05E-11 |
| GO:0006334 | nucleosome assembly | 1.06E-11 |
| GO:0046982 | protein heterodimerization activity | 2.51E-11 |
| GO:0006412 | translation | 4.72E-09 |
| GO:0015986 | ATP synthesis coupled proton transport | 2.40E-08 |
| GO:0003735 | structural constituent of ribosome | 3.02E-08 |
| GO:0005840 | ribosome | 5.89E-08 |
| GO:0015991 | ATP hydrolysis coupled proton transport | 1.28E-06 |
| GO:0008234 | cysteine-type peptidase activity | 3.73E-06 |
| GO:0046933 | proton-transporting ATP synthase activity | 7.18E-05 |
| GO:0006886 | intracellular protein transport | 9.44E-05 |
| GO:0030117 | membrane coat | 1.07E-04 |
| GO:0015031 | protein transport | 3.08E-04 |
| GO:0046034 | ATP metabolic process | 3.40E-04 |
| GO:0033180 | proton-transporting V-type ATPase | 3.70E-04 |
| GO:0016192 | vesicle-mediated transport | 4.46E-04 |
| GO:0015078 | hydrogen ion transmembrane transporter activity | 5.29E-04 |
| **Downregulated, human only** | | |
| GO:0007018 | microtubule-based movement | 4.70E-05 |
| GO:0006310 | DNA recombination | 6.58E-05 |
| **Upregulated, mouse only** | | |
| GO:0006950 | response to stress | 1.47E-12 |
| GO:0006486 | protein glycosylation | 1.09E-05 |
| GO:0008378 | galactosyltransferase activity | 3.21E-05 |
| GO:0007155 | cell adhesion | 6.70E-05 |
| GO:0008160 | protein tyrosine phosphatase activator activity | 1.15E-04 |
| GO:0042025 | host cell nucleus | 1.15E-04 |
| GO:0044081 | modulation by symbiont of host NO-mediated signaling | 1.15E-04 |
| GO:0075130 | modulation by symbiont of host protein kinase-mediated signaling | 1.15E-04 |
| GO:0006260 | DNA replication | 1.35E-04 |
| GO:0051082 | unfolded protein binding | 2.00E-04 |
| **Downregulated, mouse only** | | |
| GO:0006412 | translation | 2.04E-18 |
| GO:0005840 | ribosome | 2.06E-18 |
| GO:0003735 | structural constituent of ribosome | 3.53E-18 |
| GO:0005622 | intracellular | 3.15E-09 |
| GO:0006184 | GTP catabolic process | 3.43E-05 |
| GO:0005779 | integral component of peroxisomal membrane | 7.39E-05 |
| GO:0016559 | peroxisome fission | 7.39E-05 |
| GO:0008234 | cysteine-type peptidase activity | 1.37E-04 |

**Table 10: GO categories enriched among *L. major* genes unique to human and mouse infections.**

`GOseq` was used to identify GO categories enriched among *L. major* genes that were DE during the metacyclic promastigote to 4 hpi intracellular transition (P value < 0.05). Parasite genes that were specific to infection in the human system or specific to infection in the murine system were evaluated with up- and down-regulated genes considered separately.

and 39% and 55% of genes in the 4 hpi to 24 hpi transition for *L. major* and *L. amazonensis*, respectively. While significant numbers of *Leishmania* hypothetical

genes remain uncharacterized to date, they play an integral role in the parasite's strategy for survival and modulation of functional pathways within the host.

## Conclusion

RNA-sequencing was used to characterize the transcriptomes of primary human macrophages infected with either *Leishmania major* or *Leishmania amazonensis* and those fed latex beads over a 72-hour time course. This study differs from most previous efforts in its use of synchronized primary cells and controls designed to assess changes over time relative to both the uninfected state and the effects of large particle phagocytosis. The unambiguous mapping of reads from infected samples to both host and parasite genomes enabled the identification of differentially expressed genes within a dynamic dual biological system, and the collection of data from multiple biological replicates allowed careful statistical analysis of variation and the detection and removal of batch effects. These features provided us with a unique ability to detect biological differences between samples and timepoints with high confidence and sensitivity and to parse the potential causes of the observed changes.

While we found large numbers of differentially expressed genes between uninfected and *Leishmania*-infected cells at each timepoint, additional analyses considering expression changes in bead-containing cells reveled that almost all of the differences in gene expression from 24-72 hpi were part of a general response to phagocytosis, or, perhaps more accurately, "macrophage indigestion". Additionally, few differences existed between macrophages infected with *L. major* and those infected with *L. amazonensis*. Further comparison with

differential expression data generated in murine macrophages was used to define a signature of mammalian host infection by *L. major*. KEGG pathways enriched among upregulated genes in the signature were specific to immune activation and signaling while those enriched among downregulated genes were related to energy metabolism, lysosome structure and processes, and apoptosis.

Analysis of parasite data revealed a stabilization in the *Leishmania* response to the human intracellular environment for both *L. major* and *L. amazonensis* after 24 hpi and few meaningful differences were detected between the transcriptional programs activated by each species. Only about a third of *L. major* genes were similarly differentially expressed in both human and murine host systems, pointing to potentially meaningful differences, especially regarding previously identified virulence factors. These findings and the datasets generated by this work will be of great use to the *Leishmania* and macrophage biology communities as further attempts are made to understand the complex interplay between host and parasite during infection by *Leishmania* parasites. The comprehensive datasets generated in this study will also serve as a reference for future studies using different *Leishmania* strains (or even different pathogens altogether) to examine infections of macrophages isolated from multiple sources and in various states of activation, polarization, or rest. In conjunction with datasets that will be produced for other pathogens, a clearer picture of the signature of intracellular infection will emerge, providing additional insights into how pathogens are able to evade host defenses and modulate the biological functions of the cell in order to survive in the mammalian environment.

# Chapter 6

## Future Directions

In addition to deriving biological insights into *Leishmania* differentiation and infection, the work described here produced significant resources that will be of use to the *Leishmania* and broader host-pathogen research communities – specifically, UTR boundaries for a large majority of *L. major* genes, lists of genes that are differentially expressed across life cycle stages by *L. major* and *L. amazonensis* parasites, and differential expression datasets for murine and human macrophages upon *Leishmania* infection or the ingestion of inert particles. While a number of findings were described above regarding the general patterns of gene expression, higher-level pathways and processes implicated in infection, and the suspected roles of individual genes, additional research is needed to more fully mine and follow-up on the information contained in these datasets.

## Extend *L. major* RNA processing/gene regulation analysis

In this work, we determined UTR boundaries specific to the procyclic and metacyclic promastigote developmental stages for most *L. major* genes. The methods used should be extended to the amastigote stages/timepoints to identify and quantify UTR boundaries and their usage. We were unable to detect a correlation between differences in the steady-state mRNA levels (as assessed by RNA-seq) and differential usage of *trans*-splicing and polyadenylation sites (i.e. different UTR lengths) between the procyclic and metacyclic stages. This could

mean that differential engagement of sequence or structural motifs in UTRs by regulator proteins, rather than motif composition itself, is responsible for changes in steady-state mRNA levels between the stages. The UTRs for all stages should be systematically evaluated to identify motifs and to determine potential binding partners that may be responsible for the stage-specific differences in expression.

Additionally, while RNA-seq is a robust and high-throughput method for detecting cellular mRNAs, mRNA levels are not an ideal indicator of the biological changes taking place within *Leishmania* given the multiple levels of regulation that occur after transcription. For this reason, the RNA-seq data generated here would also be of great use when considered alongside ribosome profiling data that have been generated for the *L. major* procyclic promastigote stage (Belew *et al*., in preparation). A direct comparison of these data types will reveal the relationship between steady-state levels of RNAs (as measured by RNA-seq) and the rate of translation (as measured by ribosome profiling) to provide additional clues for how the parasite regulates cellular protein levels.

**Deepen analysis of gene expression from 0-24 hpi**

The most unexpected findings in this body of work were the lack of statistically significant changes in gene expression between subsequent timepoints after 24 hpi and that differences between uninfected and infected host cells after 24 hpi were attributable almost exclusively to phagocytosis. This adds a level of nuance to the existing hypothesis that *Leishmania* entry into macrophages is quiescent – parasite entry does in fact cause a large number of changes in host cells which have elements of both pro-inflammatory and anti-

inflammatory responses, but expression changes plateau by around a day after infection. Knowing this now, it would be interesting to collect data from additional timepoints during the first day (e.g. 2, 6, 8, 12, 16, and 20 hpi) to further tease apart the changes that occur in the very important early stages of infection. The collection of these additional timepoints would in turn open new methods for analyzing the data beyond the differential expression analyses done so far. Specifically, network analyses would be extremely useful for grouping together genes that have common expression patterns over time. This should be done separately for both parasite and host genes, with subsequent efforts to find correlations between parasite and host clusters. It is reasonable to hypothesize that some sets of parasite and host genes will show matching patterns of expression over time while other sets will show divergent expression patterns or reveal a delay in response. Network analyses also have the potential to point to the functional roles of hypothetical genes by grouping them with genes of known function and inferring guilt-by-association relationships. While network analyses can be attempted on the current dataset, the lack of dynamic changes in expression after 24 hpi limits our ability to group together genes that are likely undergoing co-regulation. An analysis of functional enrichments or motif usage in the UTRs of genes in each network cluster has the potential to be particularly informative.

**Expand infection systems and conditions studied**

In this study, we looked at the infection of C57BL/6 murine peritoneal macrophages and CD14+ monocyte-derived human macrophages by *L. major*

and/or *L. amazonensis*. While we found that there were few significant changes in parasite genes between the *Leishmania* species, it would be useful to extend this analysis to include other species, especially those that cause visceral disease, to identify parasite factors that may contribute to the gross differences in disease presentation. The use of gently heat-killed *Leishmania* would be an interesting contrast to the latex beads time course and would show how the macrophage responds to the phagocytosis of a particle that it is able to fully digest and various other experiments done using inhibitors of various genes implicated during infection (e.g. synaptotagmin family members, complement receptors) would also be useful for honing in on the roles of each during infection. It would also be useful to look at *Leishmania* infection of the THP1 human monocytic leukemia cell line, which has been used in many studies of pathogens that infect macrophages, to identify similarities and differences related to the use of primary vs. transformed cells. Experiments using other phagocytes such as dendritic cells or neutrophils (human or murine) should also be done since there is evidence that neutrophils are the first cells to encounter and take up *Leishmania* upon inoculation into the host (Peters, 2008), even though they are not the ultimate host cell type for the parasite. Data from patient lesion samples would also be very interesting to compare against the cell culture-derived data to determine if any similarities can be detected between an infection of synchronized cells of a single type and a true infection that involves interactions between multiple cell types.

**Detect parasite-host protein-protein interactions to derive biological mechanisms**

A number of genes were implicated in *Leishmania* differentiation and in the establishment and progression of an intracellular infection. While the bioinformatics approaches used here were extremely useful for the global identification of differentially expressed genes in both the parasite and its host cells, wet bench studies aimed at studying individual or small groups of genes are needed to follow-up on these observations. Direct interactions between *Leishmania* and mammalian proteins are responsible for communication between host and parasite as *Leishmania* are phagocytosed by macrophages and become contained inside parasitophorous vacuoles. Indeed, some genes in the DE lists suggest an interaction between *Leishmania* and host macrophages that affect the receptors by which *Leishmania* are taken into macrophages, with direct implications for their survival (e.g. *Leishmania* GP63 and host Fcγ receptors). In order for *Leishmania* proteins to interact with the host, the parasite must either secrete or display the peptides on their surfaces using glycosylphosphatidylinositol (GPI) anchors. The DE datasets should be analyzed to detect signal sequences that drive protein secretion or surface anchoring; additionally, transmembrane hidden Markov models (TM-HMM) can be used to predict trans-membrane proteins that may contain domains that are able to interact with host proteins. Genes identified by these analyses can be used as baits in yeast-2-hybrid (Y2H) experiments or pull down assays followed by mass spectrometry to identify host-pathogen protein-protein interactions. These

experiments can be used to identify proteins that interact with hypothetical genes, possibly providing evidence as to their function. Perturbations in the system – such as those introduced by altering experimental conditions (e.g. media components, use of serum for opsonization, etc.) or by removing genes/mRNAs using gene knock-outs or siRNA knock-downs - can then be used to test how mRNA or protein levels, protein-protein interactions, or the ability of *Leishmania* to infect cells change under various conditions. Findings from these studies can be used to develop models for the mechanisms of action of various genes that are necessary for infection. Genes involved in host-pathogen interactions may prove to be important for virulence of the pathogen and could potentially serve as candidates for vaccine or drug development.

**Establish new host-pathogen interaction resources**

Interpretation of the host DE gene lists has been hindered by the lack of robust datasets against which to compare the results. Our attempts to perform KEGG pathway analyses to identify the higher-level pathways and processes associated with *Leishmania* infection of macrophages were of limited usefulness given differences in the studies used to build the KEGG reference datasets. Specifically, KEGG pathways have largely been defined based on combinations of multiple single or few-gene studies produced under various conditions, and because of this, they are of undetermined relevance and reliability. For example, a KEGG pathway exists for leishmaniasis, but an examination of the 23 studies used to create the KEGG resource reveals a lack of consistency in scope of each study that contributes gene(s) to the pathway. Rather than relying on resources

such as this, robust RNA-seq datasets like those produced here should serve as new "gold standard" pathway/process resources that can be compared to datasets produced under other conditions or for other experimental systems. Comparison to similar datasets produced for other pathogens that infect macrophages, including *Mycobacterium sp.*, *Toxoplasma gondii*, *Francisella tularensis*, *Leptospira interrogans,* and *Legionella pneumophila*, would be very useful for determining similarities and differences in how various pathogens are able to evade immune defenses to infect immune cells.

# Chapter 7

## Experimental Procedures

### *Leishmania* Culture

Metacyclogenesis and mouse experiments

*Leishmania major* (clone V1, MHOM/IL/80/Friedlin) was isolated after passage through BALB/c mice. Promastigotes were grown in 50% M199 39% Schneider medium along with 10% FBS and 1% of Penicillin/streptomycin at 25°C. *L. major* promastigotes were not split for more than 5 passages to maintain virulence of the cultures. Enrichment for metacyclic promastigotes from stationary phase cultures was done by Ficoll density gradient centrifugation (Späth, 2001) and by negative selection with peanut agglutinin (PNA) (da Silva, 1987). Biological replicates each originated from a separate growth and were obtained on different days.

Human experiments

*Leishmania major* (clone VI, MHOM/IL/80/Friedlin) and *L. amazonensis* (IFLA/BR/67/PH8) promastigotes were cultivated at 26°C in medium 199 (Gibco, Invitrogen) supplemented with 5% penicillin/streptomycin, 0.1% hemin (25 mg/ml in 50% triethanolamine), 20% heat-inactivated fetal bovine serum (FBS), 10 mM adenine (pH 7.5), and 5 mM L-glutamine. Metacyclic forms of *L. major* were purified by agglutination of stationary-phase promastigote cultures using peanut agglutinin (Sigma) (Sacks and Melby, 2001) while enrichment for *L. amazonensis* metacyclic promastigotes was performed by Ficoll density gradient centrifugation.

**Mouse infection**

Peritoneal macrophages were isolated from C57BL/6 mice (7 weeks, female) obtained from National Cancer Institute Charles River Laboratories by flushing the peritoneal cavity using 10 mL cold DPBS without calcium and magnesium. Cells were plated in 6 well plates in 2.5 mL DMEM/F12 media supplemented with 10% FBS, 1% penicillin/streptomycin, and 1% glutamine to a density of ~2.5 x $10^6$ cells/well and incubated overnight. Two hours after washing, cells were infected with Ficoll-enriched *L. major* metacyclic promastigotes at a ratio of 5 parasites per macrophage along with 5% C5-deficient serum collected from DBA2 mice. Cells were lysed using the Trizol® reagent (Invitrogen, CA) at 4, 24, 48, and 72 hours following infection.

**Human macrophages and infection**

Human macrophages were derived from purified monocytes and purchased fully differentiated from HemaCare (Van Nuys, CA). Briefly, $CD14^+$ monocytes were positively selected via immunomagnetic beads and were cultured in the presence of 1000 IU/mL human recombinant human macrophage colony-stimulating factor M-CSF for 10 days. The cells were then harvested and transferred into hypothermosol (BioLife Solution) and shipped in suspension. Upon arrival, approximately 2.5 x $10^6$ macrophages were plated per well in a 6-well plate using X-VIVO-15 medium (Lonza) and supplemented with 1000 IU/mL M-CSF (Miltenyi Biotech) and incubated for 24 hr at 37°C, 5% $CO_2$. Macrophages were infected using a ratio of 10 parasites per macrophage for *L. amazonensis* and 5 parasites per macrophage for *L. major* for 4 hr at 34°C. Infection was performed in X-VIVO-

15 medium supplemented with 4% human AB serum. Macrophages were allowed to ingest 4.35 micrometer polystyrene beads (Spherotech) at 1:10 ratio also in the presence of 4% human AB serum (beads were pre-incubated in AB serum for 30 min). The cells were washed 3 times with PBS and further incubated at 34°C until the 24, 48 and 72 hpi timepoints. Samples intended for counting were fixed for 5 min at room temperature with Bouin solution (71.4% saturated picric acid, 23.8% formaldehyde, and 4.8% acetic acid), stained with Giemsa, and sequentially dehydrated in acetone, followed by a graded series of acetone/xylol (9:1, 7:3, and 3:7) and, finally, xylol. The number of intracellular parasites was determined by counting the total macrophages and the total intracellular parasites per microscopic field using a Nikon E200 microscope with a 100X 1.3 N.A. oil immersion objective. Counts were performed in triplicate for each period of infection.

**RNA isolation and cDNA library preparation**

For the metacyclogenesis and mouse experiments, total RNA was isolated using the Trizol® reagent (Invitrogen, CA), treated with DNase, and purified using the Qiagen RNeasy mini kit. For the human experiments, total RNA was isolated from macrophages and the metacyclic material used for the *L. major* and *L. amazonensis* infections using a NucleoSpin RNA kit (Machery-Nagel) according to manufacturer's protocol. RNA integrity was assessed using an Agilent 2100 bioanalyzer. PolyA-enriched cDNA libraries were generated using the Illumina TruSeq Sample Preparation kit (San Diego, CA) and checked for quality and quantity using the bioanalyzer and qPCR (KAPA Biosystems).

**RNA-seq data generation, pre-processing, and quality trimming**

Paired end reads (100 bp) were obtained from the Illumina HiSeq 1500 platform. `Trimmomatic` (Bolger, 2014) was used to remove any remaining Illumina adapter sequences from reads and to trim bases off the start or the end of a read when the quality score fell below a threshold of 20. Sequence quality metrics were assessed using `FastQC` (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

**Mapping cDNA fragments to the reference genome, abundance estimation, and data normalization**

`TopHat` (v 2.0.10) (Trapnell, 2009) was used to align reads to the applicable genome(s) with each genome alignment performed independently. Reads from uninfected, *L. major*-infected, *L. amazonensis*-infected, and bead-containing macrophage samples were aligned to the human genome (v. hg19/GRCh37) or mouse genome (v. mm10). Reads from *L. major*-infected or *L. amazonensis*-infected samples were additionally aligned to the *L. major* (v. 6.0) or *L. mexicana* (v. 8.1) genomes, respectively, as were reads from procyclic and metacyclic promastigote samples. The *L. mexicana* genome was used for alignment of *L. amazonensis*-containing samples since it is the most closely related well-annotated genome that was available. The human and mouse genomes were obtained from the UCSC genome browser (http://genome.ucsc.edu) and the parasite genomes were obtained from the TriTrypDB database (www.tritrypdb.org) (Aslett, 2010). Two mismatches per read were allowed and reads were allowed to map only to a single locus. The abundance of reads

mapping to each gene feature in the TriTrypDB *L. major* annotation (v 6.0) was determined using `HTSeq` (Anders, 2014). Each resulting count table was restricted to protein-coding genes (8,486 genes for *L. major*, 8,336 genes for *L. amazonensis*/*L. mexicana*, 20,956 genes for human, and 23,100 genes for mouse). For the *L. major* metacyclogenesis analysis, a count table was also generated that included the above features along with a set of novel open reading frames (ORFs) of at least 90 nt in length. These novel ORFs were identified by manual annotation of translational evidence from a ribosome profiling study of *L. major* procyclic promastigote samples. Their coordinates are provided in Table 12.

**Data quality assessment by statistical sample clustering and visualization**

Multiple approaches were used to evaluate replicates and to visualize sample-sample distances. Those included Pearson correlation, median pairwise correlation analysis, box plots, Principal Component Analysis (PCA) and Euclidean distances-based hierarchical clustering. Samples that did not pass the following quality assessment procedure were excluded from downstream analyses. For each sample, the median pairwise correlation (mpc) to all other samples in the dataset was computed. A standard outlier identification method (Hoaglin, 1983) was applied to remove samples that have low correlation with the other samples. Samples were removed if their median mpc was less than Q1 (mpc) – 1.5 IQR (mpc) where Q1 (mpc) and IQR (mpc) represent the first quartile and inter-quartile range of the mpc across all samples, respectively.

**Differential expression analysis**

Non-expressed and weakly expressed genes, defined as having less than 1 read per million in n of the samples, where n is the size of the smallest group of replicates, were removed prior to differential expression analysis (Anders, 2013). A quantile normalization scheme was applied to all samples (Bolstad, 2003). Following log2 transformation of the data, `limma` (a Bioconductor package) was used to conduct differential expression analyses. `limma` utilizes a standard variance moderated across all genes using a Bayesian model and produces *P* values with greater degrees of freedom (Smyth, 2004). The `voom` module was used to transform the data based on observational level weights derived from the mean-variance relationship prior to statistical modeling (Law, 2014). Experimental batch effects were adjusted for by including experimental batch as a covariate in the statistical model (Leek, 2010). Pairwise contrasts were done within `limma` to identify differentially expressed (DE) genes across timepoints in the parasites and within and across timepoints in mouse macrophages and in human macrophages both with and without accounting for the effect of phagocytosis. DE genes were defined as genes with a Benjamini-Hochberg multiple-testing adjusted *P* value of $< 0.05$. For comparisons of uninfected vs. infected samples without accounting for the effects of phagocytosis, *P* values from a contrast of infected minus uninfected were used as input into BH multiple testing adjustment. To determine differential expression while accounting for phagocytosis, infected samples were evaluated relative to both bead-containing macrophages and uninfected control cells. For each gene, the maximum *P* value

from these two contrasts was selected for input into BH multiple testing adjustment.

### Ortholog mapping

Orthology mapping was done between human and mouse gene identifiers and between *L. major* and *L. mexicana* gene identifiers. Human-mouse orthologs were defined using the bioMaRt package in R (Durinck, 2005). Orthology gene tables from TriTrypDB (v 8.1) (Aslett, 2010) were used to match orthologs between *L. major* and *L. mexicana*.

### Gene ontology (GO) analysis

Enriched GO categories were identified using the `GOseq` package in R (Young, 2010). `GOseq` was developed specifically to account for transcript length bias in GO analyses using RNA-seq data. For each comparison, upregulated and downregulated gene sets (no fold change cut-off) were input separately into `GOseq`. A P value cut-off of 0.05 was used.

### KEGG pathway analysis

KEGG pathway analysis using ConsensusPathDB-mouse or ConsensusPathDB-human was done to identify signaling and metabolic pathways that were over-represented in the mouse and human DE gene lists, respectively. For each KEGG pathway, a *P* value was calculated using a hypergeometric test and a cutoff of 0.01 was applied to identify enriched KEGG pathways. Genes that were DE more than 2-fold in *L. major*-infected cells relative to controls at each

timepoint were used as input with up- and down-regulated genes considered separately.

### *Trans*-splicing site detection and 5´ UTR analysis

Sequences from each sample were mapped to the *L. major* genome (v. 6.0) using `TopHat` (v 2.0.10) (Trapnell, 2009). Only one mismatch per read was allowed and the paired reads were required to be mapped for an alignment to be reported. Reads that did not align to the genome were retained to form a pool of candidate spliced leader (SL)-containing reads. These reads were filtered to keep only those containing at least 4 bases of the end of the *L. major* SL sequence (AACTAACGCTATTATTGATACAGTTTCTGTACTATATTG) or its reverse complement. This target sequence (or its reverse complement) was trimmed from the reads and `TopHat` was used to align the remaining portions to the *L. major* genome. Two mismatches per read were allowed and reads were assigned only to a single locus of the gene model annotations provided to `TopHat` (containing previously annotated genes and novel ORFs). The alignment coordinates of the trimmed reads were used to retrieve the exact locations of the putative *trans*-splicing sites. The genomic sequence neighboring each putative site was compared against the portion of the read that was removed. Reads for which the trimmed portion (4-39 nt) did not differ by at least two bases from the corresponding genomic sequence were treated as false hits and discarded. Putative sites that were located within a previously annotated CDS (from TriTrypDB, version 6.0) (Aslett, 2010) or within a novel ORF (as defined above) or those with no such feature within 7500 nt downstream of the

site were excluded. *Trans*-splicing sites that remained were assigned to the nearest downstream feature. The length of the 5' UTR was defined as the distance between the *trans*-splicing site and the start of the CDS/ORF to which it was assigned. Splice acceptor sites were identified for each gene by extracting the dinucleotide sequence in the genome upstream of each detected *trans*-splicing site using a custom Python script. Sequence composition was plotted using `WebLogo` version 3.3 (Crooks, 2004). The *trans*-splicing site detection pipeline was written in Python and made use of the Ruffus pipeline software framework (Goodstadt, 2010) and Biopython library (Cock, 2009). Data visualization was done using `ggplot2` (Wickham, 2009).

**Polypyrimidine tract characterization**

A custom Python script was used to scan a window of 250 nt upstream of each primary *trans*-splicing site to identify the corresponding polypyrimidine (polyPy) tract. A polyPy tract was defined as the longest stretch of sequence consisting of pyrimidines, allowing interruption by no more than a single purine.

**Polyadenylation site detection and 3´ UTR analysis**

Identification of the polyadenylation sites was done using a process similar to the one used for *trans*-splicing site detection. The initial filtering step performed on unmapped reads identified reads containing at least 4 nt of thymine or at least 4 nt of adenine residues. This target sequence was trimmed from the reads and `TopHat` was used to align the remaining portions of the reads to the *L. major* genome. Two mismatches per read were allowed and reads were assigned only

to a single locus in the gene model annotations provided to `TopHat` (containing previously annotated genes and novel ORFs). The alignment coordinates of the trimmed reads were used to retrieve the exact locations of the putative polyadenylation sites. The sequence neighboring each putative site was compared against the portion of the read that was removed. Any reads for which the trimmed portion (4+ nt) did not differ by at least two bases from the corresponding genomic sequence were treated as false hits and discarded. Putative sites that were located within a previously annotated CDS (from TriTrypDB, version 6.0) or within a novel ORF (as defined above) or those with no such feature within 7500 nt upstream of the site were excluded. Polyadenylation sites that remained were assigned to the nearest downstream feature (CDS or novel ORF). The length of the 3' UTR was defined as the distance between the stop of the CDS/ORF and the polyadenylation site. Sequence composition was plotted using `WebLogo` version 3.3 (Crooks, 2004). The polyadenylation site detection pipeline was written in Python and made use of the Ruffus pipeline software framework (Goodstadt, 2010) and Biopython library (Cock, 2009). Data visualization was done using `ggplot2` (Wickham, 2009).

**Alternative RNA processing site analysis**

Counts of *trans*-splicing sites or polyadenylation sites were combined from biological replicates for each developmental stage (procyclic promastigotes and metacyclic promastigotes). The site with the largest number of reads mapped was defined as “primary” for each of the developmental stages. All other sites

were considered to be “minor” with the most utilized of the minor sites designated as the “secondary” site. The ratio of reads mapping to the primary site to those mapping to the secondary site (P/S) for a given gene was used to determine the dominance (preference) of the primary site for that gene.

**Data access**

Sequence data are available at the NCBI Sequence Read Archive (SRA) under accession numbers SRR1460763-SRR1460775, SRR1460724-SRR1460747, SRR1460767, SRR1460772, SRR2136702, SRR2136703, SRR2136708, SRR2136720, SRR2136722, SRR2155070, SRR2155072- SRR2155075, SRR2155078, SRR2155082, SRR2155085, SRR2155101, SRR2155105, SRR2155143, SRR2155160-SRR2155166, SRR2156106-SRR2156110, SRR2156117, SRR2156270-SRR2156274, SRR2156854-SRR2156856, SRR2156858-SRR2156860, SRR2156862-SRR2156863, SRR2163237-SRR2163240, SRR2163242, SRR2163251, SRR2163256, SRR2163270, SRR2163272-SRR2163276, SRR2163278-SRR2163283, SRR2163285-SRR2163287, SRR2163289-SRR2163292, SRR2163399-SRR2163401, SRR2171252, SRR2171252-SRR2171253, and SRR2171255. All components of the data quality assessment statistical pipeline, named `cbcbSEQ`, were done in R and can be accessed on GitHub (https://github.com/kokrah/cbcbSEQ/). The code used for the *trans*-splicing and polyadenylation pipelines is freely available at https://github.com/elsayed-lab/utr_analysis. The code used to determine alternative RNA processing sites is available at https://github.com/elsayed-lab/lmajor_alternate_acceptor_site_usage.

# Appendices

## Appendix 1: Procyclic-metacyclic experimental design

| ID for this Manuscript | Sample ID | SRA accession number | Promast stage (enrichment method) | Batch | Number of reads that pass Illumina filter |
|---|---|---|---|---|---|
| 1 | HPGL0075 | n/a | procyclic | A | 62,051,890 |
| 2 | HPGL0076 | n/a | metacyclic (Ficoll) | A | 52,660,754 |
| 3 | HPGL0096 | SRR1460763 | procyclic | B | 103,466,044 |
| 4 | HPGL0097 | SRR1460764 | metacyclic (PNA) | B | 76,253,690 |
| 5 | HPGL0098 | SRR1460765 | metacyclic (Ficoll) | B | 93,319,752 |
| 6 | HPGL0164 | SRR1460766 | procyclic | C | 46,155,070 |
| 7 | HPGL0165 | SRR1460767 | metacyclic (Ficoll) | C | 45,492,872 |
| 8 | HPGL0192 | SRR1460768 | procyclic | D | 64,505,484 |
| 9 | HPGL0193 | SRR1460769 | metacyclic (Ficoll) | D | 70,178,176 |
| 10 | HPGL0228 | SRR1460770 | procyclic | E | 105,948,882 |
| 11 | HPGL0229 | SRR1460771 | metacyclic (PNA) | E | 77,161,294 |
| 12 | HPGL0230 | SRR1460772 | metacyclic (Ficoll) | E | 84,056,646 |
| 13 | HPGL0324 | SRR1460773 | procyclic | F | 69,215,652 |
| 14 | HPGL0325 | SRR1460774 | metacyclic (PNA) | F | 64,195,828 |
| 15 | HPGL0326 | SRR1460775 | metacyclic (Ficoll) | F | 60,167,474 |
| | | | | | |
| Total | | | | | 1,074,829,508 |

| Number of reads mapped | % of reads mapped | Number of SL-containing reads | % of reads containing SL sequence | Number of polyA-containing reads | % of reads containing polyA sequence |
|---|---|---|---|---|---|
| 55,352,314 | 89.20 | n/a | n/a | n/a | n/a |
| 47,712,265 | 90.60 | n/a | n/a | n/a | n/a |
| 92,748,560 | 89.64 | 4,095,395 | 3.96 | 58,064 | 0.06 |
| 69,557,423 | 91.22 | 2,658,344 | 3.49 | 56,582 | 0.07 |
| 84,580,995 | 90.64 | 3,634,321 | 3.89 | 47,581 | 0.05 |
| 42,801,506 | 92.73 | 1,772,321 | 3.84 | 39,718 | 0.09 |
| 41,811,082 | 91.91 | 1,360,561 | 2.99 | 44,785 | 0.10 |
| 58,857,640 | 91.24 | 2,359,763 | 3.66 | 43,034 | 0.07 |
| 63,898,517 | 91.05 | 2,439,369 | 3.48 | 48,044 | 0.07 |
| 98,120,201 | 92.61 | 3,252,367 | 3.07 | 49,416 | 0.05 |
| 71,310,256 | 92.42 | 2,645,944 | 3.43 | 41,414 | 0.05 |
| 77,364,378 | 92.04 | 3,034,977 | 3.61 | 41,102 | 0.05 |
| 63,766,203 | 92.13 | 3,752,380 | 5.42 | 13,684 | 0.02 |
| 59,048,832 | 91.98 | 3,355,270 | 5.23 | 13,606 | 0.02 |
| 55,318,601 | 91.94 | 3,064,443 | 5.09 | 12,270 | 0.02 |
| | | | | | |
| 982,248,773 | 91.39 | 37,425,455 | 3.90 | 509,300 | 0.05 |

**Table 11: Procyclic-metacyclic experimental design.**

Samples are listed using an internal lab sample identifier (HPGL----), which is referenced in the record stored at the Short Read Archive (SRA; accession numbers provided), and an additional simple identifier (1-15) for clarity in the text and figures. Experimental batches (A-F) are defined based on the start date of the experiment with each batch originating from a separate growth of cells. The number of reads sequenced, number and percentage of reads mapping to the *L. major* genome (v. 6.0), number and percentage of reads containing evidence of the SL sequence, and number and percentage of reads containing evidence of polyadenylation are also included.

# Appendix 2: Coordinates of novel ORFs

## Table 12: Coordinates of Novel ORFs.

Novel open reading frames of at least 90 nucleotides in length were identified by manual annotation of translational evidence from a ribosome profiling study of *L. major* procyclic promastigote samples.

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.01_108 | LmjF.01 | 75300 | 75404 | - |
| LmjF.01_20 | LmjF.01 | 110681 | 111175 | + |
| LmjF.01_22 | LmjF.01 | 114253 | 114546 | + |
| LmjF.01_28 | LmjF.01 | 136265 | 136417 | + |
| LmjF.01_30 | LmjF.01 | 142049 | 142294 | + |
| LmjF.01_34 | LmjF.01 | 153746 | 153916 | + |
| LmjF.01_49 | LmjF.01 | 197108 | 197224 | + |
| LmjF.01_50 | LmjF.01 | 198315 | 198512 | + |
| LmjF.01_51 | LmjF.01 | 198636 | 198767 | + |
| LmjF.01_54 | LmjF.01 | 204871 | 205035 | + |
| LmjF.01_55 | LmjF.01 | 204888 | 206231 | + |
| LmjF.01_67 | LmjF.01 | 243194 | 243376 | + |
| LmjF.01_68 | LmjF.01 | 243850 | 244035 | + |
| LmjF.02_121 | LmjF.02 | 336090 | 336365 | + |
| LmjF.02_126 | LmjF.02 | 349704 | 349862 | + |
| LmjF.02_128 | LmjF.02 | 2868 | 3155 | - |
| LmjF.02_129 | LmjF.02 | 2741 | 2896 | - |
| LmjF.02_134 | LmjF.02 | 16013 | 16486 | - |
| LmjF.02_139 | LmjF.02 | 23757 | 23873 | - |
| LmjF.02_150 | LmjF.02 | 73330 | 73431 | - |
| LmjF.02_167 | LmjF.02 | 145981 | 146163 | - |
| LmjF.02_169 | LmjF.02 | 149125 | 149301 | - |
| LmjF.02_174 | LmjF.02 | 164117 | 164605 | - |
| LmjF.03_220 | LmjF.03 | 62613 | 62861 | + |
| LmjF.03_230 | LmjF.03 | 105885 | 106739 | + |
| LmjF.03_242 | LmjF.03 | 152208 | 152321 | + |
| LmjF.03_244 | LmjF.03 | 160619 | 161200 | + |
| LmjF.03_252 | LmjF.03 | 193514 | 193633 | + |
| LmjF.03_260 | LmjF.03 | 225647 | 225829 | + |
| LmjF.03_278 | LmjF.03 | 288492 | 288599 | - |
| LmjF.03_294 | LmjF.03 | 363847 | 364215 | - |
| LmjF.04_338 | LmjF.04 | 105572 | 105733 | + |
| LmjF.04_351 | LmjF.04 | 170438 | 170656 | - |
| LmjF.04_353 | LmjF.04 | 178490 | 178645 | - |
| LmjF.04_360 | LmjF.04 | 200556 | 200759 | - |
| LmjF.04_368 | LmjF.04 | 238083 | 238238 | - |
| LmjF.04_381 | LmjF.04 | 277315 | 277410 | - |
| LmjF.04_400 | LmjF.04 | 342310 | 342519 | - |
| LmjF.04_410 | LmjF.04 | 372683 | 372889 | - |
| LmjF.04_421 | LmjF.04 | 397166 | 397357 | - |
| LmjF.04_440 | LmjF.04 | 457136 | 457396 | - |
| LmjF.05_448 | LmjF.05 | 10403 | 10681 | + |
| LmjF.05_449 | LmjF.05 | 11224 | 11319 | + |
| LmjF.05_454 | LmjF.05 | 23011 | 23118 | + |
| LmjF.05_470 | LmjF.05 | 70497 | 70976 | + |
| LmjF.05_472 | LmjF.05 | 77111 | 77257 | + |
| LmjF.05_473 | LmjF.05 | 77335 | 77466 | + |
| LmjF.05_475 | LmjF.05 | 81864 | 82109 | + |
| LmjF.05_480 | LmjF.05 | 88660 | 88932 | + |
| LmjF.05_492 | LmjF.05 | 119661 | 120101 | + |
| LmjF.05_497 | LmjF.05 | 138843 | 138968 | + |
| LmjF.05_504 | LmjF.05 | 156787 | 157500 | + |
| LmjF.05_529 | LmjF.05 | 247455 | 247832 | + |
| LmjF.05_532 | LmjF.05 | 263317 | 263409 | + |
| LmjF.05_538 | LmjF.05 | 285271 | 285750 | + |
| LmjF.05_541 | LmjF.05 | 292735 | 293013 | + |
| LmjF.05_547 | LmjF.05 | 306211 | 306378 | + |
| LmjF.05_561 | LmjF.05 | 358860 | 359045 | + |
| LmjF.05_574 | LmjF.05 | 396433 | 396798 | + |
| LmjF.05_577 | LmjF.05 | 405918 | 406031 | + |
| LmjF.05_579 | LmjF.05 | 411432 | 411776 | + |
| LmjF.05_583 | LmjF.05 | 423145 | 423273 | - |
| LmjF.05_584 | LmjF.05 | 424249 | 424743 | - |
| LmjF.05_592 | LmjF.05 | 461866 | 461991 | + |
| LmjF.06_598 | LmjF.06 | 16520 | 16744 | - |
| LmjF.06_615 | LmjF.06 | 68976 | 69209 | - |
| LmjF.06_624 | LmjF.06 | 95339 | 96148 | - |
| LmjF.06_636 | LmjF.06 | 131768 | 131908 | + |
| LmjF.06_648 | LmjF.06 | 169713 | 169913 | + |
| LmjF.06_653 | LmjF.06 | 181024 | 181203 | + |
| LmjF.06_665 | LmjF.06 | 221059 | 221355 | + |
| LmjF.06_677 | LmjF.06 | 277711 | 277929 | + |
| LmjF.06_683 | LmjF.06 | 301522 | 301737 | + |
| LmjF.06_686 | LmjF.06 | 326137 | 326256 | + |
| LmjF.06_730 | LmjF.06 | 479254 | 479595 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.06_738 | LmjF.06 | 497381 | 497902 | + |
| LmjF.06_742 | LmjF.06 | 506541 | 506759 | - |
| LmjF.06_744 | LmjF.06 | 510753 | 510905 | + |
| LmjF.07_746 | LmjF.07 | 2243 | 2380 | - |
| LmjF.07_752 | LmjF.07 | 20854 | 21009 | + |
| LmjF.07_758 | LmjF.07 | 38407 | 38625 | + |
| LmjF.07_759 | LmjF.07 | 38501 | 38767 | + |
| LmjF.07_766 | LmjF.07 | 64560 | 64787 | - |
| LmjF.07_811 | LmjF.07 | 243104 | 243364 | + |
| LmjF.07_814 | LmjF.07 | 252695 | 252793 | + |
| LmjF.07_827 | LmjF.07 | 306747 | 306842 | + |
| LmjF.07_837 | LmjF.07 | 340639 | 340980 | + |
| LmjF.07_849 | LmjF.07 | 386858 | 387064 | + |
| LmjF.07_854 | LmjF.07 | 429773 | 429907 | + |
| LmjF.07_855 | LmjF.07 | 430042 | 430587 | + |
| LmjF.07_859 | LmjF.07 | 453590 | 453724 | + |
| LmjF.07_863 | LmjF.07 | 474591 | 475604 | + |
| LmjF.07_868 | LmjF.07 | 506546 | 506662 | + |
| LmjF.07_880 | LmjF.07 | 543620 | 543928 | + |
| LmjF.07_891 | LmjF.07 | 566057 | 566206 | + |
| LmjF.07_894 | LmjF.07 | 573672 | 573800 | + |
| LmjF.07_897 | LmjF.07 | 587130 | 587240 | + |
| LmjF.07_898 | LmjF.07 | 587185 | 587358 | + |
| LmjF.07_900 | LmjF.07 | 593066 | 593353 | + |
| LmjF.07_901 | LmjF.07 | 593247 | 593480 | + |
| LmjF.08_1008 | LmjF.08 | 392394 | 392492 | + |
| LmjF.08_1020 | LmjF.08 | 436571 | 436855 | - |
| LmjF.08_1033 | LmjF.08 | 492798 | 493277 | + |
| LmjF.08_1035 | LmjF.08 | 495251 | 495427 | + |
| LmjF.08_1040 | LmjF.08 | 507408 | 507536 | + |
| LmjF.08_1043 | LmjF.08 | 516911 | 517069 | + |
| LmjF.08_1048 | LmjF.08 | 531238 | 531375 | + |
| LmjF.08_1049 | LmjF.08 | 531368 | 531469 | + |
| LmjF.08_1050 | LmjF.08 | 531402 | 531623 | + |
| LmjF.08_1056 | LmjF.08 | 541834 | 541941 | + |
| LmjF.08_1057 | LmjF.08 | 542811 | 544118 | + |
| LmjF.08_1058 | LmjF.08 | 544414 | 544515 | + |
| LmjF.08_1062 | LmjF.08 | 554975 | 555178 | + |
| LmjF.08_921 | LmjF.08 | 58067 | 58369 | + |
| LmjF.08_922 | LmjF.08 | 58676 | 58999 | + |
| LmjF.08_923 | LmjF.08 | 60807 | 61148 | + |
| LmjF.08_940 | LmjF.08 | 111881 | 112120 | + |
| LmjF.08_948 | LmjF.08 | 154201 | 154308 | + |
| LmjF.08_950 | LmjF.08 | 164345 | 164452 | + |
| LmjF.08_965 | LmjF.08 | 218722 | 218904 | + |
| LmjF.08_973 | LmjF.08 | 238798 | 238965 | + |
| LmjF.08_976 | LmjF.08 | 257665 | 257898 | + |
| LmjF.08_978 | LmjF.08 | 266223 | 266663 | + |
| LmjF.08_981 | LmjF.08 | 274422 | 274601 | + |
| LmjF.08_984 | LmjF.08 | 289943 | 290044 | + |
| LmjF.08_986 | LmjF.08 | 295799 | 295927 | + |
| LmjF.08_988 | LmjF.08 | 301711 | 301815 | + |
| LmjF.08_997 | LmjF.08 | 352965 | 353087 | + |
| LmjF.09_1073 | LmjF.09 | 6299 | 6484 | + |
| LmjF.09_1082 | LmjF.09 | 28224 | 28955 | + |
| LmjF.09_1084 | LmjF.09 | 40338 | 40820 | + |
| LmjF.09_1085 | LmjF.09 | 40951 | 41043 | + |
| LmjF.09_1088 | LmjF.09 | 53225 | 53326 | + |
| LmjF.09_1090 | LmjF.09 | 56433 | 56549 | + |
| LmjF.09_1102 | LmjF.09 | 78194 | 78409 | + |
| LmjF.09_1103 | LmjF.09 | 78409 | 78567 | + |
| LmjF.09_1121 | LmjF.09 | 124955 | 125314 | + |
| LmjF.09_1132 | LmjF.09 | 155957 | 156151 | + |
| LmjF.09_1135 | LmjF.09 | 167614 | 167709 | + |
| LmjF.09_1142 | LmjF.09 | 196413 | 196610 | + |
| LmjF.09_1144 | LmjF.09 | 199072 | 199257 | + |
| LmjF.09_1150 | LmjF.09 | 220336 | 220521 | + |
| LmjF.09_1170 | LmjF.09 | 280086 | 280292 | - |
| LmjF.09_1177 | LmjF.09 | 317221 | 317340 | - |
| LmjF.09_1179 | LmjF.09 | 321205 | 321407 | - |
| LmjF.09_1183 | LmjF.09 | 333011 | 333175 | - |
| LmjF.09_1204 | LmjF.09 | 385913 | 386071 | - |
| LmjF.09_1207 | LmjF.09 | 407194 | 407952 | + |
| LmjF.09_1209 | LmjF.09 | 416139 | 416408 | + |
| LmjF.09_1210 | LmjF.09 | 416332 | 416523 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.09_1250 | LmjF.09 | 536191 | 536463 | - |
| LmjF.09_1251 | LmjF.09 | 536444 | 536662 | - |
| LmjF.09_1255 | LmjF.09 | 545381 | 545656 | - |
| LmjF.10_1273 | LmjF.10 | 15652 | 15753 | - |
| LmjF.10_1286 | LmjF.10 | 68002 | 68100 | + |
| LmjF.10_1299 | LmjF.10 | 110871 | 111002 | + |
| LmjF.10_1307 | LmjF.10 | 133359 | 133520 | + |
| LmjF.10_1318 | LmjF.10 | 182540 | 183127 | + |
| LmjF.10_1324 | LmjF.10 | 206493 | 206693 | + |
| LmjF.10_1326 | LmjF.10 | 211877 | 211990 | + |
| LmjF.10_1331 | LmjF.10 | 233035 | 233322 | + |
| LmjF.10_1333 | LmjF.10 | 241236 | 241328 | + |
| LmjF.10_1352 | LmjF.10 | 329808 | 329981 | - |
| LmjF.10_1355 | LmjF.10 | 339436 | 341496 | - |
| LmjF.10_1358 | LmjF.10 | 351210 | 351398 | - |
| LmjF.10_1362 | LmjF.10 | 365021 | 365356 | - |
| LmjF.10_1363 | LmjF.10 | 365970 | 366203 | - |
| LmjF.10_1386 | LmjF.10 | 448939 | 449037 | - |
| LmjF.10_1400 | LmjF.10 | 488996 | 489322 | - |
| LmjF.10_1404 | LmjF.10 | 499609 | 499710 | - |
| LmjF.10_1407 | LmjF.10 | 507063 | 507278 | - |
| LmjF.10_1411 | LmjF.10 | 513365 | 513508 | - |
| LmjF.10_1412 | LmjF.10 | 512495 | 512656 | - |
| LmjF.10_1417 | LmjF.10 | 526964 | 527110 | - |
| LmjF.10_1420 | LmjF.10 | 530183 | 530386 | - |
| LmjF.10_1421 | LmjF.10 | 530897 | 531301 | - |
| LmjF.10_1433 | LmjF.10 | 558882 | 559106 | + |
| LmjF.11_1460 | LmjF.11 | 58477 | 58683 | + |
| LmjF.11_1461 | LmjF.11 | 58757 | 58861 | + |
| LmjF.11_1468 | LmjF.11 | 76083 | 76637 | + |
| LmjF.11_1477 | LmjF.11 | 112451 | 112552 | + |
| LmjF.11_1482 | LmjF.11 | 126347 | 126649 | + |
| LmjF.11_1495 | LmjF.11 | 177970 | 178116 | + |
| LmjF.11_1500 | LmjF.11 | 192857 | 193102 | + |
| LmjF.11_1501 | LmjF.11 | 193293 | 193394 | + |
| LmjF.11_1503 | LmjF.11 | 202072 | 202290 | + |
| LmjF.11_1505 | LmjF.11 | 207549 | 207863 | + |
| LmjF.11_1506 | LmjF.11 | 210393 | 210596 | + |
| LmjF.11_1509 | LmjF.11 | 218361 | 218684 | + |
| LmjF.11_1511 | LmjF.11 | 232934 | 233074 | + |
| LmjF.11_1532 | LmjF.11 | 314026 | 314268 | + |
| LmjF.11_1537 | LmjF.11 | 327389 | 327706 | + |
| LmjF.11_1538 | LmjF.11 | 328201 | 328293 | + |
| LmjF.11_1540 | LmjF.11 | 335098 | 335211 | + |
| LmjF.11_1547 | LmjF.11 | 362870 | 364081 | + |
| LmjF.11_1561 | LmjF.11 | 412194 | 412361 | + |
| LmjF.11_1565 | LmjF.11 | 427564 | 427860 | + |
| LmjF.11_1566 | LmjF.11 | 427628 | 427981 | + |
| LmjF.11_1567 | LmjF.11 | 433530 | 433751 | + |
| LmjF.11_1570 | LmjF.11 | 447364 | 447957 | + |
| LmjF.11_1574 | LmjF.11 | 467632 | 467847 | + |
| LmjF.11_1582 | LmjF.11 | 486870 | 487418 | + |
| LmjF.11_1605 | LmjF.11 | 578324 | 578545 | + |
| LmjF.12_1607 | LmjF.12 | 1861 | 2577 | + |
| LmjF.12_1608 | LmjF.12 | 2084 | 2218 | + |
| LmjF.12_1639 | LmjF.12 | 90060 | 90188 | + |
| LmjF.12_1644 | LmjF.12 | 111234 | 111479 | + |
| LmjF.12_1649 | LmjF.12 | 137706 | 137825 | + |
| LmjF.12_1655 | LmjF.12 | 161139 | 161375 | + |
| LmjF.12_1662 | LmjF.12 | 183196 | 183294 | - |
| LmjF.12_1667 | LmjF.12 | 210150 | 210257 | - |
| LmjF.12_1671 | LmjF.12 | 235040 | 235255 | - |
| LmjF.12_1672 | LmjF.12 | 245380 | 245595 | - |
| LmjF.12_1676 | LmjF.12 | 292149 | 292439 | + |
| LmjF.12_1681 | LmjF.12 | 303405 | 303575 | + |
| LmjF.12_1684 | LmjF.12 | 319163 | 319357 | + |
| LmjF.12_1702 | LmjF.12 | 376698 | 378473 | + |
| LmjF.12_1710 | LmjF.12 | 406846 | 407169 | + |
| LmjF.12_1711 | LmjF.12 | 409010 | 409918 | + |
| LmjF.12_1712 | LmjF.12 | 409661 | 409918 | + |
| LmjF.12_1730 | LmjF.12 | 473486 | 473914 | + |
| LmjF.12_1734 | LmjF.12 | 486488 | 486895 | + |
| LmjF.12_1742 | LmjF.12 | 517518 | 517748 | + |
| LmjF.12_1744 | LmjF.12 | 524129 | 524536 | + |
| LmjF.12_1755 | LmjF.12 | 571648 | 571755 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.12_1770 | LmjF.12 | 636127 | 636408 | + |
| LmjF.12_1783 | LmjF.12 | 672560 | 673522 | + |
| LmjF.13_1792 | LmjF.13 | 21697 | 21840 | - |
| LmjF.13_1798 | LmjF.13 | 34292 | 34732 | - |
| LmjF.13_1799 | LmjF.13 | 35607 | 35783 | - |
| LmjF.13_1806 | LmjF.13 | 56869 | 57240 | - |
| LmjF.13_1810 | LmjF.13 | 69781 | 70047 | - |
| LmjF.13_1814 | LmjF.13 | 78532 | 78732 | - |
| LmjF.13_1834 | LmjF.13 | 140813 | 141076 | - |
| LmjF.13_1838 | LmjF.13 | 152824 | 152946 | + |
| LmjF.13_1839 | LmjF.13 | 153353 | 153445 | + |
| LmjF.13_1845 | LmjF.13 | 171075 | 171257 | + |
| LmjF.13_1846 | LmjF.13 | 171578 | 171685 | + |
| LmjF.13_1847 | LmjF.13 | 171733 | 171867 | + |
| LmjF.13_1848 | LmjF.13 | 172427 | 172615 | + |
| LmjF.13_1854 | LmjF.13 | 184247 | 184690 | + |
| LmjF.13_1860 | LmjF.13 | 198656 | 199069 | + |
| LmjF.13_1905 | LmjF.13 | 359020 | 359310 | - |
| LmjF.13_1912 | LmjF.13 | 393022 | 393300 | - |
| LmjF.13_1922 | LmjF.13 | 413593 | 413898 | - |
| LmjF.13_1926 | LmjF.13 | 425323 | 425577 | - |
| LmjF.13_1931 | LmjF.13 | 434985 | 435446 | - |
| LmjF.13_1932 | LmjF.13 | 437047 | 437265 | - |
| LmjF.13_1933 | LmjF.13 | 437435 | 437569 | - |
| LmjF.13_1935 | LmjF.13 | 443405 | 443560 | - |
| LmjF.13_1936 | LmjF.13 | 443253 | 443345 | - |
| LmjF.13_1937 | LmjF.13 | 443060 | 443209 | - |
| LmjF.13_1945 | LmjF.13 | 469909 | 470046 | - |
| LmjF.13_1953 | LmjF.13 | 503627 | 503923 | - |
| LmjF.13_1964 | LmjF.13 | 551601 | 551711 | + |
| LmjF.13_1966 | LmjF.13 | 558691 | 558789 | - |
| LmjF.13_1971 | LmjF.13 | 572816 | 572983 | - |
| LmjF.13_1974 | LmjF.13 | 582146 | 582313 | - |
| LmjF.13_1977 | LmjF.13 | 591645 | 591824 | - |
| LmjF.14_1991 | LmjF.14 | 4457 | 4567 | + |
| LmjF.14_1996 | LmjF.14 | 16816 | 17031 | + |
| LmjF.14_2035 | LmjF.14 | 134974 | 135123 | + |
| LmjF.14_2043 | LmjF.14 | 169522 | 169749 | - |
| LmjF.14_2045 | LmjF.14 | 177135 | 177272 | - |
| LmjF.14_2053 | LmjF.14 | 203333 | 203581 | - |
| LmjF.14_2054 | LmjF.14 | 204098 | 204880 | - |
| LmjF.14_2057 | LmjF.14 | 216683 | 216853 | - |
| LmjF.14_2062 | LmjF.14 | 226190 | 226387 | - |
| LmjF.14_2067 | LmjF.14 | 239895 | 240128 | - |
| LmjF.14_2072 | LmjF.14 | 263428 | 263571 | - |
| LmjF.14_2073 | LmjF.14 | 262933 | 263070 | + |
| LmjF.14_2078 | LmjF.14 | 285571 | 285711 | - |
| LmjF.14_2083 | LmjF.14 | 297012 | 297191 | - |
| LmjF.14_2091 | LmjF.14 | 341241 | 341378 | - |
| LmjF.14_2096 | LmjF.14 | 363090 | 363314 | - |
| LmjF.14_2103 | LmjF.14 | 393681 | 393944 | - |
| LmjF.14_2106 | LmjF.14 | 404441 | 404647 | - |
| LmjF.14_2113 | LmjF.14 | 418425 | 418577 | - |
| LmjF.14_2114 | LmjF.14 | 419811 | 420311 | - |
| LmjF.14_2122 | LmjF.14 | 454237 | 454350 | + |
| LmjF.14_2125 | LmjF.14 | 476695 | 476913 | + |
| LmjF.14_2126 | LmjF.14 | 476695 | 476913 | + |
| LmjF.14_2129 | LmjF.14 | 492510 | 492827 | + |
| LmjF.14_2154 | LmjF.14 | 569202 | 569417 | + |
| LmjF.14_2159 | LmjF.14 | 580132 | 580305 | + |
| LmjF.14_2161 | LmjF.14 | 586647 | 586817 | + |
| LmjF.14_2162 | LmjF.14 | 587062 | 587340 | + |
| LmjF.15_2179 | LmjF.15 | 23840 | 24139 | - |
| LmjF.15_2210 | LmjF.15 | 122991 | 123191 | + |
| LmjF.15_2248 | LmjF.15 | 291511 | 291942 | + |
| LmjF.15_2260 | LmjF.15 | 343961 | 344272 | - |
| LmjF.15_2262 | LmjF.15 | 352004 | 353059 | - |
| LmjF.15_2264 | LmjF.15 | 362507 | 362671 | - |
| LmjF.15_2269 | LmjF.15 | 378222 | 378407 | - |
| LmjF.15_2280 | LmjF.15 | 419770 | 419931 | - |
| LmjF.15_2283 | LmjF.15 | 426044 | 426226 | - |
| LmjF.15_2285 | LmjF.15 | 429383 | 429544 | - |
| LmjF.15_2298 | LmjF.15 | 458033 | 458314 | - |
| LmjF.15_2303 | LmjF.15 | 463752 | 464204 | - |
| LmjF.15_2319 | LmjF.15 | 512813 | 513031 | - |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.15_2325 | LmjF.15 | 531727 | 531924 | - |
| LmjF.15_2355 | LmjF.15 | 608359 | 608547 | - |
| LmjF.16_2373 | LmjF.16 | 21974 | 22225 | - |
| LmjF.16_2385 | LmjF.16 | 71928 | 72065 | - |
| LmjF.16_2390 | LmjF.16 | 86236 | 86379 | - |
| LmjF.16_2433 | LmjF.16 | 231395 | 231847 | - |
| LmjF.16_2445 | LmjF.16 | 277015 | 277146 | - |
| LmjF.16_2446 | LmjF.16 | 277279 | 277476 | - |
| LmjF.16_2474 | LmjF.16 | 400452 | 400712 | + |
| LmjF.16_2477 | LmjF.16 | 413775 | 413999 | + |
| LmjF.16_2481 | LmjF.16 | 421938 | 422372 | + |
| LmjF.16_2483 | LmjF.16 | 429096 | 430079 | + |
| LmjF.16_2493 | LmjF.16 | 465231 | 465347 | - |
| LmjF.16_2504 | LmjF.16 | 520917 | 521219 | - |
| LmjF.16_2531 | LmjF.16 | 641441 | 641533 | - |
| LmjF.16_2533 | LmjF.16 | 647237 | 647569 | - |
| LmjF.16_2534 | LmjF.16 | 647237 | 647569 | - |
| LmjF.16_2550 | LmjF.16 | 711667 | 712083 | + |
| LmjF.17_2552 | LmjF.17 | 13059 | 13424 | - |
| LmjF.17_2563 | LmjF.17 | 30495 | 30647 | - |
| LmjF.17_2585 | LmjF.17 | 123229 | 123327 | - |
| LmjF.17_2609 | LmjF.17 | 239997 | 240203 | - |
| LmjF.17_2616 | LmjF.17 | 275946 | 276266 | - |
| LmjF.17_2650 | LmjF.17 | 409339 | 409566 | - |
| LmjF.17_2655 | LmjF.17 | 418442 | 418642 | + |
| LmjF.17_2659 | LmjF.17 | 423627 | 423884 | + |
| LmjF.17_2664 | LmjF.17 | 446335 | 446535 | + |
| LmjF.17_2668 | LmjF.17 | 459865 | 460029 | + |
| LmjF.17_2671 | LmjF.17 | 470319 | 470570 | + |
| LmjF.17_2676 | LmjF.17 | 503959 | 504141 | + |
| LmjF.17_2682 | LmjF.17 | 528322 | 528600 | + |
| LmjF.17_2700 | LmjF.17 | 594656 | 595075 | + |
| LmjF.17_2714 | LmjF.17 | 636615 | 636785 | + |
| LmjF.17_2717 | LmjF.17 | 639649 | 639774 | + |
| LmjF.17_2725 | LmjF.17 | 663858 | 664577 | + |
| LmjF.18_2733 | LmjF.18 | 5454 | 6002 | - |
| LmjF.18_2747 | LmjF.18 | 44260 | 44556 | - |
| LmjF.18_2753 | LmjF.18 | 96600 | 96743 | - |
| LmjF.18_2761 | LmjF.18 | 119694 | 119888 | - |
| LmjF.18_2765 | LmjF.18 | 132763 | 132924 | - |
| LmjF.18_2781 | LmjF.18 | 191537 | 191986 | - |
| LmjF.18_2790 | LmjF.18 | 215800 | 215988 | - |
| LmjF.18_2798 | LmjF.18 | 247137 | 248000 | + |
| LmjF.18_2801 | LmjF.18 | 254683 | 254943 | + |
| LmjF.18_2802 | LmjF.18 | 256220 | 256822 | + |
| LmjF.18_2814 | LmjF.18 | 300800 | 300898 | + |
| LmjF.18_2846 | LmjF.18 | 414571 | 414783 | + |
| LmjF.18_2850 | LmjF.18 | 425281 | 425604 | + |
| LmjF.18_2865 | LmjF.18 | 492933 | 493067 | + |
| LmjF.18_2874 | LmjF.18 | 520283 | 520417 | + |
| LmjF.18_2885 | LmjF.18 | 553402 | 553500 | + |
| LmjF.18_2889 | LmjF.18 | 572618 | 572896 | + |
| LmjF.18_2892 | LmjF.18 | 579167 | 579331 | + |
| LmjF.18_2893 | LmjF.18 | 582458 | 583801 | + |
| LmjF.18_2894 | LmjF.18 | 585697 | 586041 | + |
| LmjF.18_2917 | LmjF.18 | 691562 | 691738 | + |
| LmjF.18_2930 | LmjF.18 | 722127 | 722258 | + |
| LmjF.19_2950 | LmjF.19 | 32411 | 32698 | - |
| LmjF.19_2951 | LmjF.19 | 32214 | 32498 | - |
| LmjF.19_2957 | LmjF.19 | 54517 | 54966 | - |
| LmjF.19_2965 | LmjF.19 | 81284 | 81529 | + |
| LmjF.19_2979 | LmjF.19 | 158573 | 158887 | + |
| LmjF.19_2997 | LmjF.19 | 223005 | 223277 | + |
| LmjF.19_2999 | LmjF.19 | 230087 | 230230 | + |
| LmjF.19_3000 | LmjF.19 | 230277 | 230423 | + |
| LmjF.19_3004 | LmjF.19 | 263059 | 263406 | + |
| LmjF.19_3014 | LmjF.19 | 304932 | 305084 | + |
| LmjF.19_3017 | LmjF.19 | 315264 | 315431 | + |
| LmjF.19_3018 | LmjF.19 | 315530 | 315625 | + |
| LmjF.19_3019 | LmjF.19 | 316267 | 316482 | + |
| LmjF.19_3024 | LmjF.19 | 325082 | 325177 | + |
| LmjF.19_3042 | LmjF.19 | 360014 | 360202 | + |
| LmjF.19_3047 | LmjF.19 | 366770 | 366943 | + |
| LmjF.19_3054 | LmjF.19 | 382655 | 382816 | + |
| LmjF.19_3059 | LmjF.19 | 395719 | 395889 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.19_3062 | LmjF.19 | 407105 | 407311 | + |
| LmjF.19_3063 | LmjF.19 | 406987 | 407484 | + |
| LmjF.19_3075 | LmjF.19 | 466828 | 466989 | + |
| LmjF.19_3088 | LmjF.19 | 529105 | 529260 | + |
| LmjF.19_3098 | LmjF.19 | 561020 | 561187 | + |
| LmjF.19_3099 | LmjF.19 | 561947 | 562501 | + |
| LmjF.19_3101 | LmjF.19 | 567210 | 567320 | + |
| LmjF.19_3102 | LmjF.19 | 567764 | 567964 | + |
| LmjF.19_3105 | LmjF.19 | 577105 | 577206 | + |
| LmjF.19_3109 | LmjF.19 | 589049 | 589171 | + |
| LmjF.19_3114 | LmjF.19 | 599803 | 600054 | + |
| LmjF.19_3117 | LmjF.19 | 607282 | 607374 | + |
| LmjF.19_3130 | LmjF.19 | 644230 | 644529 | + |
| LmjF.19_3135 | LmjF.19 | 650227 | 650418 | + |
| LmjF.19_3146 | LmjF.19 | 684352 | 684468 | + |
| LmjF.19_3150 | LmjF.19 | 696902 | 697354 | + |
| LmjF.19_3151 | LmjF.19 | 698510 | 698695 | + |
| LmjF.19_3152 | LmjF.19 | 698646 | 698771 | + |
| LmjF.20_3165 | LmjF.20 | 37994 | 38371 | - |
| LmjF.20_3186 | LmjF.20 | 111335 | 111922 | + |
| LmjF.20_3189 | LmjF.20 | 116568 | 116699 | + |
| LmjF.20_3193 | LmjF.20 | 127340 | 127699 | + |
| LmjF.20_3200 | LmjF.20 | 154486 | 155364 | + |
| LmjF.20_3206 | LmjF.20 | 175571 | 175684 | + |
| LmjF.20_3210 | LmjF.20 | 182013 | 182249 | + |
| LmjF.20_3228 | LmjF.20 | 240435 | 240653 | + |
| LmjF.20_3233 | LmjF.20 | 284034 | 284630 | + |
| LmjF.20_3235 | LmjF.20 | 300054 | 300371 | + |
| LmjF.20_3236 | LmjF.20 | 300545 | 300688 | + |
| LmjF.20_3256 | LmjF.20 | 389823 | 390137 | + |
| LmjF.20_3266 | LmjF.20 | 445786 | 445881 | + |
| LmjF.20_3267 | LmjF.20 | 445905 | 446066 | + |
| LmjF.20_3280 | LmjF.20 | 508514 | 508654 | + |
| LmjF.20_3293 | LmjF.20 | 558534 | 558629 | + |
| LmjF.20_3312 | LmjF.20 | 623030 | 623170 | + |
| LmjF.20_3314 | LmjF.20 | 629765 | 630016 | + |
| LmjF.20_3320 | LmjF.20 | 648946 | 649065 | + |
| LmjF.20_3326 | LmjF.20 | 661652 | 661864 | - |
| LmjF.20_3327 | LmjF.20 | 662064 | 662219 | - |
| LmjF.20_3346 | LmjF.20 | 718564 | 718707 | - |
| LmjF.20_3347 | LmjF.20 | 719427 | 719636 | - |
| LmjF.21_3431 | LmjF.21 | 222179 | 222307 | - |
| LmjF.21_3432 | LmjF.21 | 221650 | 222171 | - |
| LmjF.21_3434 | LmjF.21 | 223859 | 224041 | - |
| LmjF.21_3468 | LmjF.21 | 364852 | 364968 | + |
| LmjF.21_3484 | LmjF.21 | 418628 | 418804 | + |
| LmjF.21_3520 | LmjF.21 | 553382 | 553534 | - |
| LmjF.21_3539 | LmjF.21 | 622129 | 622299 | - |
| LmjF.21_3544 | LmjF.21 | 628508 | 628777 | - |
| LmjF.21_3547 | LmjF.21 | 637441 | 637713 | - |
| LmjF.21_3555 | LmjF.21 | 672358 | 672456 | - |
| LmjF.21_3560 | LmjF.21 | 683782 | 683997 | - |
| LmjF.21_3564 | LmjF.21 | 698741 | 699049 | - |
| LmjF.21_3580 | LmjF.21 | 731092 | 731466 | - |
| LmjF.22_3599 | LmjF.22 | 10032 | 10319 | + |
| LmjF.22_3602 | LmjF.22 | 13174 | 13467 | - |
| LmjF.22_3607 | LmjF.22 | 23382 | 23861 | - |
| LmjF.22_3612 | LmjF.22 | 49173 | 49421 | - |
| LmjF.22_3618 | LmjF.22 | 68399 | 68665 | - |
| LmjF.22_3646 | LmjF.22 | 179715 | 179810 | - |
| LmjF.22_3659 | LmjF.22 | 218830 | 219273 | - |
| LmjF.22_3665 | LmjF.22 | 229566 | 229856 | - |
| LmjF.22_3674 | LmjF.22 | 264753 | 265016 | - |
| LmjF.22_3675 | LmjF.22 | 268157 | 268258 | - |
| LmjF.22_3688 | LmjF.22 | 328838 | 328936 | + |
| LmjF.22_3692 | LmjF.22 | 341723 | 341878 | + |
| LmjF.22_3693 | LmjF.22 | 345016 | 345279 | + |
| LmjF.22_3699 | LmjF.22 | 362149 | 362286 | + |
| LmjF.22_3703 | LmjF.22 | 376335 | 376547 | + |
| LmjF.22_3706 | LmjF.22 | 384782 | 385144 | + |
| LmjF.22_3710 | LmjF.22 | 405487 | 405690 | + |
| LmjF.22_3744 | LmjF.22 | 538893 | 539480 | - |
| LmjF.22_3750 | LmjF.22 | 571290 | 571397 | - |
| LmjF.22_3770 | LmjF.22 | 628368 | 628460 | + |
| LmjF.22_3773 | LmjF.22 | 634724 | 635710 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.22_3780 | LmjF.22 | 672680 | 672796 | + |
| LmjF.22_3785 | LmjF.22 | 682270 | 682677 | + |
| LmjF.22_3789 | LmjF.22 | 695364 | 701963 | + |
| LmjF.23_3803 | LmjF.23 | 23243 | 23347 | + |
| LmjF.23_3808 | LmjF.23 | 43506 | 43664 | + |
| LmjF.23_3819 | LmjF.23 | 79049 | 79762 | + |
| LmjF.23_3822 | LmjF.23 | 83916 | 84179 | + |
| LmjF.23_3839 | LmjF.23 | 128812 | 129513 | + |
| LmjF.23_3845 | LmjF.23 | 151418 | 151558 | + |
| LmjF.23_3853 | LmjF.23 | 182985 | 183653 | + |
| LmjF.23_3873 | LmjF.23 | 245005 | 245238 | - |
| LmjF.23_3874 | LmjF.23 | 245299 | 245421 | - |
| LmjF.23_3883 | LmjF.23 | 277976 | 278086 | - |
| LmjF.23_3886 | LmjF.23 | 293212 | 294165 | - |
| LmjF.23_3901 | LmjF.23 | 379220 | 379594 | - |
| LmjF.23_3903 | LmjF.23 | 386824 | 387168 | - |
| LmjF.23_3911 | LmjF.23 | 411585 | 411758 | - |
| LmjF.23_3914 | LmjF.23 | 426935 | 427075 | - |
| LmjF.23_3919 | LmjF.23 | 439816 | 439923 | - |
| LmjF.23_3921 | LmjF.23 | 454869 | 455375 | - |
| LmjF.23_3927 | LmjF.23 | 467353 | 467508 | - |
| LmjF.23_3931 | LmjF.23 | 477090 | 477218 | - |
| LmjF.23_3934 | LmjF.23 | 488234 | 488377 | - |
| LmjF.23_3942 | LmjF.23 | 501014 | 501418 | - |
| LmjF.23_3955 | LmjF.23 | 545674 | 545787 | - |
| LmjF.23_3968 | LmjF.23 | 578165 | 578416 | + |
| LmjF.23_3976 | LmjF.23 | 602798 | 602926 | + |
| LmjF.23_3979 | LmjF.23 | 616169 | 616366 | + |
| LmjF.23_3981 | LmjF.23 | 619735 | 619998 | + |
| LmjF.23_3982 | LmjF.23 | 620215 | 620478 | + |
| LmjF.23_3989 | LmjF.23 | 658003 | 658212 | + |
| LmjF.23_4010 | LmjF.23 | 721765 | 721893 | + |
| LmjF.23_4012 | LmjF.23 | 726943 | 727392 | + |
| LmjF.23_4023 | LmjF.23 | 760711 | 760965 | + |
| LmjF.24_4094 | LmjF.24 | 250509 | 250787 | + |
| LmjF.24_4099 | LmjF.24 | 264282 | 264479 | + |
| LmjF.24_4136 | LmjF.24 | 390235 | 390909 | + |
| LmjF.24_4137 | LmjF.24 | 392389 | 393135 | + |
| LmjF.24_4144 | LmjF.24 | 418293 | 419021 | + |
| LmjF.24_4157 | LmjF.24 | 457136 | 457267 | + |
| LmjF.24_4167 | LmjF.24 | 485496 | 485963 | + |
| LmjF.24_4169 | LmjF.24 | 489797 | 489919 | + |
| LmjF.24_4198 | LmjF.24 | 613487 | 613783 | + |
| LmjF.24_4212 | LmjF.24 | 657382 | 657546 | - |
| LmjF.24_4215 | LmjF.24 | 666770 | 666940 | - |
| LmjF.24_4261 | LmjF.24 | 805848 | 805952 | - |
| LmjF.24_4262 | LmjF.24 | 806683 | 806787 | - |
| LmjF.24_4263 | LmjF.24 | 807597 | 807701 | - |
| LmjF.25_4279 | LmjF.25 | 11694 | 11933 | - |
| LmjF.25_4285 | LmjF.25 | 22606 | 22962 | - |
| LmjF.25_4288 | LmjF.25 | 29089 | 29349 | - |
| LmjF.25_4291 | LmjF.25 | 31550 | 31720 | - |
| LmjF.25_4294 | LmjF.25 | 37558 | 37713 | - |
| LmjF.25_4323 | LmjF.25 | 123710 | 123955 | - |
| LmjF.25_4338 | LmjF.25 | 197208 | 197540 | - |
| LmjF.25_4345 | LmjF.25 | 210385 | 210690 | - |
| LmjF.25_4357 | LmjF.25 | 260248 | 260394 | - |
| LmjF.25_4359 | LmjF.25 | 265010 | 265474 | - |
| LmjF.25_4361 | LmjF.25 | 267284 | 267835 | - |
| LmjF.25_4377 | LmjF.25 | 332131 | 332976 | + |
| LmjF.25_4378 | LmjF.25 | 334316 | 334741 | + |
| LmjF.25_4396 | LmjF.25 | 397670 | 397981 | + |
| LmjF.25_4426 | LmjF.25 | 527057 | 527290 | - |
| LmjF.25_4431 | LmjF.25 | 550595 | 550750 | - |
| LmjF.25_4440 | LmjF.25 | 576360 | 576482 | - |
| LmjF.25_4441 | LmjF.25 | 576781 | 577041 | - |
| LmjF.25_4461 | LmjF.25 | 659305 | 659397 | - |
| LmjF.25_4463 | LmjF.25 | 663816 | 664130 | - |
| LmjF.25_4479 | LmjF.25 | 707273 | 707398 | - |
| LmjF.25_4491 | LmjF.25 | 745007 | 745204 | - |
| LmjF.25_4545 | LmjF.25 | 880921 | 881397 | + |
| LmjF.25_4550 | LmjF.25 | 898007 | 898162 | + |
| LmjF.25_4554 | LmjF.25 | 903271 | 903393 | + |
| LmjF.25_4556 | LmjF.25 | 908439 | 908726 | + |
| LmjF.25_4557 | LmjF.25 | 908620 | 908853 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.26_4564 | LmjF.26 | 15464 | 15982 | - |
| LmjF.26_4570 | LmjF.26 | 23329 | 23508 | - |
| LmjF.26_4585 | LmjF.26 | 58735 | 58977 | - |
| LmjF.26_4609 | LmjF.26 | 122940 | 123269 | - |
| LmjF.26_4636 | LmjF.26 | 211169 | 211762 | - |
| LmjF.26_4641 | LmjF.26 | 229533 | 229745 | - |
| LmjF.26_4662 | LmjF.26 | 297493 | 297588 | - |
| LmjF.26_4707 | LmjF.26 | 492712 | 492975 | + |
| LmjF.26_4711 | LmjF.26 | 522463 | 522699 | + |
| LmjF.26_4716 | LmjF.26 | 537949 | 538416 | + |
| LmjF.26_4724 | LmjF.26 | 573222 | 573857 | + |
| LmjF.26_4738 | LmjF.26 | 613651 | 613788 | + |
| LmjF.26_4740 | LmjF.26 | 616024 | 616269 | + |
| LmjF.26_4744 | LmjF.26 | 621941 | 622357 | + |
| LmjF.26_4745 | LmjF.26 | 622907 | 623011 | + |
| LmjF.26_4763 | LmjF.26 | 737645 | 737794 | + |
| LmjF.26_4767 | LmjF.26 | 747329 | 747568 | + |
| LmjF.26_4772 | LmjF.26 | 759582 | 759680 | + |
| LmjF.26_4799 | LmjF.26 | 894825 | 895196 | + |
| LmjF.26_4838 | LmjF.26 | 1028616 | 1028711 | + |
| LmjF.26_4840 | LmjF.26 | 1032257 | 1032469 | + |
| LmjF.26_4845 | LmjF.26 | 1047130 | 1047249 | + |
| LmjF.26_4846 | LmjF.26 | 1047250 | 1047738 | + |
| LmjF.26_4852 | LmjF.26 | 1061703 | 1061921 | + |
| LmjF.27_4860 | LmjF.27 | 2769 | 3149 | - |
| LmjF.27_4876 | LmjF.27 | 39042 | 39182 | - |
| LmjF.27_4886 | LmjF.27 | 56209 | 56532 | - |
| LmjF.27_4915 | LmjF.27 | 218674 | 218811 | + |
| LmjF.27_4938 | LmjF.27 | 297197 | 297304 | + |
| LmjF.27_4947 | LmjF.27 | 321960 | 322058 | + |
| LmjF.27_4948 | LmjF.27 | 324096 | 324389 | + |
| LmjF.27_4960 | LmjF.27 | 369406 | 369519 | + |
| LmjF.27_4965 | LmjF.27 | 393742 | 393858 | - |
| LmjF.27_4968 | LmjF.27 | 404639 | 405034 | - |
| LmjF.27_4994 | LmjF.27 | 499834 | 500124 | - |
| LmjF.27_4995 | LmjF.27 | 501001 | 501150 | - |
| LmjF.27_5002 | LmjF.27 | 522967 | 523143 | - |
| LmjF.27_5018 | LmjF.27 | 571090 | 571251 | + |
| LmjF.27_5026 | LmjF.27 | 592628 | 592861 | + |
| LmjF.27_5042 | LmjF.27 | 652580 | 652738 | + |
| LmjF.27_5044 | LmjF.27 | 655113 | 655301 | + |
| LmjF.27_5045 | LmjF.27 | 655443 | 655592 | + |
| LmjF.27_5047 | LmjF.27 | 657578 | 657772 | + |
| LmjF.27_5058 | LmjF.27 | 717537 | 717752 | + |
| LmjF.27_5081 | LmjF.27 | 836255 | 836371 | - |
| LmjF.27_5088 | LmjF.27 | 860992 | 861156 | - |
| LmjF.27_5114 | LmjF.27 | 946873 | 947031 | - |
| LmjF.27_5115 | LmjF.27 | 948319 | 948591 | - |
| LmjF.27_5147 | LmjF.27 | 1108703 | 1108978 | + |
| LmjF.28_5161 | LmjF.28 | 18568 | 18687 | + |
| LmjF.28_5165 | LmjF.28 | 33383 | 33502 | + |
| LmjF.28_5171 | LmjF.28 | 46533 | 46844 | + |
| LmjF.28_5174 | LmjF.28 | 53169 | 53297 | + |
| LmjF.28_5175 | LmjF.28 | 53468 | 53602 | + |
| LmjF.28_5177 | LmjF.28 | 59133 | 59453 | + |
| LmjF.28_5210 | LmjF.28 | 162314 | 162445 | - |
| LmjF.28_5213 | LmjF.28 | 177018 | 177170 | - |
| LmjF.28_5219 | LmjF.28 | 194878 | 195252 | - |
| LmjF.28_5220 | LmjF.28 | 195932 | 196156 | - |
| LmjF.28_5232 | LmjF.28 | 247385 | 247531 | - |
| LmjF.28_5259 | LmjF.28 | 325860 | 325955 | + |
| LmjF.28_5269 | LmjF.28 | 356788 | 357234 | + |
| LmjF.28_5271 | LmjF.28 | 362430 | 362870 | + |
| LmjF.28_5307 | LmjF.28 | 497691 | 498077 | + |
| LmjF.28_5313 | LmjF.28 | 512498 | 512701 | + |
| LmjF.28_5320 | LmjF.28 | 540861 | 541091 | + |
| LmjF.28_5321 | LmjF.28 | 541099 | 541215 | + |
| LmjF.28_5322 | LmjF.28 | 543194 | 543304 | + |
| LmjF.28_5323 | LmjF.28 | 543309 | 543431 | + |
| LmjF.28_5332 | LmjF.28 | 565396 | 565500 | + |
| LmjF.28_5340 | LmjF.28 | 593705 | 594034 | + |
| LmjF.28_5344 | LmjF.28 | 603920 | 604126 | - |
| LmjF.28_5383 | LmjF.28 | 783213 | 783755 | - |
| LmjF.28_5384 | LmjF.28 | 785857 | 786171 | - |
| LmjF.28_5386 | LmjF.28 | 796462 | 796617 | - |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.28_5409 | LmjF.28 | 855796 | 856122 | + |
| LmjF.28_5414 | LmjF.28 | 873612 | 873725 | + |
| LmjF.28_5415 | LmjF.28 | 873022 | 873261 | + |
| LmjF.28_5416 | LmjF.28 | 873880 | 874194 | + |
| LmjF.28_5419 | LmjF.28 | 879816 | 880046 | + |
| LmjF.28_5439 | LmjF.28 | 952514 | 952618 | + |
| LmjF.28_5472 | LmjF.28 | 1054881 | 1054997 | - |
| LmjF.28_5487 | LmjF.28 | 1110389 | 1110598 | - |
| LmjF.28_5490 | LmjF.28 | 1119601 | 1119867 | - |
| LmjF.28_5500 | LmjF.28 | 1154797 | 1155570 | - |
| LmjF.29_5532 | LmjF.29 | 102534 | 103037 | - |
| LmjF.29_5533 | LmjF.29 | 105309 | 105680 | - |
| LmjF.29_5548 | LmjF.29 | 153682 | 153918 | - |
| LmjF.29_5578 | LmjF.29 | 270939 | 271628 | - |
| LmjF.29_5587 | LmjF.29 | 303429 | 303548 | - |
| LmjF.29_5596 | LmjF.29 | 330789 | 331310 | - |
| LmjF.29_5613 | LmjF.29 | 392478 | 392603 | + |
| LmjF.29_5632 | LmjF.29 | 462844 | 462987 | + |
| LmjF.29_5641 | LmjF.29 | 501533 | 501685 | + |
| LmjF.29_5644 | LmjF.29 | 513955 | 514065 | + |
| LmjF.29_5649 | LmjF.29 | 531237 | 531524 | + |
| LmjF.29_5658 | LmjF.29 | 631449 | 631628 | + |
| LmjF.29_5667 | LmjF.29 | 676994 | 677137 | - |
| LmjF.29_5684 | LmjF.29 | 749984 | 750280 | - |
| LmjF.29_5717 | LmjF.29 | 859640 | 860014 | - |
| LmjF.29_5734 | LmjF.29 | 929491 | 929745 | - |
| LmjF.29_5745 | LmjF.29 | 982742 | 982837 | - |
| LmjF.29_5749 | LmjF.29 | 992244 | 992342 | - |
| LmjF.29_5777 | LmjF.29 | 1075281 | 1075373 | + |
| LmjF.29_5779 | LmjF.29 | 1079507 | 1079710 | + |
| LmjF.29_5783 | LmjF.29 | 1092543 | 1092830 | + |
| LmjF.29_5796 | LmjF.29 | 1128590 | 1128682 | + |
| LmjF.29_5797 | LmjF.29 | 1130176 | 1130364 | + |
| LmjF.29_5809 | LmjF.29 | 1169377 | 1169532 | + |
| LmjF.29_5820 | LmjF.29 | 1193346 | 1193567 | + |
| LmjF.30_5826 | LmjF.30 | 3511 | 3921 | - |
| LmjF.30_5836 | LmjF.30 | 27478 | 27846 | - |
| LmjF.30_5839 | LmjF.30 | 37984 | 38076 | - |
| LmjF.30_5846 | LmjF.30 | 52157 | 52399 | - |
| LmjF.30_5851 | LmjF.30 | 62613 | 63149 | - |
| LmjF.30_5884 | LmjF.30 | 165514 | 165639 | - |
| LmjF.30_5896 | LmjF.30 | 205310 | 205552 | - |
| LmjF.30_5900 | LmjF.30 | 213272 | 213886 | - |
| LmjF.30_5909 | LmjF.30 | 225965 | 226486 | - |
| LmjF.30_5932 | LmjF.30 | 286667 | 286825 | + |
| LmjF.30_5956 | LmjF.30 | 377085 | 377540 | + |
| LmjF.30_5964 | LmjF.30 | 403326 | 403592 | + |
| LmjF.30_5978 | LmjF.30 | 446104 | 446232 | + |
| LmjF.30_5989 | LmjF.30 | 488721 | 488906 | + |
| LmjF.30_5993 | LmjF.30 | 509584 | 509811 | + |
| LmjF.30_5994 | LmjF.30 | 516765 | 517199 | + |
| LmjF.30_5997 | LmjF.30 | 525189 | 525329 | + |
| LmjF.30_6006 | LmjF.30 | 545633 | 545809 | + |
| LmjF.30_6007 | LmjF.30 | 547754 | 547888 | + |
| LmjF.30_6034 | LmjF.30 | 617231 | 617452 | + |
| LmjF.30_6043 | LmjF.30 | 641964 | 642134 | + |
| LmjF.30_6060 | LmjF.30 | 722220 | 722372 | + |
| LmjF.30_6063 | LmjF.30 | 730009 | 730107 | + |
| LmjF.30_6069 | LmjF.30 | 749100 | 749285 | + |
| LmjF.30_6078 | LmjF.30 | 782363 | 783034 | + |
| LmjF.30_6086 | LmjF.30 | 805843 | 806043 | - |
| LmjF.30_6091 | LmjF.30 | 834520 | 834630 | - |
| LmjF.30_6092 | LmjF.30 | 836452 | 836859 | - |
| LmjF.30_6109 | LmjF.30 | 903633 | 903944 | - |
| LmjF.30_6115 | LmjF.30 | 928397 | 928498 | - |
| LmjF.30_6135 | LmjF.30 | 1009117 | 1009239 | - |
| LmjF.30_6141 | LmjF.30 | 1024307 | 1024717 | - |
| LmjF.30_6147 | LmjF.30 | 1033976 | 1034077 | - |
| LmjF.30_6165 | LmjF.30 | 1079227 | 1079370 | - |
| LmjF.30_6176 | LmjF.30 | 1107826 | 1108050 | - |
| LmjF.30_6205 | LmjF.30 | 1214568 | 1214672 | - |
| LmjF.30_6227 | LmjF.30 | 1288884 | 1289387 | + |
| LmjF.30_6234 | LmjF.30 | 1314082 | 1314360 | + |
| LmjF.30_6238 | LmjF.30 | 1320895 | 1321521 | + |
| LmjF.30_6253 | LmjF.30 | 1358433 | 1358588 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.30_6262 | LmjF.30 | 1382044 | 1382136 | + |
| LmjF.30_6264 | LmjF.30 | 1389754 | 1389999 | + |
| LmjF.31_6279 | LmjF.31 | 28489 | 28740 | - |
| LmjF.31_6291 | LmjF.31 | 53651 | 53776 | - |
| LmjF.31_6317 | LmjF.31 | 149183 | 149401 | - |
| LmjF.31_6346 | LmjF.31 | 249183 | 249359 | - |
| LmjF.31_6347 | LmjF.31 | 250484 | 250606 | - |
| LmjF.31_6353 | LmjF.31 | 271888 | 272007 | - |
| LmjF.31_6355 | LmjF.31 | 278862 | 279320 | - |
| LmjF.31_6365 | LmjF.31 | 311657 | 311815 | - |
| LmjF.31_6366 | LmjF.31 | 314762 | 315199 | - |
| LmjF.31_6372 | LmjF.31 | 334569 | 334820 | - |
| LmjF.31_6374 | LmjF.31 | 345010 | 345921 | - |
| LmjF.31_6377 | LmjF.31 | 361432 | 362343 | - |
| LmjF.31_6379 | LmjF.31 | 371246 | 371485 | - |
| LmjF.31_6381 | LmjF.31 | 374768 | 374866 | - |
| LmjF.31_6388 | LmjF.31 | 404914 | 405030 | - |
| LmjF.31_6400 | LmjF.31 | 451625 | 451759 | - |
| LmjF.31_6402 | LmjF.31 | 458381 | 458515 | - |
| LmjF.31_6403 | LmjF.31 | 458988 | 459194 | - |
| LmjF.31_6404 | LmjF.31 | 459635 | 459901 | - |
| LmjF.31_6417 | LmjF.31 | 516172 | 516339 | - |
| LmjF.31_6425 | LmjF.31 | 584579 | 585256 | - |
| LmjF.31_6445 | LmjF.31 | 714872 | 714988 | - |
| LmjF.31_6452 | LmjF.31 | 763758 | 764342 | - |
| LmjF.31_6460 | LmjF.31 | 774843 | 775013 | - |
| LmjF.31_6472 | LmjF.31 | 815080 | 815307 | - |
| LmjF.31_6476 | LmjF.31 | 830485 | 830766 | - |
| LmjF.31_6477 | LmjF.31 | 835978 | 836469 | - |
| LmjF.31_6485 | LmjF.31 | 874807 | 874983 | - |
| LmjF.31_6489 | LmjF.31 | 888636 | 888839 | - |
| LmjF.31_6490 | LmjF.31 | 892030 | 892509 | - |
| LmjF.31_6491 | LmjF.31 | 893894 | 895051 | - |
| LmjF.31_6493 | LmjF.31 | 904472 | 904693 | - |
| LmjF.31_6494 | LmjF.31 | 907110 | 907610 | - |
| LmjF.31_6497 | LmjF.31 | 923555 | 923743 | - |
| LmjF.31_6502 | LmjF.31 | 938638 | 938892 | - |
| LmjF.31_6504 | LmjF.31 | 948638 | 948916 | - |
| LmjF.31_6507 | LmjF.31 | 958736 | 959134 | - |
| LmjF.31_6510 | LmjF.31 | 974171 | 974362 | - |
| LmjF.31_6528 | LmjF.31 | 1051831 | 1051950 | - |
| LmjF.31_6534 | LmjF.31 | 1094511 | 1094708 | - |
| LmjF.31_6539 | LmjF.31 | 1118926 | 1119126 | - |
| LmjF.31_6551 | LmjF.31 | 1157449 | 1157631 | - |
| LmjF.31_6552 | LmjF.31 | 1163514 | 1163666 | - |
| LmjF.31_6583 | LmjF.31 | 1273984 | 1274109 | - |
| LmjF.31_6599 | LmjF.31 | 1335460 | 1335960 | - |
| LmjF.31_6600 | LmjF.31 | 1337797 | 1337988 | - |
| LmjF.31_6616 | LmjF.31 | 1379543 | 1379767 | - |
| LmjF.31_6625 | LmjF.31 | 1417384 | 1417485 | - |
| LmjF.31_6632 | LmjF.31 | 1433266 | 1433730 | - |
| LmjF.31_6637 | LmjF.31 | 1443464 | 1443610 | - |
| LmjF.32_6654 | LmjF.32 | 12916 | 13044 | - |
| LmjF.32_6655 | LmjF.32 | 13154 | 13267 | - |
| LmjF.32_6657 | LmjF.32 | 15906 | 16313 | - |
| LmjF.32_6663 | LmjF.32 | 26330 | 26521 | - |
| LmjF.32_6673 | LmjF.32 | 54755 | 54928 | - |
| LmjF.32_6681 | LmjF.32 | 72864 | 72986 | - |
| LmjF.32_6683 | LmjF.32 | 73791 | 74030 | - |
| LmjF.32_6719 | LmjF.32 | 197223 | 197348 | + |
| LmjF.32_6733 | LmjF.32 | 228486 | 228734 | + |
| LmjF.32_6748 | LmjF.32 | 278523 | 278945 | + |
| LmjF.32_6750 | LmjF.32 | 288007 | 288222 | + |
| LmjF.32_6753 | LmjF.32 | 303230 | 303478 | + |
| LmjF.32_6758 | LmjF.32 | 324524 | 324718 | + |
| LmjF.32_6766 | LmjF.32 | 344566 | 344940 | + |
| LmjF.32_6775 | LmjF.32 | 393728 | 393958 | + |
| LmjF.32_6803 | LmjF.32 | 483095 | 483334 | + |
| LmjF.32_6804 | LmjF.32 | 483545 | 483709 | + |
| LmjF.32_6807 | LmjF.32 | 490734 | 490958 | + |
| LmjF.32_6808 | LmjF.32 | 492718 | 493101 | + |
| LmjF.32_6811 | LmjF.32 | 501111 | 501248 | + |
| LmjF.32_6832 | LmjF.32 | 577072 | 577215 | - |
| LmjF.32_6834 | LmjF.32 | 584438 | 584647 | - |
| LmjF.32_6843 | LmjF.32 | 609543 | 609641 | - |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.32_6865 | LmjF.32 | 680526 | 680663 | - |
| LmjF.32_6874 | LmjF.32 | 709162 | 709281 | - |
| LmjF.32_6888 | LmjF.32 | 752872 | 753144 | - |
| LmjF.32_6899 | LmjF.32 | 787074 | 787208 | - |
| LmjF.32_6906 | LmjF.32 | 804421 | 804522 | - |
| LmjF.32_6916 | LmjF.32 | 841775 | 842152 | - |
| LmjF.32_6923 | LmjF.32 | 861193 | 861348 | - |
| LmjF.32_6945 | LmjF.32 | 949586 | 949684 | - |
| LmjF.32_6951 | LmjF.32 | 960860 | 961027 | - |
| LmjF.32_6952 | LmjF.32 | 966516 | 966740 | - |
| LmjF.32_6956 | LmjF.32 | 979339 | 980139 | - |
| LmjF.32_6972 | LmjF.32 | 1045104 | 1045859 | - |
| LmjF.32_6987 | LmjF.32 | 1097752 | 1098168 | - |
| LmjF.32_7004 | LmjF.32 | 1161019 | 1161147 | - |
| LmjF.32_7010 | LmjF.32 | 1173277 | 1174002 | + |
| LmjF.32_7012 | LmjF.32 | 1174241 | 1174351 | + |
| LmjF.32_7013 | LmjF.32 | 1177162 | 1177695 | + |
| LmjF.32_7014 | LmjF.32 | 1178115 | 1178330 | + |
| LmjF.32_7026 | LmjF.32 | 1210135 | 1210245 | + |
| LmjF.32_7039 | LmjF.32 | 1259010 | 1259255 | + |
| LmjF.32_7042 | LmjF.32 | 1268617 | 1268823 | + |
| LmjF.32_7050 | LmjF.32 | 1314654 | 1315043 | + |
| LmjF.32_7058 | LmjF.32 | 1338987 | 1339202 | + |
| LmjF.32_7068 | LmjF.32 | 1387272 | 1387541 | + |
| LmjF.32_7073 | LmjF.32 | 1402015 | 1402239 | + |
| LmjF.32_7106 | LmjF.32 | 1500702 | 1501307 | + |
| LmjF.32_7132 | LmjF.32 | 1565617 | 1565901 | + |
| LmjF.32_7133 | LmjF.32 | 1566491 | 1566679 | + |
| LmjF.32_7140 | LmjF.32 | 1596351 | 1596704 | + |
| LmjF.32_7142 | LmjF.32 | 1600951 | 1601208 | + |
| LmjF.33_7170 | LmjF.33 | 70742 | 70879 | - |
| LmjF.33_7174 | LmjF.33 | 78560 | 78763 | - |
| LmjF.33_7201 | LmjF.33 | 181566 | 181943 | - |
| LmjF.33_7204 | LmjF.33 | 185730 | 185990 | - |
| LmjF.33_7206 | LmjF.33 | 190020 | 190157 | - |
| LmjF.33_7207 | LmjF.33 | 192236 | 192943 | - |
| LmjF.33_7213 | LmjF.33 | 209032 | 209127 | - |
| LmjF.33_7221 | LmjF.33 | 232640 | 232747 | - |
| LmjF.33_7233 | LmjF.33 | 264068 | 264253 | - |
| LmjF.33_7234 | LmjF.33 | 264012 | 264263 | - |
| LmjF.33_7235 | LmjF.33 | 264486 | 264653 | - |
| LmjF.33_7240 | LmjF.33 | 272910 | 273224 | + |
| LmjF.33_7242 | LmjF.33 | 272957 | 273493 | + |
| LmjF.33_7259 | LmjF.33 | 323981 | 324364 | + |
| LmjF.33_7261 | LmjF.33 | 333320 | 333424 | + |
| LmjF.33_7294 | LmjF.33 | 434166 | 434312 | + |
| LmjF.33_7296 | LmjF.33 | 442631 | 442849 | + |
| LmjF.33_7301 | LmjF.33 | 459297 | 459506 | + |
| LmjF.33_7304 | LmjF.33 | 481625 | 481759 | + |
| LmjF.33_7305 | LmjF.33 | 481785 | 481958 | + |
| LmjF.33_7319 | LmjF.33 | 525888 | 526103 | + |
| LmjF.33_7335 | LmjF.33 | 583100 | 583201 | + |
| LmjF.33_7336 | LmjF.33 | 588524 | 588910 | + |
| LmjF.33_7343 | LmjF.33 | 612828 | 613430 | + |
| LmjF.33_7348 | LmjF.33 | 627958 | 628332 | + |
| LmjF.33_7353 | LmjF.33 | 663733 | 663954 | + |
| LmjF.33_7365 | LmjF.33 | 742969 | 743112 | + |
| LmjF.33_7370 | LmjF.33 | 754185 | 754439 | + |
| LmjF.33_7380 | LmjF.33 | 782259 | 782600 | + |
| LmjF.33_7383 | LmjF.33 | 793708 | 793827 | + |
| LmjF.33_7396 | LmjF.33 | 832722 | 833105 | + |
| LmjF.33_7407 | LmjF.33 | 859658 | 860182 | + |
| LmjF.33_7416 | LmjF.33 | 892722 | 892973 | + |
| LmjF.33_7420 | LmjF.33 | 911331 | 911444 | + |
| LmjF.33_7421 | LmjF.33 | 916223 | 916531 | + |
| LmjF.33_7429 | LmjF.33 | 945613 | 945729 | + |
| LmjF.33_7432 | LmjF.33 | 954344 | 954562 | + |
| LmjF.33_7440 | LmjF.33 | 992287 | 992505 | + |
| LmjF.33_7449 | LmjF.33 | 1023674 | 1023904 | + |
| LmjF.33_7450 | LmjF.33 | 1023380 | 1023508 | + |
| LmjF.33_7454 | LmjF.33 | 1032371 | 1032637 | + |
| LmjF.33_7459 | LmjF.33 | 1057320 | 1057580 | + |
| LmjF.33_7464 | LmjF.33 | 1073630 | 1074520 | + |
| LmjF.33_7470 | LmjF.33 | 1094581 | 1094901 | + |
| LmjF.33_7501 | LmjF.33 | 1219899 | 1221326 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.33_7506 | LmjF.33 | 1243800 | 1244342 | + |
| LmjF.33_7515 | LmjF.33 | 1289644 | 1290105 | + |
| LmjF.33_7521 | LmjF.33 | 1300512 | 1300856 | + |
| LmjF.33_7522 | LmjF.33 | 1302311 | 1302985 | + |
| LmjF.33_7523 | LmjF.33 | 1305980 | 1306234 | + |
| LmjF.33_7524 | LmjF.33 | 1310341 | 1310955 | + |
| LmjF.33_7525 | LmjF.33 | 1314985 | 1315320 | + |
| LmjF.33_7526 | LmjF.33 | 1316916 | 1317284 | + |
| LmjF.33_7527 | LmjF.33 | 1319105 | 1319386 | + |
| LmjF.33_7528 | LmjF.33 | 1320502 | 1321179 | + |
| LmjF.33_7529 | LmjF.33 | 1323981 | 1324133 | + |
| LmjF.33_7530 | LmjF.33 | 1324130 | 1324384 | + |
| LmjF.33_7531 | LmjF.33 | 1328456 | 1329109 | + |
| LmjF.33_7532 | LmjF.33 | 1331221 | 1331472 | + |
| LmjF.33_7533 | LmjF.33 | 1333103 | 1333468 | + |
| LmjF.33_7536 | LmjF.33 | 1367027 | 1367122 | + |
| LmjF.33_7543 | LmjF.33 | 1419498 | 1419839 | + |
| LmjF.33_7547 | LmjF.33 | 1441882 | 1441983 | + |
| LmjF.33_7555 | LmjF.33 | 1487380 | 1487475 | + |
| LmjF.33_7559 | LmjF.33 | 1525876 | 1526025 | + |
| LmjF.33_7560 | LmjF.33 | 1526031 | 1526255 | + |
| LmjF.33_7563 | LmjF.33 | 1534126 | 1534269 | + |
| LmjF.33_7571 | LmjF.33 | 1551991 | 1552173 | + |
| LmjF.33_7574 | LmjF.33 | 1559966 | 1560109 | + |
| LmjF.33_7575 | LmjF.33 | 1560224 | 1561138 | + |
| LmjF.33_7576 | LmjF.33 | 1561773 | 1561901 | + |
| LmjF.33_7584 | LmjF.33 | 1577475 | 1577702 | + |
| LmjF.34_7585 | LmjF.34 | 3520 | 3627 | - |
| LmjF.34_7591 | LmjF.34 | 17041 | 17751 | + |
| LmjF.34_7614 | LmjF.34 | 91749 | 91958 | + |
| LmjF.34_7616 | LmjF.34 | 96814 | 97035 | + |
| LmjF.34_7633 | LmjF.34 | 152654 | 152821 | + |
| LmjF.34_7639 | LmjF.34 | 191561 | 191719 | + |
| LmjF.34_7641 | LmjF.34 | 196668 | 197066 | + |
| LmjF.34_7648 | LmjF.34 | 235203 | 235400 | + |
| LmjF.34_7655 | LmjF.34 | 258547 | 258849 | + |
| LmjF.34_7656 | LmjF.34 | 261107 | 261568 | + |
| LmjF.34_7663 | LmjF.34 | 303090 | 303368 | - |
| LmjF.34_7689 | LmjF.34 | 392611 | 392811 | + |
| LmjF.34_7694 | LmjF.34 | 408337 | 408480 | + |
| LmjF.34_7705 | LmjF.34 | 456122 | 456214 | + |
| LmjF.34_7711 | LmjF.34 | 496793 | 497002 | - |
| LmjF.34_7720 | LmjF.34 | 524418 | 524957 | - |
| LmjF.34_7728 | LmjF.34 | 551098 | 551328 | - |
| LmjF.34_7729 | LmjF.34 | 551277 | 551570 | - |
| LmjF.34_7730 | LmjF.34 | 551611 | 552171 | - |
| LmjF.34_7753 | LmjF.34 | 652640 | 652774 | - |
| LmjF.34_7760 | LmjF.34 | 678223 | 678324 | - |
| LmjF.34_7761 | LmjF.34 | 678090 | 678245 | - |
| LmjF.34_7765 | LmjF.34 | 691013 | 691144 | - |
| LmjF.34_7767 | LmjF.34 | 692857 | 693042 | - |
| LmjF.34_7769 | LmjF.34 | 696893 | 697150 | - |
| LmjF.34_7772 | LmjF.34 | 701626 | 702021 | - |
| LmjF.34_7813 | LmjF.34 | 874419 | 875024 | - |
| LmjF.34_7815 | LmjF.34 | 877547 | 877708 | - |
| LmjF.34_7824 | LmjF.34 | 915966 | 916916 | - |
| LmjF.34_7839 | LmjF.34 | 980128 | 980298 | - |
| LmjF.34_7866 | LmjF.34 | 1086900 | 1087097 | - |
| LmjF.34_7880 | LmjF.34 | 1163684 | 1163992 | + |
| LmjF.34_7894 | LmjF.34 | 1229804 | 1230139 | + |
| LmjF.34_7898 | LmjF.34 | 1237690 | 1237929 | + |
| LmjF.34_7905 | LmjF.34 | 1269808 | 1270020 | + |
| LmjF.34_7982 | LmjF.34 | 1529138 | 1529254 | + |
| LmjF.34_8004 | LmjF.34 | 1594167 | 1594367 | + |
| LmjF.34_8019 | LmjF.34 | 1630662 | 1630766 | + |
| LmjF.34_8034 | LmjF.34 | 1671199 | 1671570 | + |
| LmjF.34_8035 | LmjF.34 | 1671199 | 1671570 | + |
| LmjF.34_8039 | LmjF.34 | 1689603 | 1689815 | + |
| LmjF.34_8040 | LmjF.34 | 1691974 | 1692117 | + |
| LmjF.34_8067 | LmjF.34 | 1780258 | 1780500 | + |
| LmjF.35_8115 | LmjF.35 | 7688 | 7810 | - |
| LmjF.35_8121 | LmjF.35 | 19701 | 20405 | - |
| LmjF.35_8122 | LmjF.35 | 20598 | 20849 | - |
| LmjF.35_8145 | LmjF.35 | 72768 | 73385 | + |
| LmjF.35_8153 | LmjF.35 | 92057 | 92233 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.35_8157 | LmjF.35 | 102552 | 102995 | + |
| LmjF.35_8159 | LmjF.35 | 111077 | 112000 | + |
| LmjF.35_8168 | LmjF.35 | 139595 | 140566 | + |
| LmjF.35_8173 | LmjF.35 | 247758 | 248342 | + |
| LmjF.35_8181 | LmjF.35 | 287149 | 287658 | + |
| LmjF.35_8185 | LmjF.35 | 301592 | 301978 | + |
| LmjF.35_8190 | LmjF.35 | 331226 | 331471 | + |
| LmjF.35_8199 | LmjF.35 | 356546 | 356833 | + |
| LmjF.35_8206 | LmjF.35 | 372298 | 372579 | + |
| LmjF.35_8209 | LmjF.35 | 383704 | 383991 | + |
| LmjF.35_8212 | LmjF.35 | 392494 | 392664 | + |
| LmjF.35_8242 | LmjF.35 | 511163 | 511321 | + |
| LmjF.35_8249 | LmjF.35 | 542292 | 542462 | + |
| LmjF.35_8255 | LmjF.35 | 557578 | 558132 | + |
| LmjF.35_8262 | LmjF.35 | 571602 | 571766 | + |
| LmjF.35_8269 | LmjF.35 | 590186 | 590347 | + |
| LmjF.35_8283 | LmjF.35 | 652624 | 652800 | - |
| LmjF.35_8285 | LmjF.35 | 649781 | 650113 | - |
| LmjF.35_8287 | LmjF.35 | 659091 | 659258 | - |
| LmjF.35_8296 | LmjF.35 | 685830 | 685943 | - |
| LmjF.35_8303 | LmjF.35 | 711040 | 711234 | - |
| LmjF.35_8323 | LmjF.35 | 756958 | 757113 | + |
| LmjF.35_8344 | LmjF.35 | 833018 | 833284 | + |
| LmjF.35_8354 | LmjF.35 | 877847 | 877972 | + |
| LmjF.35_8421 | LmjF.35 | 1131910 | 1132059 | - |
| LmjF.35_8422 | LmjF.35 | 1131765 | 1131923 | - |
| LmjF.35_8424 | LmjF.35 | 1138950 | 1139450 | - |
| LmjF.35_8427 | LmjF.35 | 1152336 | 1152440 | - |
| LmjF.35_8429 | LmjF.35 | 1156802 | 1156984 | - |
| LmjF.35_8441 | LmjF.35 | 1196139 | 1196639 | - |
| LmjF.35_8458 | LmjF.35 | 1254516 | 1254848 | - |
| LmjF.35_8459 | LmjF.35 | 1256287 | 1256589 | - |
| LmjF.35_8468 | LmjF.35 | 1292593 | 1293399 | - |
| LmjF.35_8469 | LmjF.35 | 1293462 | 1293656 | - |
| LmjF.35_8482 | LmjF.35 | 1339793 | 1339984 | - |
| LmjF.35_8483 | LmjF.35 | 1339881 | 1339994 | - |
| LmjF.35_8528 | LmjF.35 | 1483673 | 1483768 | - |
| LmjF.35_8534 | LmjF.35 | 1507040 | 1507264 | - |
| LmjF.35_8582 | LmjF.35 | 1668024 | 1668197 | + |
| LmjF.35_8639 | LmjF.35 | 1883459 | 1883575 | + |
| LmjF.35_8652 | LmjF.35 | 1925329 | 1925985 | + |
| LmjF.35_8655 | LmjF.35 | 1936843 | 1937154 | + |
| LmjF.35_8677 | LmjF.35 | 2013532 | 2013726 | + |
| LmjF.35_8683 | LmjF.35 | 2026034 | 2026297 | + |
| LmjF.35_8685 | LmjF.35 | 2036303 | 2036485 | + |
| LmjF.35_8686 | LmjF.35 | 2036403 | 2036660 | + |
| LmjF.35_8687 | LmjF.35 | 2037086 | 2037343 | + |
| LmjF.35_8697 | LmjF.35 | 2052367 | 2052462 | + |
| LmjF.35_8703 | LmjF.35 | 2066649 | 2066936 | + |
| LmjF.36_8706 | LmjF.36 | 3057 | 3290 | - |
| LmjF.36_8716 | LmjF.36 | 27722 | 28474 | - |
| LmjF.36_8739 | LmjF.36 | 88109 | 88282 | - |
| LmjF.36_8748 | LmjF.36 | 109146 | 109421 | - |
| LmjF.36_8750 | LmjF.36 | 113226 | 113603 | - |
| LmjF.36_8752 | LmjF.36 | 116705 | 116977 | - |
| LmjF.36_8780 | LmjF.36 | 195130 | 195285 | + |
| LmjF.36_8793 | LmjF.36 | 278971 | 279117 | + |
| LmjF.36_8810 | LmjF.36 | 357841 | 358062 | + |
| LmjF.36_8821 | LmjF.36 | 408361 | 408513 | + |
| LmjF.36_8839 | LmjF.36 | 454548 | 454679 | + |
| LmjF.36_8844 | LmjF.36 | 465629 | 465847 | + |
| LmjF.36_8847 | LmjF.36 | 471444 | 471545 | + |
| LmjF.36_8848 | LmjF.36 | 473984 | 474910 | + |
| LmjF.36_8858 | LmjF.36 | 502251 | 502367 | - |
| LmjF.36_8859 | LmjF.36 | 502548 | 502817 | - |
| LmjF.36_8873 | LmjF.36 | 562958 | 563098 | - |
| LmjF.36_8876 | LmjF.36 | 570044 | 570160 | - |
| LmjF.36_8891 | LmjF.36 | 647433 | 647693 | - |
| LmjF.36_8911 | LmjF.36 | 726209 | 726394 | - |
| LmjF.36_8925 | LmjF.36 | 766358 | 766480 | - |
| LmjF.36_8926 | LmjF.36 | 768283 | 768750 | - |
| LmjF.36_8933 | LmjF.36 | 785300 | 785515 | + |
| LmjF.36_8938 | LmjF.36 | 801872 | 802165 | + |
| LmjF.36_8940 | LmjF.36 | 809192 | 809506 | + |
| LmjF.36_8945 | LmjF.36 | 825526 | 825642 | + |

| Gene ID | Chromosome | Start coordinate | Stop coordinate | Strand |
|---|---|---|---|---|
| LmjF.36_8947 | LmjF.36 | 832242 | 832715 | + |
| LmjF.36_8953 | LmjF.36 | 847825 | 848001 | + |
| LmjF.36_8977 | LmjF.36 | 938578 | 938694 | + |
| LmjF.36_8980 | LmjF.36 | 952996 | 953910 | + |
| LmjF.36_8981 | LmjF.36 | 954569 | 954913 | + |
| LmjF.36_9003 | LmjF.36 | 1038671 | 1038826 | - |
| LmjF.36_9009 | LmjF.36 | 1064136 | 1064288 | - |
| LmjF.36_9011 | LmjF.36 | 1084782 | 1084937 | - |
| LmjF.36_9013 | LmjF.36 | 1086056 | 1086373 | - |
| LmjF.36_9025 | LmjF.36 | 1129887 | 1130135 | - |
| LmjF.36_9026 | LmjF.36 | 1131248 | 1131577 | - |
| LmjF.36_9032 | LmjF.36 | 1153798 | 1154010 | - |
| LmjF.36_9037 | LmjF.36 | 1167774 | 1168001 | - |
| LmjF.36_9043 | LmjF.36 | 1187998 | 1188240 | - |
| LmjF.36_9048 | LmjF.36 | 1210882 | 1211073 | - |
| LmjF.36_9050 | LmjF.36 | 1213140 | 1213460 | - |
| LmjF.36_9067 | LmjF.36 | 1244192 | 1244782 | - |
| LmjF.36_9069 | LmjF.36 | 1247798 | 1248163 | - |
| LmjF.36_9078 | LmjF.36 | 1271328 | 1271660 | - |
| LmjF.36_9090 | LmjF.36 | 1308305 | 1308832 | - |
| LmjF.36_9097 | LmjF.36 | 1330640 | 1330972 | - |
| LmjF.36_9098 | LmjF.36 | 1330640 | 1330972 | - |
| LmjF.36_9107 | LmjF.36 | 1341383 | 1341532 | - |
| LmjF.36_9117 | LmjF.36 | 1364886 | 1365023 | - |
| LmjF.36_9124 | LmjF.36 | 1384817 | 1385164 | - |
| LmjF.36_9150 | LmjF.36 | 1443089 | 1443190 | + |
| LmjF.36_9173 | LmjF.36 | 1521065 | 1521169 | + |
| LmjF.36_9174 | LmjF.36 | 1521186 | 1521569 | + |
| LmjF.36_9178 | LmjF.36 | 1530894 | 1530986 | + |
| LmjF.36_9196 | LmjF.36 | 1590191 | 1590316 | + |
| LmjF.36_9197 | LmjF.36 | 1590725 | 1590820 | + |
| LmjF.36_9198 | LmjF.36 | 1590922 | 1591164 | + |
| LmjF.36_9211 | LmjF.36 | 1656043 | 1656531 | + |
| LmjF.36_9215 | LmjF.36 | 1673744 | 1673896 | + |
| LmjF.36_9224 | LmjF.36 | 1716495 | 1716641 | + |
| LmjF.36_9240 | LmjF.36 | 1767587 | 1767703 | + |
| LmjF.36_9253 | LmjF.36 | 1804405 | 1804536 | + |
| LmjF.36_9255 | LmjF.36 | 1807821 | 1808138 | + |
| LmjF.36_9257 | LmjF.36 | 1815545 | 1815769 | + |
| LmjF.36_9261 | LmjF.36 | 1826404 | 1826865 | + |
| LmjF.36_9269 | LmjF.36 | 1853699 | 1853824 | + |
| LmjF.36_9289 | LmjF.36 | 1921984 | 1922226 | - |
| LmjF.36_9307 | LmjF.36 | 1979234 | 1979332 | - |
| LmjF.36_9308 | LmjF.36 | 1980690 | 1980821 | - |
| LmjF.36_9327 | LmjF.36 | 2054519 | 2054746 | - |
| LmjF.36_9330 | LmjF.36 | 2060464 | 2060745 | - |
| LmjF.36_9347 | LmjF.36 | 2111532 | 2111687 | - |
| LmjF.36_9372 | LmjF.36 | 2193280 | 2193531 | - |
| LmjF.36_9380 | LmjF.36 | 2213553 | 2213696 | - |
| LmjF.36_9400 | LmjF.36 | 2289198 | 2289302 | - |
| LmjF.36_9406 | LmjF.36 | 2310785 | 2310970 | - |
| LmjF.36_9415 | LmjF.36 | 2342137 | 2342271 | - |
| LmjF.36_9416 | LmjF.36 | 2343096 | 2343245 | - |
| LmjF.36_9418 | LmjF.36 | 2347826 | 2347930 | - |
| LmjF.36_9419 | LmjF.36 | 2347649 | 2347804 | - |
| LmjF.36_9422 | LmjF.36 | 2355651 | 2356004 | - |
| LmjF.36_9426 | LmjF.36 | 2365418 | 2365945 | - |
| LmjF.36_9436 | LmjF.36 | 2396231 | 2396374 | - |
| LmjF.36_9440 | LmjF.36 | 2407793 | 2408038 | - |
| LmjF.36_9447 | LmjF.36 | 2440572 | 2440724 | - |
| LmjF.36_9451 | LmjF.36 | 2462399 | 2462629 | - |
| LmjF.36_9453 | LmjF.36 | 2465939 | 2466082 | - |
| LmjF.36_9460 | LmjF.36 | 2484516 | 2484968 | - |
| LmjF.36_9487 | LmjF.36 | 2567032 | 2567313 | - |
| LmjF.36_9494 | LmjF.36 | 2583135 | 2583392 | - |
| LmjF.36_9498 | LmjF.36 | 2595086 | 2595283 | - |
| LmjF.36_9515 | LmjF.36 | 2645478 | 2645816 | - |
| LmjF.36_9520 | LmjF.36 | 2657059 | 2657193 | - |
| LmjF.36_9528 | LmjF.36 | 2672836 | 2673333 | - |

# Appendix 3: Mouse infection experimental design

| Sample ID | SRA accession | Infection status/timepoint | Batch | Number of reads that pass Illumina | Number of reads after trimming | Number of reads mapped - mouse | % of reads mapped - mouse | Number of reads mapped - *L. major* | % of reads mapped - *L. major* | % macrophages infected | # parasites per 100 macrophages |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HPGL0165 | SRR1460767 | metacyclic | A | 45,492,872 | 44,325,986 | n/a | n/a | 41,811,082 | 94.3 | n/a | n/a |
| HPGL0166 | SRR1460724 | uninfected, 4 hrs | A | 40,524,394 | 39,910,722 | 38,605,849 | 96.7 | n/a | n/a | n/a | n/a |
| HPGL0167 | SRR1460725 | infected, 4 hrs | A | 90,990,046 | 88,980,750 | 51,799,068 | 58.2 | 33,630,041 | 37.8 | unknown | unknown |
| HPGL0168 | SRR1460726 | uninfected, 24 hrs | A | 48,172,914 | 47,278,686 | 45,655,539 | 96.6 | n/a | n/a | n/a | n/a |
| HPGL0169 | SRR1460727 | infected, 24 hrs | A | 100,702,572 | 98,641,188 | 73,902,795 | 74.9 | 20,709,933 | 21.0 | unknown | unknown |
| HPGL0170 | SRR1460728 | uninfected, 48 hrs | A | 47,092,436 | 46,134,592 | 44,569,557 | 96.6 | n/a | n/a | n/a | n/a |
| HPGL0171 | SRR1460729 | infected, 48 hrs | A | 104,431,682 | 102,285,498 | 83,340,313 | 81.5 | 14,995,629 | 14.7 | unknown | unknown |
| HPGL0172 | SRR1460730 | uninfected, 72 hrs | A | 42,210,532 | 41,341,096 | 40,014,106 | 96.8 | n/a | n/a | n/a | n/a |
| HPGL0173 | SRR1460731 | infected, 72 hrs | A | 72,318,516 | 70,901,794 | 63,595,008 | 89.7 | 4,795,444 | 6.8 | unknown | unknown |
| HPGL0230 | SRR1460772 | metacyclic | B | 84,056,646 | 82,084,958 | n/a | n/a | 77,364,378 | 94.2 | n/a | n/a |
| HPGL0231 | SRR1460732 | uninfected, 4 hrs | B | 96,215,864 | 94,854,552 | 92,101,618 | 97.1 | n/a | n/a | n/a | n/a |
| HPGL0232 | SRR1460733 | infected, 4 hrs | B | 92,160,718 | 90,299,686 | 39,344,747 | 43.6 | 46,979,374 | 52.0 | 100.0 | 407 |
| HPGL0233 | SRR1460734 | uninfected, 24 hrs | B | 82,830,256 | 81,677,132 | 79,383,322 | 97.2 | n/a | n/a | n/a | n/a |
| HPGL0234 | SRR1460735 | infected, 24 hrs | B | 103,092,448 | 101,118,460 | 64,280,705 | 63.6 | 33,073,618 | 32.7 | 89.8 | 194 |
| HPGL0235 | SRR1460736 | uninfected, 48 hrs | B | 81,211,534 | 80,061,410 | 77,743,701 | 97.1 | n/a | n/a | n/a | n/a |
| HPGL0236 | SRR1460737 | infected, 48 hrs | B | 111,687,434 | 109,753,574 | 90,222,254 | 82.2 | 15,852,226 | 14.4 | 79.6 | 226 |
| HPGL0237 | SRR1460738 | uninfected, 72 hrs | B | 92,915,856 | 91,606,470 | 89,142,777 | 97.3 | n/a | n/a | n/a | n/a |
| HPGL0238 | SRR1460739 | infected, 72 hrs | B | 138,070,730 | 135,814,536 | 119,555,532 | 88.0 | 12,313,978 | 9.1 | 88.9 | 279 |
| HPGL0329 | SRR2136702 | metacyclic | C | 64,860,308 | 63,200,662 | n/a | n/a | 59,727,854 | 94.5 | n/a | n/a |
| HPGL0388 | SRR1460740 | uninfected, 4 hrs | C | 126,914,156 | 122,765,288 | 117,718,796 | 95.9 | n/a | n/a | n/a | n/a |
| HPGL0389 | SRR1460741 | infected, 4 hrs | C | 109,064,742 | 105,845,790 | 87,737,777 | 82.9 | 13,715,013 | 13.0 | 100.0 | 716 |
| HPGL0390 | SRR1460742 | uninfected, 24 hrs | C | 100,298,240 | 97,739,428 | 95,367,829 | 97.6 | n/a | n/a | n/a | n/a |
| HPGL0391 | SRR1460743 | infected, 24 hrs | C | 120,601,610 | 116,947,098 | 102,128,561 | 87.3 | 10,499,445 | 9.0 | 60.0 | 150 |
| HPGL0392 | SRR1460744 | uninfected, 48 hrs | C | 116,767,874 | 113,646,410 | 110,780,532 | 97.5 | n/a | n/a | n/a | n/a |
| HPGL0393 | SRR1460745 | infected, 48 hrs | C | 106,096,990 | 102,294,678 | 92,253,573 | 90.2 | 4,919,065 | 4.8 | 69.1 | 204 |
| HPGL0394 | SRR1460746 | uninfected, 72 hrs | C | 106,521,540 | 101,703,022 | 94,661,771 | 93.1 | n/a | n/a | n/a | n/a |
| HPGL0395 | SRR1460747 | infected, 72 hrs | C | 122,268,922 | 110,361,056 | 97,588,136 | 88.4 | 3,856,201 | 3.5 | 72.6 | 201 |

**Table 13: Mouse infection experimental design.**

Samples are listed using a sample identifier (HPGL----), which is referenced in the record stored at the SRA. Sample type is provided as metacyclic (contains only *L. major* metacyclic promastigote RNA), uninfected (contains only mouse RNA), or infected (contains both mouse and *L. major* RNA). Experimental batches (A, B, and C) are defined based on the start date of the experiment. The number of reads sequenced, number of reads that passed quality trimming, and number and percentage of trimmed reads that mapped to the mouse (mm10) and/or *L. major* genomes (v. 6.0) are provided, as are the percentage of infected macrophages and number of parasites observed per 100 macrophages for the 2 replicates for which these data were collected.

# Appendix 4: Human infection experimental design

### Table 14: Human infection experimental design.

Samples are listed using an internal lab sample identifier (HPGL----), which is referenced in the record stored at the Short Read Archive (SRA; accession numbers specified). Sample type is provided as metacyclic (contains only *Leishmania* metacyclic promastigote RNA, "LM" for *L. major* or "LA" for *L. amazonensis*), uninfected (contains only human RNA), infected (contains both human RNA and *Leishmania* RNA), or beads (contains only human RNA). Experimental batches (A-F) are defined based on the start date of the experiment with each batch originating from a separate growth of cells. The patient identifier indicates the human donor who supplied the monocytes used for the experiment. The percentage of infected human macrophages and number of parasites observed per 100 macrophages are reported, as are the number of reads sequenced, number of reads that passed quality trimming, and number and percentage of trimmed reads that mapped to the human (hg19) and/or *Leishmania (L. major* v. 6.0 or *L. mexicana* v 8.1) genomes.

| Sample ID | SRA accession | Infection status/timepoint | Batch: Patient | % macrophages infected | # parasites per 100 macrophages |
|---|---|---|---|---|---|
| HPGL0364 | SRR2136703 | LM metacyclic | A:n/a | n/a | n/a |
| HPGL0365 | SRR2155070 | uninfected, 4 hrs | A:H2 | n/a | n/a |
| HPGL0366 | SRR2155072 | LM-infected, 4 hrs | A:H2 | 100.0 | 524 |
| HPGL0367 | SRR2155073 | uninfected, 24 hrs | A:H2 | n/a | n/a |
| HPGL0368 | SRR2155074 | LM-infected, 24 hrs | A:H2 | 92.0 | 1220 |
| HPGL0369 | SRR2155075 | uninfected, 48 hrs | A:H2 | n/a | n/a |
| HPGL0370 | SRR2155078 | LM-infected, 48 hrs | A:H2 | 100.0 | 1778 |
| HPGL0371 | SRR2155082 | uninfected, 72 hrs | A:H2 | n/a | n/a |
| HPGL0372 | SRR2155085 | LM-infected, 72 hrs | A:H2 | 91.0 | 1896 |
| HPGL0374 | SRR2136708 | LM metacyclic | B:n/a | n/a | n/a |
| HPGL0375 | SRR2155101 | uninfected, 4 hrs | B:H3 | n/a | n/a |
| HPGL0376 | SRR2155105 | LM-infected, 4 hrs | B:H3 | 100.0 | 842 |
| HPGL0377 | SRR2155143 | uninfected, 24 hrs | B:H3 | n/a | n/a |
| HPGL0378 | SRR2155160 | LM-infected, 24 hrs | B:H3 | 100.0 | 2011 |
| HPGL0379 | SRR2155161 | uninfected, 48 hrs | B:H3 | n/a | n/a |
| HPGL0380 | SRR2155162 | LM-infected, 48 hrs | B:H3 | 98.0 | 2471 |
| HPGL0381 | SRR2155163 | uninfected, 72 hrs | B:H3 | n/a | n/a |
| HPGL0382 | SRR2155164 | LM-infected, 72 hrs | B:H3 | 96.0 | 1753 |
| HPGL0397 | SRR2136720 | LM metacyclic | C:n/a | n/a | n/a |
| HPGL0398 | SRR2155165 | uninfected, 4 hrs | C:H4 | n/a | n/a |
| HPGL0399 | SRR2155166 | LM-infected, 4 hrs | C:H4 | 96.0 | 752 |
| HPGL0400 | SRR2156106 | uninfected, 24 hrs | C:H4 | n/a | n/a |
| HPGL0401 | SRR2156107 | LM-infected, 24 hrs | C:H4 | 90.0 | 1245 |
| HPGL0402 | SRR2156108 | uninfected, 48 hrs | C:H4 | n/a | n/a |
| HPGL0403 | SRR2156109 | LM-infected, 48 hrs | C:H4 | 96.0 | 1996 |
| HPGL0404 | SRR2156110 | uninfected, 72 hrs | C:H4 | n/a | n/a |
| HPGL0405 | SRR2156117 | LM-infected, 72 hrs | C:H4 | 98.0 | 1471 |

| Sample ID | SRA accession | Infection status/timepoint | Batch: Patient | % macrophages infected | # parasites per 100 macrophages |
|---|---|---|---|---|---|
| HPGL0435 | SRR2171253 | LA metacyclic | D:n/a | n/a | n/a |
| HPGL0436 | SRR2156270 | uninfected, 4 hrs | D:H5 | n/a | n/a |
| HPGL0437 | SRR2156271 | LA-infected, 4 hrs | D:H5 | 75.0 | 453 |
| HPGL0438 | SRR2156272 | beads, 4 hrs | D:H5 | n/a | n/a |
| HPGL0439 | SRR2156273 | uninfected, 24 hrs | D:H5 | n/a | n/a |
| HPGL0440 | SRR2156274 | LA-infected, 24 hrs | D:H5 | 73.0 | 304 |
| HPGL0441 | SRR2156854 | beads, 24 hrs | D:H5 | n/a | n/a |
| HPGL0442 | SRR2156855 | uninfected, 48 hrs | D:H5 | n/a | n/a |
| HPGL0443 | SRR2156856 | LA-infected, 48 hrs | D:H5 | 65.0 | 225 |
| HPGL0444 | SRR2156858 | beads, 48 hrs | D:H5 | n/a | n/a |
| HPGL0445 | SRR2156859 | uninfected, 72 hrs | D:H5 | n/a | n/a |
| HPGL0446 | SRR2156860 | LA-infected, 72 hrs | D:H5 | 55.0 | 246 |
| HPGL0452 | SRR2156862 | beads, 72 hrs | D:H5 | n/a | n/a |
| HPGL0454 | SRR2171254 | LA metacyclic | E:n/a | n/a | n/a |
| HPGL0456 | SRR2171252 | LM metacyclic | E:n/a | n/a | n/a |
| HPGL0457 | SRR2156863 | uninfected, 4 hrs | E:H2 | n/a | n/a |
| HPGL0458 | SRR2163237 | LA-infected, 4 hrs | E:H2 | 100.0 | 1164 |
| HPGL0459 | SRR2163238 | LM-infected, 4 hrs | E:H2 | 96.0 | 752 |
| HPGL0460 | SRR2163239 | beads, 4 hrs | E:H2 | n/a | n/a |
| HPGL0461 | SRR2163240 | uninfected, 24 hrs | E:H2 | n/a | n/a |
| HPGL0462 | SRR2163242 | LA-infected, 24 hrs | E:H2 | 100.0 | 1723 |
| HPGL0463 | SRR2163251 | LM-infected, 24 hrs | E:H2 | 93.0 | 1527 |
| HPGL0464 | SRR2163256 | beads, 24 hrs | E:H2 | n/a | n/a |
| HPGL0465 | SRR2163270 | uninfected, 48 hrs | E:H2 | n/a | n/a |
| HPGL0466 | SRR2163272 | LA-infected, 48 hrs | E:H2 | 95.0 | 1455 |
| HPGL0467 | SRR2163273 | LM-infected, 48 hrs | E:H2 | 91.0 | 2485 |
| HPGL0468 | SRR2163274 | beads, 48 hrs | E:H2 | n/a | n/a |
| HPGL0469 | SRR2163275 | uninfected, 72 hrs | E:H2 | n/a | n/a |
| HPGL0470 | SRR2163276 | LA-infected, 72 hrs | E:H2 | 98.0 | 1149 |
| HPGL0492 | SRR2171255 | LA metacyclic | F:n/a | n/a | n/a |
| HPGL0494 | SRR2136722 | LM metacyclic | F:n/a | n/a | n/a |
| HPGL0495 | SRR2163278 | uninfected, 4 hrs | F:H3 | n/a | n/a |
| HPGL0496 | SRR2163279 | LA-infected, 4 hrs | F:H3 | 61.0 | 156 |
| HPGL0497 | SRR2163280 | LM-infected, 4 hrs | F:H3 | 69.0 | 237 |
| HPGL0498 | SRR2163281 | beads, 4 hrs | F:H3 | n/a | n/a |
| HPGL0499 | SRR2163282 | uninfected, 24 hrs | F:H3 | n/a | n/a |
| HPGL0500 | SRR2163283 | LA-infected, 24 hrs | F:H3 | 70.0 | 202 |
| HPGL0501 | SRR2163285 | LM-infected, 24 hrs | F:H3 | 71.0 | 243 |
| HPGL0502 | SRR2163286 | beads, 24 hrs | F:H3 | n/a | n/a |
| HPGL0503 | SRR2163287 | uninfected, 48 hrs | F:H3 | n/a | n/a |
| HPGL0504 | SRR2163289 | LA-infected, 48 hrs | F:H3 | 71.0 | 195 |
| HPGL0505 | SRR2163290 | LM-infected, 48 hrs | F:H3 | 60.0 | 186 |
| HPGL0506 | SRR2163291 | beads, 48 hrs | F:H3 | n/a | n/a |
| HPGL0507 | SRR2163292 | uninfected, 72 hrs | F:H3 | n/a | n/a |
| HPGL0508 | SRR2163399 | LA-infected, 72 hrs | F:H3 | 82.0 | 284 |
| HPGL0509 | SRR2163400 | LM-infected, 72 hrs | F:H3 | 62.0 | 258 |
| HPGL0510 | SRR2163401 | beads, 72 hrs | F:H3 | n/a | n/a |

| Sample ID | # reads passing Illumina filter | # reads after trimming | # reads mapped - human | % reads map - human | # reads mapped - *Leish.* | % reads map - *Leishmania* |
|---|---|---|---|---|---|---|
| HPGL0364 | 78,758,128 | 75,672,222 | n/a | n/a | 72,113,919 | 95.30 |
| HPGL0365 | 91,388,692 | 88,599,546 | 82,744,345 | 93.4 | n/a | n/a |
| HPGL0366 | 86,023,690 | 83,531,092 | 57,118,665 | 68.4 | 22,370,540 | 26.8 |
| HPGL0367 | 85,514,140 | 83,432,828 | 79,367,982 | 95.1 | n/a | n/a |
| HPGL0368 | 72,432,090 | 69,867,190 | 50,790,608 | 72.7 | 15,839,498 | 22.7 |
| HPGL0369 | 77,830,926 | 76,033,626 | 71,860,400 | 94.5 | n/a | n/a |
| HPGL0370 | 78,550,138 | 76,390,670 | 58,163,491 | 76.1 | 13,096,474 | 17.1 |
| HPGL0371 | 79,985,738 | 77,733,188 | 72,054,284 | 92.7 | n/a | n/a |
| HPGL0372 | 67,952,622 | 65,888,202 | 51,427,294 | 78.1 | 11,428,289 | 17.3 |
| HPGL0374 | 78,737,648 | 75,494,056 | n/a | n/a | 71,993,760 | 95.36 |
| HPGL0375 | 78,471,106 | 76,111,004 | 71,737,258 | 94.3 | n/a | n/a |
| HPGL0376 | 83,905,020 | 81,439,468 | 53,650,676 | 65.9 | 23,925,977 | 29.4 |
| HPGL0377 | 76,653,764 | 74,425,690 | 69,948,299 | 94.0 | n/a | n/a |
| HPGL0378 | 76,572,998 | 74,313,368 | 42,840,828 | 57.6 | 18,783,218 | 25.3 |
| HPGL0379 | 76,283,630 | 74,104,972 | 69,384,638 | 93.6 | n/a | n/a |
| HPGL0380 | 75,140,814 | 72,663,760 | 49,226,213 | 67.7 | 20,019,048 | 27.6 |
| HPGL0381 | 73,605,514 | 71,071,712 | 67,162,430 | 94.5 | n/a | n/a |
| HPGL0382 | 71,380,904 | 69,295,114 | 49,294,945 | 71.1 | 16,888,389 | 24.4 |
| HPGL0397 | 97,701,708 | 92,092,826 | n/a | n/a | 85,865,457 | 93.24 |
| HPGL0398 | 100,195,200 | 96,731,678 | 89,365,001 | 92.4 | n/a | n/a |
| HPGL0399 | 90,467,044 | 87,516,500 | 72,210,158 | 82.5 | 11,113,911 | 12.7 |
| HPGL0400 | 98,373,732 | 94,028,218 | 86,604,115 | 92.1 | n/a | n/a |
| HPGL0401 | 85,316,742 | 81,857,110 | 64,375,678 | 78.6 | 12,366,309 | 15.1 |
| HPGL0402 | 86,638,390 | 84,086,304 | 78,432,710 | 93.3 | n/a | n/a |
| HPGL0403 | 98,025,186 | 94,787,210 | 77,527,704 | 81.8 | 12,446,371 | 13.1 |
| HPGL0404 | 104,376,924 | 100,715,178 | 92,856,028 | 92.2 | n/a | n/a |
| HPGL0405 | 89,000,552 | 84,864,346 | 61,622,607 | 72.6 | 16,923,589 | 19.9 |
| HPGL0435 | 88,108,512 | 85,602,946 | n/a | n/a | 55,592,305 | 64.94 |
| HPGL0436 | 87,679,330 | 86,252,844 | 81,956,083 | 95.0 | n/a | n/a |
| HPGL0437 | 98,036,962 | 96,501,364 | 77,319,838 | 80.1 | 10,321,591 | 10.7 |
| HPGL0438 | 64,871,106 | 63,825,098 | 60,629,949 | 95.0 | n/a | n/a |
| HPGL0439 | 51,292,194 | 50,291,050 | 47,140,587 | 93.7 | n/a | n/a |
| HPGL0440 | 54,855,220 | 54,035,386 | 46,743,009 | 86.5 | 3,299,237 | 6.1 |
| HPGL0441 | 55,086,382 | 54,242,308 | 51,571,492 | 95.1 | n/a | n/a |
| HPGL0442 | 79,454,644 | 78,067,616 | 74,307,932 | 95.2 | n/a | n/a |
| HPGL0443 | 85,079,824 | 83,475,144 | 71,611,331 | 85.8 | 5,562,379 | 6.7 |
| HPGL0444 | 95,550,506 | 93,420,086 | 88,784,659 | 95.0 | n/a | n/a |
| HPGL0445 | 71,491,834 | 70,297,022 | 66,254,603 | 94.2 | n/a | n/a |
| HPGL0446 | 79,552,452 | 78,237,034 | 64,119,530 | 82.0 | 7,390,792 | 9.4 |
| HPGL0452 | 86,600,982 | 85,198,052 | 81,416,206 | 95.6 | n/a | n/a |
| HPGL0454 | 87,184,878 | 85,009,962 | n/a | n/a | 57,045,902 | 67.10 |
| HPGL0456 | 102,691,716 | 99,950,410 | n/a | n/a | 95,206,399 | 95.25 |
| HPGL0457 | 107,773,908 | 102,908,504 | 96,570,364 | 93.8 | n/a | n/a |
| HPGL0458 | 99,857,328 | 97,489,652 | 54,703,033 | 56.1 | 26,626,832 | 27.3 |
| HPGL0459 | 86,248,244 | 84,299,440 | 55,033,854 | 65.3 | 24,434,847 | 29.0 |
| HPGL0460 | 98,522,766 | 93,688,002 | 87,781,217 | 93.7 | n/a | n/a |
| HPGL0461 | 103,760,218 | 100,592,660 | 95,298,430 | 94.7 | n/a | n/a |
| HPGL0462 | 89,674,336 | 87,415,806 | 39,536,537 | 45.2 | 30,840,179 | 35.3 |
| HPGL0463 | 89,165,916 | 84,658,058 | 56,051,998 | 66.2 | 23,740,347 | 28.0 |
| HPGL0464 | 98,425,396 | 96,227,604 | 90,804,430 | 94.4 | n/a | n/a |
| HPGL0465 | 99,389,800 | 95,816,874 | 91,110,475 | 95.1 | n/a | n/a |
| HPGL0466 | 98,728,132 | 95,668,548 | 46,719,663 | 48.8 | 32,161,999 | 33.6 |
| HPGL0467 | 82,501,414 | 79,798,494 | 57,139,084 | 71.6 | 18,560,152 | 23.3 |
| HPGL0468 | 97,391,724 | 92,107,018 | 87,076,039 | 94.5 | n/a | n/a |
| HPGL0469 | 102,498,020 | 100,814,392 | 95,736,413 | 95.0 | n/a | n/a |
| HPGL0470 | 88,416,478 | 86,556,644 | 41,360,179 | 47.8 | 29,712,764 | 34.3 |
| HPGL0492 | 90,858,848 | 73,250,388 | n/a | n/a | 48,081,033 | 65.64 |
| HPGL0494 | 137,414,688 | 104,282,894 | n/a | n/a | 86,802,339 | 83.24 |
| HPGL0495 | 101,126,484 | 91,249,136 | 84,560,144 | 92.7 | n/a | n/a |
| HPGL0496 | 87,931,978 | 72,505,814 | 60,960,166 | 84.1 | 4,211,132 | 5.8 |
| HPGL0497 | 86,696,524 | 82,757,690 | 74,272,939 | 89.7 | 3,548,297 | 4.3 |
| HPGL0498 | 96,026,066 | 92,353,698 | 86,906,401 | 94.1 | n/a | n/a |
| HPGL0499 | 97,186,514 | 85,554,738 | 79,374,626 | 92.8 | n/a | n/a |
| HPGL0500 | 90,560,302 | 86,774,262 | 73,298,146 | 84.5 | 6,515,813 | 7.5 |
| HPGL0501 | 81,197,850 | 78,566,274 | 71,846,173 | 91.4 | 2,724,553 | 3.5 |
| HPGL0502 | 98,522,858 | 84,187,684 | 77,370,844 | 91.9 | n/a | n/a |
| HPGL0503 | 94,794,662 | 87,256,992 | 82,034,097 | 94.0 | n/a | n/a |
| HPGL0504 | 64,195,758 | 57,625,032 | 51,037,001 | 88.6 | 1,905,629 | 3.3 |
| HPGL0505 | 93,036,854 | 85,931,898 | 79,106,619 | 92.1 | 1,488,979 | 1.7 |
| HPGL0506 | 108,582,528 | 89,718,944 | 84,005,453 | 93.6 | n/a | n/a |
| HPGL0507 | 55,576,648 | 52,618,126 | 49,188,871 | 93.5 | n/a | n/a |
| HPGL0508 | 50,760,444 | 45,837,500 | 39,593,275 | 86.4 | 2,366,019 | 5.2 |
| HPGL0509 | 61,900,532 | 60,479,728 | 55,555,352 | 91.9 | 1,490,397 | 2.5 |
| HPGL0510 | 59,588,422 | 57,386,442 | 54,056,078 | 94.2 | n/a | n/a |

# References


Akopyants NS, Matlib RS, Bukanova EN, Smeds MR, Brownstein BH, Stormo GD, Beverley SM. 2004. Expression profiling using random genomic DNA microarrays identifies differentially expressed genes associated with three major developmental stages of the protozoan parasite *leishmania major*. *Mol Biochem Parasitol.* 136: 71-86.

Alexander J, Coombs GH, Mottram JC. 1998. *Leishmania* mexicana cysteine proteinase-deficient mutants have attenuated virulence for mice and potentiate a th1 response. *J Immunol.* 161: 6794-6801.

Allocco JJ, Donald R, Zhong T, Lee A, Tang YS, Hendrickson RC, Liberator P, Nare B. 2006. Inhibitors of casein kinase 1 block the growth of *leishmania major* promastigotes in vitro. *Int J Parasitol.* 36: 1249-1259.

Almeida R, Gilmartin BJ, McCann SH, Norrish A, Ivens AC, Lawson D, Levick MP, Smith DF, Dyall SD, Vetrie D, et al. 2004. Expression profiling of the *leishmania* life cycle: CDNA arrays identify developmentally regulated genes present but not annotated in the genome. *Mol Biochem Parasitol.* 136: 87-100.

Almeida RP, Barral-Netto M, De Jesus AM, De Freitas LA, Carvalho EM, Barral A. 1996. Biological behavior of *leishmania amazonensis* isolated from humans with cutaneous, mucosal, or visceral leishmaniasis in BALB/C mice. *Am J Trop Med Hyg.* 54: 178-184.

Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M, WHO Leishmaniasis Control Team. 2012. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One.* 7: e35671.

Ambit A, Woods KL, Cull B, Coombs GH, Mottram JC. 2011. Morphological events during the cell cycle of *leishmania major*. *Eukaryot Cell.* 10: 1429-1438.

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013. Count-based differential expression analysis of RNA sequencing data using R and bioconductor. *Nat Protoc.* 8: 1765-1786.

Anders S, Pyl PT, Huber W. 2014. HTSeq-a python framework to work with high-throughput sequencing data. *Bioinformatics.*doi:10.1093/bioinformatics/btu638.

Andrews NW, Chakrabarti S. 2005. There's more to life than neurotransmission: The regulation of exocytosis by synaptotagmin VII. *Trends Cell Biol.* 15: 626-631.

Arango Duque G, Fukuda M, Descoteaux A. 2013. Synaptotagmin XI regulates phagocytosis and cytokine secretion in macrophages. *J Immunol.* 190: 1737-1745.

Arango Duque G, Fukuda M, Turco SJ, Stäger S, Descoteaux A. 2014. *Leishmania* promastigotes induce cytokine secretion in macrophages through the degradation of synaptotagmin XI. *J Immunol.* 193: 2363-2372.

Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, et al. 2010. TriTrypDB: A functional genomic resource for the trypanosomatidae. *Nucleic Acids Res.* 38: D457-D462.

Babadjanova Z, Wiedinger K, Gosselin EJ, Bitsaktsis C. 2015. Targeting of a fixed bacterial immunogen to fc receptors reverses the anti-inflammatory properties of the gram-negative bacterium, *francisella tularensis*, during the early stages of infection. *PLoS One.* 10: e0129981.

Baram D, Adachi R, Medalia O, Tuvim M, Dickey BF, Mekori YA, Sagi-Eisenberg R. 1999. Synaptotagmin II negatively regulates ca2+-triggered exocytosis of lysosomes in mast cells. *J Exp Med.* 189: 1649-1658.

Bates PA. 2007. Transmission of *leishmania* metacyclic promastigotes by phlebotomine sand flies. *Int J Parasitol.* 37: 1097-1106.

Bengs F, Scholz A, Kuhn D, Wiese M. 2005. LmxMPK9, a mitogen-activated protein kinase homologue affects flagellar length in *leishmania mexicana*. *Mol Microbiol.* 55: 1606-1615.

Bennett CL, Misslitz A, Colledge L, Aebischer T, Blackburn CC. 2001. Silent infection of bone marrow-derived dendritic cells by *leishmania mexicana* amastigotes. *Eur J Immunol.* 31: 876-883.

Bent ZW, Poorey K, Brazel DM, LaBauve AE, Sinha A, Curtis DJ, House SE, Tew KE, Hamblin RY, Williams KP, et al. 2015. Transcriptomic analysis of *yersinia enterocolitica* biovar 1B infecting murine macrophages reveals new mechanisms of extracellular and intracellular survival. *Infect Immun.* 83: 2672-2685.

Beverley SM, Turco SJ. 1998. Lipophosphoglycan (LPG) and the identification of virulence genes in the protozoan parasite *leishmania*. *Trends Microbiol.* 6: 35-40.

Biswas A, Bhattacharya A, Das PK. 2011. Role of camp signaling in the survival and infectivity of the protozoan parasite, *leishmania donovani*. *Mol Biol Int.* 2011: 782971.

Blaineau C, Tessier M, Dubessay P, Tasse L, Crobu L, Pagès M, Bastien P. 2007. A novel microtubule-depolymerizing kinesin involved in length control of a eukaryotic flagellum. *Curr Biol.* 17: 778-782.

Blumenthal A, Lauber J, Hoffmann R, Ernst M, Keller C, Buer J, Ehlers S, Reiling N. 2005. Common and unique gene expression signatures of human macrophages in response to four strains of *mycobacterium avium* that differ in their growth and persistence characteristics. *Infect Immun.* 73: 3330-3341.

Bogdan C, Donhauser N, Döring R, Röllinghoff M, Diefenbach A, Rittig MG. 2000. Fibroblasts as host cells in latent leishmaniosis. *J Exp Med.* 191: 2121-2130.

Bogdan C, Gessner A, Solbach W, Röllinghoff M. 1996. Invasion, control and persistence of *leishmania* parasites. *Curr Opin Immunol.* 8: 517-525.

Boitz JM, Ullman B. 2013. Adenine and adenosine salvage in *leishmania donovani*. *Mol Biochem Parasitol.* 190: 51-55.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics.* 30: 2114-2120.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 19: 185-193.

Brittingham A, Miller MA, Donelson JE, Wilson ME. 2001. Regulation of GP63 mRNA stability in promastigotes of virulent and attenuated *leishmania chagasi. Mol Biochem Parasitol.* 112: 51-59.

Brittingham A, Morrison CJ, McMaster WR, McGwire BS, Chang KP, Mosser DM. 1995. Role of the leishmania surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *J Immunol.* 155: 3102-3111.

Buates S, Matlashewski G. 2001. General suppression of macrophage gene expression during *leishmania donovani* infection. *J Immunol.* 166: 3416-3422.

Cao M, Li G, Pan J. 2009. Regulation of cilia assembly, disassembly, and length by protein phosphorylation. *Methods Cell Biol.* 94: 333-346.

Chakraborty A, Das I, Datta R, Sen B, Bhattacharyya D, Mandal C, Datta AK. 2002. A single-domain cyclophilin from *leishmania donovani* reactivates soluble aggregates of adenosine kinase by isomerase-independent chaperone function. *J Biol Chem.* 277: 47451-47460.

Charest H, Zhang WW, Matlashewski G. 1996. The developmental expression of *leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3'-untranslated region. *J Biol Chem.* 271: 17081-17090.

Charmoy M, Brunner-Agten S, Aebischer D, Auderset F, Launois P, Milon G, Proudfoot AEI, Tacchini-Cottier F. 2010. Neutrophil-derived CCL3 is essential for the rapid recruitment of dendritic cells to the site of *leishmania major* inoculation in resistant mice. *PLoS Pathog.* 6: e1000755.

Chaussabel D, Semnani RT, McDowell MA, Sacks D, Sher A, Nutman TB. 2003. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites. *Blood.* 102: 672-681.

Clayton C, Shapira M. 2007. Post-transcriptional regulation of gene expression in trypanosomes and *leishmanias*. *Mol Biochem Parasitol.* 156: 93-101.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 25: 1422-1423.

Cohen-Freue G, Holzer TR, Forney JD, McMaster WR. 2007. Global gene expression in *leishmania*. *Int J Parasitol.* 37: 1077-1086.

Corradin S, Ransijn A, Corradin G, Roggero MA, Schmitz AA, Schneider P, Mauël J, Vergères G. 1999. MARCKS-related protein (MRP) is a substrate for the *leishmania major* surface protease leishmanolysin (gp63). *J Biol Chem.* 274: 25411-25418.

Coughlin BC, Teixeira SM, Kirchhoff LV, Donelson JE. 2000. Amastin mRNA abundance in *trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. *J Biol Chem.* 275: 12051-12060.

Coulson RM, Connor V, Chen JC, Ajioka JW. 1996. Differential expression of *leishmania major* beta-tubulin genes during the acquisition of promastigote infectivity. *Mol Biochem Parasitol.* 82: 227-236.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res.* 14: 1188-1190.

Czibener C, Sherer NM, Becker SM, Pypaert M, Hui E, Chapman ER, Mothes W, Andrews NW. 2006. Ca2+ and synaptotagmin vii-dependent delivery of lysosomal membrane to nascent phagosomes. *J Cell Biol.* 174: 997-1007.

Dan-Goor M, Nasereddin A, Jaber H, Jaffe CL. 2013. Identification of a secreted casein kinase 1 in *leishmania donovani*: Effect of protein over expression on parasite growth and virulence. *PLoS One.* 8: e79287.

da Silva R, Sacks DL. 1987. Metacyclogenesis is a major determinant of *leishmania* promastigote virulence and attenuation. *Infect Immun.* 55: 2802-2806.

Da Silva RP, Hall BF, Joiner KA, Sacks DL. 1989. CR1, the c3b receptor, mediates binding of infective *leishmania major* metacyclic promastigotes to human macrophages. *J Immunol.* 143: 617-622.

Depledge DP, Evans KJ, Ivens AC, Aziz N, Maroof A, Kaye PM, Smith DF. 2009. Comparative expression profiling of *leishmania*: Modulation in gene expression between species and in different host genetic backgrounds. *PLoS Negl Trop Dis.* 3: e476.

Derelle R, Momose T, Manuel M, Da Silva C, Wincker P, Houliston E. 2010. Convergent origins and rapid evolution of spliced leader *trans*-splicing in metazoa: Insights from the *ctenophora* and *hydrozoa*. *RNA.* 16: 696-707.

Desai PV, Patny A, Gut J, Rosenthal PJ, Tekwani B, Srivastava A, Avery M. 2006. Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory. *J Med Chem.* 49: 1576-1584.

Dey R, Meneses C, Salotra P, Kamhawi S, Nakhasi HL, Duncan R. 2010. Characterization of a *leishmania* stage-specific mitochondrial membrane protein that enhances the activity of cytochrome c oxidase and its role in virulence. *Mol Microbiol.* 77: 399-414.

Dillon LAL, Okrah K, Hughitt VK, Suresh R, Li Y, Fernandes MC, Belew AT, Corrada Bravo H, Mosser DM, El-Sayed NM. 2015. Transcriptomic profiling of gene expression and RNA processing during *leishmania major* differentiation. *Nucleic Acids Res.*doi:10.1093/nar/gkv656.

Dobson DE, Scholtes LD, Valdez KE, Sullivan DR, Mengeling BJ, Cilmi S, Turco SJ, Beverley SM. 2003. Functional identification of galactosyltransferases (scgs) required for species-specific modifications of the lipophosphoglycan adhesin controlling *leishmania major*-sand fly interactions. *J Biol Chem.* 278: 15523-15531.

Drew ME, Langford CK, Klamo EM, Russell DG, Kavanaugh MP, Landfear SM. 1995. Functional expression of a myo-inositol/H+ symporter from *leishmania donovani*. *Mol Cell Biol.* 15: 5508-5515.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics.* 21: 3439-3440.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, et al. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science.* 309: 404-409.

Emeny RT, Kasten-Jolly J, Mondal T, Lynes MA, Lawrence DA. 2015. Metallothionein differentially affects the host response to *listeria* infection both with and without an additional stress from cold-restraint. *Cell Stress Chaperones.*doi:10.1007/s12192-015-0630-z.

Erdmann M, Scholz A, Melzer IM, Schmetz C, Wiese M. 2006. Interacting protein kinases involved in the regulation of flagellar length. *Mol Biol Cell.* 17: 2035-2045.

Etges R, Müller I. 1998. Progressive disease or protective immunity to *leishmania major* infection: The result of a network of stimulatory and inhibitory interactions. *Journal of Molecular Medicine.* 76: 372-390.

Ettinger NA, Wilson ME. 2008. Macrophage and t-cell gene expression in a model of early infection with the protozoan *leishmania chagasi*. *PLoS Negl Trop Dis.* 2: e252.

Fadda A, Ryten M, Droll D, Rojas F, Färber V, Haanstra JR, Merce C, Bakker BM, Matthews K, Clayton C. 2014. Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol Microbiol.*doi:10.1111/mmi.12764.

Fadul CE, Channon JY, Kasper LH. 1995. Survival of immunoglobulin g-opsonized *toxoplasma gondii* in nonadherent human monocytes. *Infect Immun.* 63: 4290-4294.

Fleming BD, Chandrasekaran P, Dillon LAL, Dalby E, Suresh R, Sarkar A, El-Sayed NM, Mosser DM. 2015. The generation of macrophages with anti-inflammatory activity in the absence of STAT6 signaling. *J Leukoc Biol.*doi:10.1189/jlb.2A1114-560R.

Flinn HM, Rangarajan D, Smith DF. 1994. Expression of a hydrophilic surface protein in infective stages of *leishmania major*. *Mol Biochem Parasitol.* 65: 259-270.

Folgueira C, Quijada L, Soto M, Abanades DR, Alonso C, Requena JM. 2005. The translational efficiencies of the two *leishmania infantum* HSP70 mRNAs, differing in their 3'-untranslated regions, are affected by shifts in the temperature of growth through different mechanisms. *J Biol Chem.* 280: 35172-35183.

Forget G, Siminovitch KA, Brochu S, Rivest S, Radzioch D, Olivier M. 2001. Role of host phosphotyrosine phosphatase SHP-1 in the development of murine leishmaniasis. *Eur J Immunol.* 31: 3185-3196.

Genske JE, Cairns BR, Stack SP, Landfear SM. 1991. Structure and regulation of histone H2B mRNAs from *leishmania enriettii*. *Mol Cell Biol.* 11: 240-249.

Gerbaulet SP, van Wijnen AJ, Aronin N, Tassinari MS, Lian JB, Stein JL, Stein GS. 1992. Downregulation of histone H4 gene transcription during postnatal development in transgenic mice and at the onset of differentiation in transgenically derived calvarial osteoblast cultures. *J Cell Biochem.* 49: 137-147.

Ghoshal K, Jacob ST. 2001. Regulation of metallothionein gene expression. *Prog Nucleic Acid Res Mol Biol.* 66: 357-384.

Ginger ML, Collingridge PW, Brown RWB, Sproat R, Shaw MK, Gull K. 2013. Calmodulin is required for paraflagellar rod assembly and flagellum-cell body attachment in trypanosomes. *Protist.* 164: 528-540.

Giraud E, Lecoeur H, Soubigou G, Coppée J-Y, Milon G, Prina E, Lang T. 2012. Distinct transcriptional signatures of bone marrow-derived C57BL/6 and DBA/2 dendritic leucocytes hosting live *leishmania amazonensis* amastigotes. *PLoS Negl Trop Dis.* 6: e1980.

Goodstadt L. 2010. Ruffus: A lightweight python library for computational pipelines. *Bioinformatics.* 26: 2778-2779.

Gómez MA, Navas A, Márquez R, Rojas LJ, Vargas DA, Blanco VM, Koren R, Zilberstein D, Saravia NG. 2014. *Leishmania panamensis* infection and antimonial drugs modulate expression of macrophage drug transporters and metabolizing enzymes: Impact on intracellular parasite survival. *J Antimicrob Chemother.* 69: 139-149.

Gregory DJ, Sladek R, Olivier M, Matlashewski G. 2008. Comparison of the effects of *leishmania major* or *leishmania donovani* infection on macrophage gene expression. *Infect Immun.* 76: 1186-1192.

Greif G, Ponce de Leon M, Lamolle G, Rodriguez M, Piñeyro D, Tavares-Marques LM, Reyna-Bello A, Robello C, Alvarez-Valin F. 2013. Transcriptome analysis of the bloodstream stage from the parasite *trypanosoma vivax*. *BMC Genomics.* 14: 149.

Guerfali FZ, Laouini D, Guizani-Tabbane L, Ottones F, Ben-Aissa K, Benkahla A, Manchon L, Piquemal D, Smandi S, Mghirbi O, et al. 2008. Simultaneous gene expression profiling in human macrophages infected with *leishmania major* parasites using SAGE. *BMC Genomics.* 9: 238.

Günzl A. 2010. The pre-mRNA splicing machinery of trypanosomes: Complex or simplified? *Eukaryot Cell.* 9: 1159-1170.

Harder S, Thiel M, Clos J, Bruchhaus I. 2010. Characterization of a subunit of the outer dynein arm docking complex necessary for correct flagellar assembly in *leishmania donovani*. *PLoS Negl Trop Dis.* 4: e586.

Herwaldt BL. 1999. Leishmaniasis. *Lancet.* 354: 1191-1199.

Hoaglin, DC, Mosteller F, Tukey JW. (ed.). 1983. *Understanding Robust and Exploratory Data Analysis.* New York.

Hochegger H, Takeda S, Hunt T. 2008. Cyclin-dependent kinases and cell-cycle transitions: Does one fit all? *Nat Rev Mol Cell Biol.* 9: 910-916.

Holzer TR, McMaster WR, Forney JD. 2006. Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *leishmania mexicana*. *Mol Biochem Parasitol.* 146: 198-218.

Hombach A, Clos J. 2014. No stress--hsp90 and signal transduction in *leishmania*. *Parasitology.* 141: 1156-1166.

Huang J, Van der Ploeg LH. 1991. Requirement of a polypyrimidine tract for *trans*-splicing in trypanosomes: Discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. *EMBO J.* 10: 3877-3885.

Huynh C, Sacks DL, Andrews NW. 2006. A *leishmania amazonensis* ZIP family iron transporter is essential for parasite replication within macrophage phagolysosomes. *J Exp Med.* 203: 2363-2375.

Ilg T. 2002. Generation of myo-inositol-auxotrophic *leishmania mexicana* mutants by targeted replacement of the myo-inositol-1-phosphate synthase gene. *Mol Biochem Parasitol.* 120: 151-156.

Iyer JP, Kaprakkaden A, Choudhary ML, Shaha C. 2008. Crucial role of cytosolic tryparedoxin peroxidase in *leishmania donovani* survival, drug response and virulence. *Mol Microbiol.* 68: 372-391.

Jolliffe IT. 2002. *Principal Component Analysis, Second Edition.*.

Kabara E, Kloss CC, Wilson M, Tempelman RJ, Sreevatsan S, Janagama H, Coussens PM. 2010. A large-scale study of differential gene expression in monocyte-derived macrophages infected with several strains of *mycobacterium avium* subspecies paratuberculosis. *Brief Funct Genomics.* 9: 220-237.

Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. 2011. ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Res.* 39: D712-D717.

Kaye P, Scott P. 2011. Leishmaniasis: Complexity at the host-pathogen interface. *Nat Rev Microbiol.* 9: 604-615.

Kim W, Choi M, Kim J-E. 2014. The histone methyltransferase dot1/DOT1L as a critical regulator of the cell cycle. *Cell Cycle.* 13: 726-738.

Knuepfer E, Stierhof YD, McKean PG, Smith DF. 2001. Characterization of a differentially expressed protein that shows an unusual localization to intracellular membranes in *leishmania major*. *Biochem J.* 356: 335-344.

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. 2010. The transcriptome of the human pathogen *trypanosoma brucei* at single-nucleotide resolution. *PLoS Pathog.* 6: e1001090.

Kramer S. 2012. Developmental regulation of gene expression in the absence of transcriptional control: The case of kinetoplastids. *Mol Biochem Parasitol.* 181: 61-72.

Kwan WC, McMaster WR, Wong N, Reiner NE. 1992. Inhibition of expression of major histocompatibility complex class II molecules in macrophages infected with *leishmania donovani* occurs at the level of gene transcription via a cyclic amp-independent mechanism. *Infect Immun.* 60: 2115-2120.

Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, Holland N, Kuzyk M, Borchers C, Zilberstein D, et al. 2011. Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *leishmania*. *FASEB J.* 25: 515-525.

Lander N, Li Z-H, Niyogi S, Docampo R. 2015. CRISPR/cas9-induced disruption of paraflagellar rod protein 1 and 2 genes in *trypanosoma cruzi* reveals their role in flagellar attachment. *MBio.* 6: e01012.

Laskay T, van Zandbergen G, Solbach W. 2003. Neutrophil granulocytes--trojan horses for *leishmania major* and other intracellular microbes? *Trends Microbiol.* 11: 210-214.

Laufs H, Müller K, Fleischer J, Reiling N, Jahnke N, Jensenius JC, Solbach W, Laskay T. 2002. Intracellular survival of *leishmania major* in neutrophil granulocytes after uptake in the absence of heat-labile serum factors. *Infect Immun.* 70: 826-835.

Law CW, Chen Y, Shi W, Smyth GK. 2014. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15: R29.

LeBowitz JH, Smith HQ, Rusche L, Beverley SM. 1993. Coupling of poly (A) site selection and *trans*-splicing in *leishmania*. *Genes & Development.* 7: 996-1007.

Leifso K, Cohen-Freue G, Dogra N, Murray A, McMaster WR. 2007. Genomic and proteomic expression analysis of *leishmania* promastigote and amastigote life stages: The *leishmania* genome is constitutively expressed. *Mol Biochem Parasitol.* 152: 35-46.

León B, López-Bravo M, Ardavín C. 2007. Monocyte-derived dendritic cells formed at the infection site control the induction of protective T helper 1 responses against *leishmania*. *Immunity.* 26: 519-531.

Locksley RM, Heinzel FP, Fankhauser JE, Nelson CS, Sadick MD. 1988. Cutaneous host defense in leishmaniasis: Interaction of isolated dermal macrophages and epidermal langerhans cells with the insect-stage promastigote. *Infect Immun.* 56: 336-342.

Lynes MA, Borghesi LA, Youn J, Olson EA. 1993. Immunomodulatory activities of extracellular metallothionein. I. Metallothionein effects on antibody production. *Toxicology.* 85: 161-177.

Magee DA, Taraktsoglou M, Killick KE, Nalpas NC, Browne JA, Park SDE, Conlon KM, Lynn DJ, Hokamp K, Gordon SV, et al. 2012. Global gene expression and systems biology analysis of bovine monocyte-derived macrophages in response to in vitro challenge with *mycobacterium bovis*. *PLoS One.* 7: e32034.

Mair G, Shi H, Li H, Djikeng A, Aviles HO, Bishop JR, Falcone FH, Gavrilescu C, Montgomery JL, Santori MI, et al. 2000. A new twist in trypanosome RNA metabolism: Cis-splicing of pre-mRNA. *RNA.* 6: 163-169.

Manful T, Fadda A, Clayton C. 2011. The role of the 5'-3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA.* 17: 2039-2047.

Maretti-Mira AC, Bittner J, Oliveira-Neto MP, Liu M, Kang D, Li H, Pirmez C, Craft N. 2012. Transcriptome patterns from primary cutaneous *leishmania braziliensis* infections associate with eventual development of mucosal disease in humans. *PLoS Negl Trop Dis.* 6: e1816.

Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. 2003. Transcription of *leishmania major* friedlin chromosome 1 initiates in both directions within a single region. *Molecular Cell.* 11: 1291-1299.

Matheoud D, Moradin N, Bellemare-Pelletier A, Shio MT, Hong WJ, Olivier M, Gagnon E, Desjardins M, Descoteaux A. 2013. *Leishmania* evades host immunity by inhibiting antigen cross-presentation through direct cleavage of the SNARE VAMP8. *Cell Host Microbe.* 14: 15-25.

Matthews KR, Tschudi C, Ullu E. 1994. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* 8: 491-501.

Mavromatis CH, Bokil NJ, Totsika M, Kakkanat A, Schaale K, Cannistraci CV, Ryu T, Beatson SA, Ulett GC, Schembri MA, et al. 2015. The co-transcriptome of uropathogenic *escherichia coli*-infected mouse macrophages reveals new insights into host-pathogen interactions. *Cell Microbiol.* 17: 730-746.

McCall L-I, Matlashewski G. 2012. Involvement of the *leishmania donovani* virulence factor A2 in protection against heat and oxidative stress. *Exp Parasitol.* 132: 109-115.

McDonagh PD, Myler PJ, Stuart K. 2000. The unusual gene organization of *leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res.* 28: 2800-2803.

McNicoll F, Drummelsmith J, Müller M, Madore E, Boilard N, Ouellette M, Papadopoulou B. 2006. A combined proteomic and transcriptomic approach to the study of stage differentiation in *leishmania infantum*. *Proteomics.* 6: 3567-3581.

Michaeli S. 2011. *Trans*-splicing in trypanosomes: Machinery and its impact on the parasite transcriptome. *Future Microbiol.* 6: 459-474.

Moll H, Flohé S, Röllinghoff M. 1995. Dendritic cells in *leishmania major*-immune mice harbor persistent parasites and mediate an antigen-specific T cell immune response. *Eur J Immunol.* 25: 693-699.

Mosser DM, Edelson PJ. 1985. The mouse macrophage receptor for c3bi (CR3) is a major mechanism in the phagocytosis of *leishmania* promastigotes. *J Immunol.* 135: 2785-2789.

Mosser DM, Edelson PJ. 1987. The third component of complement (C3) is responsible for the intracellular survival of *leishmania major*. *Nature.* 327: 329-331.

Mottram JC, Coombs GH, Alexander J. 2004. Cysteine peptidases as virulence factors of *leishmania*. *Curr Opin Microbiol.* 7: 375-381.

Murray HW, Berman JD, Davies CR, Saravia NG. 2005. Advances in leishmaniasis. *Lancet.* 366: 1561-1577.

Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, Gordon S, Hamilton JA, Ivashkiv LB, Lawrence T, et al. 2014. Macrophage activation and polarization: Nomenclature and experimental guidelines. *Immunity.* 41: 14-20.

Myler PJ, Audleman L, deVos T, Hixson G, Kiser P, Lemley C, Magness C, Rickel E, Sisk E, Sunkin S, et al. 1999. *Leishmania major* friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A.* 96: 2902-2906.

Nalpas NC, Park SDE, Magee DA, Taraktsoglou M, Browne JA, Conlon KM, Rue-Albrecht K, Killick KE, Hokamp K, Lohan AJ, et al. 2013. Whole-transcriptome, high-throughput RNA sequence analysis of the bovine macrophage response to *mycobacterium bovis* infection in vitro. *BMC Genomics.* 14: 230.

Nandan D, Lo R, Reiner NE. 1999. Activation of phosphotyrosine phosphatase activity attenuates mitogen-activated protein kinase signaling and inhibits c-fos and nitric oxide synthase expression in macrophages infected with *leishmania donovani*. *Infect Immun.* 67: 4055-4063.

Nandan D, Reiner NE. 1995. Attenuation of gamma interferon-induced tyrosine phosphorylation in mononuclear phagocytes infected with *leishmania donovani*: Selective inhibition of signaling through janus kinases and stat1. *Infect Immun.* 63: 4495-4500.

Ng LG, Hsu A, Mandell MA, Roediger B, Hoeller C, Mrass P, Iparraguirre A, Cavanagh LL, Triccas JA, Beverley SM, et al. 2008. Migratory dermal dendritic cells act as rapid sensors of protozoan parasites. *PLoS Pathog.* 4: e1000222.

Nourbakhsh F, Uliana SR, Smith DF. 1996. Characterisation and expression of a stage-regulated gene of *leishmania major*. *Mol Biochem Parasitol.* 76: 201-213.

Novais FO, Carvalho LP, Passos S, Roos DS, Carvalho EM, Scott P, Beiting DP. 2015. Genomic profiling of human *leishmania braziliensis* lesions identifies transcriptional modules associated with cutaneous immunopathology. *J Invest Dermatol.* 135: 94-101.

Olivier M, Atayde VD, Isnard A, Hassani K, Shio MT. 2012. *Leishmania* virulence factors: Focus on the metalloprotease GP63. *Microbes Infect.* 14: 1377-1389.

Olivier M, Brownsey RW, Reiner NE. 1992. Defective stimulus-response coupling in human monocytes infected with *leishmania donovani* is associated with altered activation and translocation of protein kinase C. *Proc Natl Acad Sci U S A.* 89: 7481-7485.

Olivier M, Gregory DJ, Forget G. 2005. Subversion mechanisms by which *leishmania* parasites can escape the host immune response: A signaling point of view. *Clin Microbiol Rev.* 18: 293-305.

Osorio y Fortéa J, de La Llave E, Regnault B, Coppée J-Y, Milon G, Lang T, Prina E. 2009. Transcriptional signatures of BALB/c mouse macrophages housing multiplying *leishmania amazonensis* amastigotes. *BMC Genomics.* 10: 119.

Ovalle-Bracho C, Franco-Muñoz C, Londoño-Barbosa D, Restrepo-Montoya D, Clavijo-Ramírez C. 2015. Changes in macrophage gene expression associated with *leishmania (viannia) braziliensis* infection. *PLoS One.* 10: e0128934.

Pal S, Dolai S, Yadav RK, Adak S. 2010. Ascorbate peroxidase from *leishmania major* controls the virulence of infective stage of promastigotes by regulating oxidative stress. *PLoS One.* 5: e11271.

Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M, et al. 2007. Comparative genomic analysis of three *leishmania* species that cause diverse human disease. *Nat Genet.* 39: 839-847.

Peters NC, Egen JG, Secundino N, Debrabant A, Kimblin N, Kamhawi S, Lawyer P, Fay MP, Germain RN, Sacks D. 2008. In vivo imaging reveals an essential role for neutrophils in leishmaniasis transmitted by sand flies. *Science.* 321: 970-974.

Price CTD, Abu Kwaik Y. 2014. The transcriptome of *legionella pneumophila*-infected human monocyte-derived macrophages. *PLoS One.* 9: e114914.

Probst CM, Silva RA, Menezes JPB, Almeida TF, Gomes IN, Dallabona AC, Ozaki LS, Buck GA, Pavoni DP, Krieger MA, et al. 2012. A comparison of two distinct murine macrophage gene expression profiles in response to *leishmania amazonensis* infection. *BMC Microbiol.* 12: 22.

Puri V, Goyal A, Sankaranarayanan R, Enright AJ, Vaidya T. 2011. Evolutionary and functional insights into *leishmania* META1: Evidence for lateral gene transfer and a role for META1 in secretion. *BMC Evol Biol.* 11: 334.

Quijada L, Soto M, Alonso C, Requena JM. 2000. Identification of a putative regulatory element in the 3'-untranslated region that controls expression of HSP70 in *leishmania infantum*. *Mol Biochem Parasitol.* 110: 79-91.

Quijada L, Soto M, Requena JM. 2005. Genomic DNA macroarrays as a tool for analysis of gene expression in *leishmania*. *Exp Parasitol.* 111: 64-70.

Rabhi I, Rabhi S, Ben-Othman R, Rasche A, Daskalaki A, Trentin B, Piquemal D, Regnault B, Descoteaux A, Guizani-Tabbane L, et al. 2012. Transcriptomic signature of *leishmania* infected mice macrophages: A metabolic point of view. *PLoS Negl Trop Dis.* 6: e1763.

Ramakrishnan S, Serricchio M, Striepen B, Bütikofer P. 2013. Lipid synthesis in protozoan parasites: A comparison between kinetoplastids and apicomplexans. *Prog Lipid Res.* 52: 488-512.

Ramírez C, Díaz-Toro Y, Tellez J, Castilho TM, Rojas R, Ettinger NA, Tikhonova I, Alexander ND, Valderrama L, Hager J, et al. 2012. Human macrophage response to *L. (Viannia) panamensis*: Microarray evidence for an early inflammatory response. *PLoS Negl Trop Dis.* 6: e1866.

Rastrojo A, Carrasco-Ramiro F, Martín D, Crespillo A, Reguera RM, Aguado B, Requena JM. 2013. The transcriptome of *leishmania major* in the axenic promastigote stage: Transcript annotation and relative expression levels by RNA-seq. *BMC Genomics.* 14: 223.

Real F, Florentino PTV, Reis LC, Ramos-Sanchez EM, Veras PST, Goto H, Mortara RA. 2014. Cell-to-cell transfer of *leishmania amazonensis* amastigotes is mediated by immunomodulatory lamp-rich parasitophorous extrusions. *Cell Microbiol.* 16: 1549-1564.

Real F, Vidal RO, Carazzolle MF, Mondego JMC, Costa GGL, Herai RH, Würtele M, de Carvalho LM, Carmona e Ferreira R, Mortara RA, et al. 2013. The genome sequence of *leishmania (leishmania) amazonensis*: Functional annotation and extended analysis of gene models. *DNA Res.* 20: 567-581.

Reiner NE. 1994. Altered cell signaling and mononuclear phagocyte deactivation during intracellular infection. *Immunol Today.* 15: 374-381.

Reiner SL, Locksley RM. 1995. The regulation of immunity to *leishmania major*. *Annu Rev Immunol.* 13: 151-177.

Requena JM, Montalvo AM, Fraga J. 2015. Molecular chaperones of *leishmania*: Central players in many stress-related and -unrelated physiological processes. *Biomed Res Int.* 2015: 301326.

Requena JM, Quijada L, Soto M, Alonso C. 2003. Conserved nucleotides surrounding the *trans*-splicing acceptor site and the translation initiation codon in *leishmania* genes. *Exp Parasitol.* 103: 78-81.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol.* 29: 24-26.

Rochette A, Raymond F, Corbeil J, Ouellette M, Papadopoulou B. 2009. Whole-genome comparative RNA expression profiling of axenic and intracellular amastigote forms of *leishmania infantum*. *Mol Biochem Parasitol.* 165: 32-47.

Rochette A, Raymond F, Ubeda J-M, Smith M, Messier N, Boisvert S, Rigault P, Corbeil J, Ouellette M, Papadopoulou B. 2008. Genome-wide gene expression profiling analysis of *leishmania major* and *leishmania infantum* developmental stages reveals substantial differences between the two species. *BMC Genomics.* 9: 255.

Rodriguez NE, Chang HK, Wilson ME. 2004. Novel program of macrophage gene expression induced by phagocytosis of *leishmania chagasi*. *Infect Immun.* 72: 2111-2122.

Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, et al. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of *leishmania*. *Genome Res.* 21: 2129-2142.

Sacks D, Sher A. 2002. Evasion of innate immunity by parasitic protozoa. *Nature Immunology.* 3: 1041-1047.

Sacks DL. 1989. Metacyclogenesis in *leishmania* promastigotes. *Exp Parasitol.* 69: 100-103.

Sacks DL, Perkins PV. 1984. Identification of an infective stage of *leishmania* promastigotes. *Science.* 223: 1417-1419.

Sanchez MA, Zeoli D, Klamo EM, Kavanaugh MP, Landfear SM. 1995. A family of putative receptor-adenylate cyclases from *leishmania donovani*. *J Biol Chem.* 270: 17551-17558.

Sarkar A, Aga E, Bussmeyer U, Bhattacharyya A, Möller S, Hellberg L, Behnen M, Solbach W, Laskay T. 2013. Infection of neutrophil granulocytes with *leishmania major* activates ERK 1/2 and modulates multiple apoptotic pathways to inhibit apoptosis. *Med Microbiol Immunol.* 202: 25-35.

Saxena A, Lahav T, Holland N, Aggarwal G, Anupama A, Huang Y, Volpin H, Myler PJ, Zilberstein D. 2007. Analysis of the *leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Mol Biochem Parasitol.* 152: 53-65.

Saxena A, Worthey EA, Yan S, Leland A, Stuart KD, Myler PJ. 2003. Evaluation of differential gene expression in *leishmania major* friedlin procyclics and metacyclics using DNA microarray analysis. *Mol Biochem Parasitol.* 129: 103-114.

Sádlová J, Price HP, Smith BA, Votýpka J, Volf P, Smith DF. 2010. The stage-regulated HASPB and SHERP proteins are essential for differentiation of the protozoan parasite *leishmania major* in its sand fly vector, *phlebotomus papatasi*. *Cell Microbiol.* 12: 1765-1779.

Scharton-Kersten T, Scott P. 1995. The role of the innate immune response in th1 cell development following *leishmania major* infection. *J Leukoc Biol.* 57: 515-522.

Siegel TN, Hekstra DR, Wang X, Dewell S, Cross GAM. 2010. Genome-wide analysis of mRNA abundance in two life-cycle stages of *trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* 38: 4946-4957.

Siegel TN, Tan KSW, Cross GAM. 2005. Systematic study of sequence motifs for RNA trans splicing in *trypanosoma brucei*. *Mol Cell Biol.* 25: 9586-9594.

Smith CW, Chu TT, Nadal-Ginard B. 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol Cell Biol.* 13: 4939-4952.

Smith CW, Porro EB, Patton JG, Nadal-Ginard B. 1989. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature.* 342: 243-247.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 3: Article3.

Soto M, Iborra S, Quijada L, Folgueira C, Alonso C, Requena JM. 2004. Cell-cycle-dependent translation of histone mRNAs is the key control point for regulation of histone biosynthesis in *leishmania infantum*. *Biochem J.* 379: 617-625.

de Sousa LR, Wu H, Nebo L, Fernandes JB, da Silva MF, Kiefer W, Schirmeister T, Vieira PC. 2015. Natural products as inhibitors of recombinant cathepsin L of *leishmania mexicana*. *Exp Parasitol.* 156: 42-48.

Späth GF, Beverley SM. 2001. A lipophosphoglycan-independent method for isolation of infective *leishmania* metacyclic promastigotes by density gradient centrifugation. *Exp Parasitol.* 99: 97-103.

Stein JL, van Wijnen AJ, Lian JB, Stein GS. 1996. Control of cell cycle regulated histone genes during proliferation and differentiation. *Int J Obes Relat Metab Disord.* 20 Suppl 3: S84-S90.

Subbian S, Tsenova L, Kim M-J, Wainwright HC, Visser A, Bandyopadhyay N, Bader JS, Karakousis PC, Murrmann GB, Bekker L-G, et al. 2015. Lesion-Specific immune response in granulomas of patients with pulmonary tuberculosis: A pilot study. *PLoS One.* 10: e0132249.

Sutton RE, Boothroyd JC. 1986. Evidence for *trans* splicing in trypanosomes. *Cell.* 47: 527-535.

Südhof TC, Rothman JE. 2009. Membrane fusion: Grappling with SNARE and SM proteins. *Science.* 323: 474-477.

Tannahill GM, Curtis AM, Adamik J, Palsson-McDermott EM, McGettrick AF, Goel G, Frezza C, Bernard NJ, Kelly B, Foley NH, et al. 2013. Succinate is an inflammatory signal that induces IL-1β through HIF-1α. *Nature.* 496: 238-242.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics.* 25: 1105-1111.

Tucker WC, Chapman ER. 2002. Role of synaptotagmin in Ca2+-triggered exocytosis. *Biochem J.* 366: 1-13.

Ullu E, Matthews KR, Tschudi C. 1993. Temporal order of RNA-processing reactions in trypanosomes: Rapid *trans* splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol.* 13: 720-725.

Vicente JJ, Wordeman L. 2015. Mitosis, microtubule dynamics and the evolution of kinesins. *Exp Cell Res.* 334: 61-69.

Vinet AF, Fukuda M, Descoteaux A. 2008. The exocytosis regulator synaptotagmin V controls phagocytosis in macrophages. *J Immunol.* 181: 5289-5295.

Vinet AF, Fukuda M, Turco SJ, Descoteaux A. 2009. The *leishmania donovani* lipophosphoglycan excludes the vesicular proton-atpase from phagosomes by impairing the recruitment of synaptotagmin V. *PLoS Pathog.* 5: e1000628.

Vinet AF, Jananji S, Turco SJ, Fukuda M, Descoteaux A. 2011. Exclusion of synaptotagmin V at the phagocytic cup by *leishmania donovani* lipophosphoglycan results in decreased promastigote internalization. *Microbiology.* 157: 2619-2628.

Walker J, Vasquez J-J, Gomez MA, Drummelsmith J, Burchmore R, Girard I, Ouellette M. 2006. Identification of developmentally-regulated proteins in *leishmania panamensis* by proteome profiling of promastigotes and axenic amastigotes. *Mol Biochem Parasitol.* 147: 64-73.

Wheeler RJ, Gluenz E, Gull K. 2011. The cell cycle of *leishmania*: Morphogenetic events and their implications for parasite biology. *Mol Microbiol.* 79: 647-662.

Wickham H. 2009. *Ggplot2 : Elegant Graphics for Data Analysis.*.

Wiese M, Kuhn D, Grünfelder CG. 2003. Protein kinase involved in flagellar-length control. *Eukaryot Cell.* 2: 769-777.

Woelbing F, Kostka SL, Moelle K, Belkaid Y, Sunderkoetter C, Verbeek S, Waisman A, Nigg AP, Knop J, Udey MC, et al. 2006. Uptake of *leishmania major* by dendritic cells is mediated by fcgamma receptors and facilitates acquisition of protective immunity. *J Exp Med.* 203: 177-188.

Wu Y, El Fakhry Y, Sereno D, Tamar S, Papadopoulou B. 2000. A new developmentally regulated gene family in *leishmania* amastigotes encoding a homolog of amastin surface proteins. *Mol Biochem Parasitol.* 110: 345-357.

Xue F, Zhao X, Yang Y, Zhao J, Yang Y, Cao Y, Hong C, Liu Y, Sun L, Huang M, et al. 2013. Responses of murine and human macrophages to leptospiral infection: A study using comparative array analysis. *PLoS Negl Trop Dis.* 7: e2477.

Xu W, Xin L, Soong L, Zhang K. 2011. Sphingolipid degradation by *leishmania major* is required for its resistance to acidic ph in the mammalian host. *Infect Immun.* 79: 3377-3387.

Zhang S, Kim CC, Batra S, McKerrow JH, Loke P. 2010. Delineation of diverse macrophage activation programs in response to intracellular parasites and cytokines. *PLoS Negl Trop Dis.* 4: e648.

Zhang WW, Matlashewski G. 2001. Characterization of the a2-a2rel gene cluster in *leishmania donovani*: Involvement of A2 in visceralization during infection. *Mol Microbiol.* 39: 935-948.

Zhang W-W, Mendez S, Ghosh A, Myler P, Ivens A, Clos J, Sacks DL, Matlashewski G. 2003. Comparison of the A2 gene locus in *leishmania donovani* and *leishmania major* and its control over cutaneous infection. *J Biol Chem.* 278: 35508-35515.

Zilka A, Garlapati S, Dahan E, Yaolsky V, Shapira M. 2001. Developmental regulation of heat shock protein 83 in *leishmania*. 3' processing and mRNA stability control transcript abundance, and translation id directed by a determinant in the 3'-untranslated region. *J Biol Chem.* 276: 47922-47929.