



Warren, Felix Sean Lee (2023) *Life and cell cycle progression analysis in Leishmania mexicana by single cell RNA-sequencing*. PhD thesis.

<http://theses.gla.ac.uk/83695/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University
of Glasgow

**Life and cell cycle progression analysis in
Leishmania mexicana by single cell RNA-
sequencing**

Felix Sean Lee Warren

M.Sci. (Hons)

Submitted in fulfilment of the requirement for the Degree of Doctor
of Philosophy

School of Biodiversity, One Health & Veterinary Medicine

College of Medical, Veterinary and Life Sciences

University of Glasgow

December 2022

Abstract

Leishmaniasis is a neglected tropical disease, which is estimated to produce approximately 1.3 million new cases annually. As yet, no vaccines are available and current effective chemotherapeutic strategies are lacking, due to drug toxicity, resistance, and social and economic barriers affecting availability. Pivotal in the discovery of new interventions for infectious diseases is the understanding of the fundamental biology of the pathogen and its role in infection.

Here, the application of single cell RNA sequencing (scRNA-seq) was employed to investigate the fundamental biology behind transcriptomic changes as the parasite progresses through life cycle stages, from promastigote to metacyclic forms before differentiating into amastigotes. The transcriptomic dynamics underpinning these multifarious developmental transitions through life cycle stages are yet to be fully described.

To investigate these changes, we employed scRNA-seq over five time-points as *Leishmania mexicana* (*L. mexicana*) differentiated from promastigotes to axenic-amastigotes *in vitro*. With clustering and marker analysis of over 16,500 parasites across three experiments, revealing thousands of stage specific markers. Of note during gene marker analysis was the discovery of a transitional cluster placed between promastigote and amastigote stages. This cluster, named here Trans A, displayed overlap between both promastigote and amastigote markers, and potentially representing a new intermediate life cycle stage, defined transcriptionally. Additionally in these data was the identification of over 1,500 differentially expressed markers for a metacyclic-like cluster.

To examine the timing and the patterns of gene expression over the life and cell cycle, pseudotime analysis was used for the first time in *Leishmania*, so that we may further explain the order of these transitional events and any potential stage specific patterns in gene expression, providing an unprecedented understanding of the life cycle transitions. Analysis indicated the development between promastigote and amastigote stages was possible by progression through this Trans A cluster, circumventing the infective metacyclic form. Thus, potentially revealing a new developmental strategy for life cycle progression.

Furthermore, cell cycle labelling analysis was performed using phase marker orthologues, also revealing new cell cycle phase specific markers. When combined with promastigote stage marker orthologues, a striking overlap between promastigote life stages and cell cycle stages was found, potentially providing further evidence that promastigote morphologies are intrinsically linked with cell cycle stages.

To validated how transient gene expression changes may be represented at the protein level, 96 fluorescently tagged cell lines were produced using a high-throughput CRISPR-Cas9 system. Of the 96 cell lines produced, 91 were hypothetical proteins identified as having transient expression patterns for developmental trajectories drawn over promastigote to metacyclic and axenic amastigote stages. Initial assessments by fluorescent microscopy revealed 80 out of the 96 tagged proteins matched overall pseudotime expression profiles, indicating that dynamics of RNA levels detected by scRNA-seq could reflect changes in protein levels. Furthermore, these 91 cell lines of tagged hypothetical proteins provide the opportunity for further research into new biology.

So that this transitional gene progression may be explored in broader contexts, scRNA-seq of *L. mexicana*-infected human macrophages was undertaken. Allowing the comparison of the RNA populations found in differentiation to axenic amastigotes *in vitro*, to those of the amastigote forms infecting a host cell. This crucial life-cycle development stage of *Leishmania* within the macrophage revealed infection response associated genes, linked with parasite removal strategies.

This project aimed to provide an example of investigating fundamental *Leishmania* biology by applying scRNA-seq for the study of life and cell cycle transitions, with the application of pseudotime analysis for the first time. Results of differential gene expression analysis revealed new biology not previously described in previous methods. Thereby demonstrating how the application of scRNA-seq may further disseminate parasite biology and infection dynamics.

Table of Contents

Abstract.....	3
List of Tables	8
List of Figures.....	9
Conference Proceedings	13
Acknowledgements	14
Author's Declaration.....	16
Definitions/Abbreviations.....	17
Chapter 1 Introduction	20
1.1 Leishmaniasis	21
1.2 The Biology of the <i>Leishmania</i> parasite	24
1.2.1 Strategies of <i>Leishmania</i> immune system avoidance, survival, and transmission	29
1.3 Transcriptomics in <i>Leishmania</i> biology using bulk RNA-sequencing	34
1.4 The macrophage response to <i>Leishmania</i> infection at the transcriptomic level	37
1.5 Single cell RNA-sequencing	42
1.5.1 10x Chromium platform for single cell RNA-sequencing	45
1.6 Single cell RNA-sequencing in kinetoplastids	48
1.6.1 Trajectory inference	49
1.7 High-throughput gene tagging using CRISPR-Cas9	50
1.8 Thesis summary and aims	51
Chapter 2 Methods & materials	54
2.1 Methods and lab procedures.....	55
2.1.1 <i>Leishmania in vitro</i> culture and differentiation	55
2.1.2 Growth curves	55
2.1.3 Stabilate preparation and retrieval.....	56
2.1.4 Suspended sample single cell RNA-sequencing.....	56
2.1.5 Structural bioinformatics of hypothetical proteins.....	62
2.1.6 Flow cytometry	62
2.1.7 Transforming <i>E. coli</i> for plasmid synthesis.....	63
2.1.8 Plasmid validation by restriction enzyme digest.....	65
2.1.9 Agarose gel electrophoresis	65
2.1.10 High-throughput CRISPR-Cas9 tagging	65
2.1.11 Fluorescence microscopy	70
2.1.12 THP-1 culture and differentiation.....	71
2.1.13 CFSE infection assays.....	71
2.1.14 Adherent sample single cell RNA-sequencing	72

Chapter 3	Life and cell cycle progression analysis by single cell RNA-sequencing in <i>Leishmania mexicana</i>	75
3.1	Introduction	76
3.1.1	Life and cell cycle progression analysis in <i>Leishmania</i> using single cell RNA-sequencing	76
3.2	Aims	77
3.3	Results	78
3.3.1	Single cell RNA-sequencing sample culture and preparation	78
3.3.2	Single cell RNA-sequencing analysis for first <i>L. mexicana</i> replicate 79	
3.3.3	Single cell RNA-sequencing analysis for second promastigote only replicate.....	98
3.3.4	Single cell RNA-sequencing analysis for second, axenic amastigote only, replicate.....	118
3.3.5	Single cell RNA-sequencing analysis of integrated <i>L. mexicana</i> replicates	127
3.4	Discussion.....	159
3.5	Summary	165
Chapter 4	Validation of transiently expressed hypothetical proteins by CRISPR-Cas9 mediated tagging	166
4.1	Introduction	167
4.1.1	CRISPR tagging.....	167
4.2	Aims	169
4.3	Results	169
4.3.1	Generation of 96 tagged cell lines by CRISPR-Cas9.....	169
4.3.2	Comparing pseudotime gene expression with tagged fluorescent expression of proteins.....	172
4.4	Discussion.....	177
4.4.1	mRNA and protein expression in <i>Leishmania</i>	177
4.5	Summary	181
Chapter 5	Single cell RNA-sequencing analysis of <i>Leishmania</i> infecting human macrophages	182
5.1	Introduction	183
5.2	Aims	186
5.3	Results	187
5.3.1	Single cell RNA-sequencing sample culture and preparation for macrophage infections.....	187
5.3.2	Single cell RNA-sequencing analysis of infected macrophages ...	189
5.3.3	Single cell RNA-sequencing analysis for <i>L. mexicana</i> within infected macrophage sample	205
5.3.4	Integrated samples from promastigote, axenic amastigote and amastigotes from macrophage infections.....	208

5.4	Discussion.....	218
5.4.1	Markers for <i>L. mexicana</i> infection process revealed by scRNA-seq 221	
5.4.2	Novel markers for <i>L. mexicana</i> life cycle stages and prospective new life cycle stage defined by transcriptomic profile.....	222
5.5	Summary.....	224
Chapter 6	Concluding remarks.....	225
6.1	Spatial transcriptomics.....	229
6.2	Cell atlases.....	230
6.3	Summary.....	231
Appendices	232
Appendix I	232
6.4	Rep2_Pro quality control plots.....	232
6.5	Rep2_Axa quality control plots.....	237
6.6	Human macrophage Cell Ranger plots, pre-quality control.....	242
Appendix II	245
Appendix III	252
Appendix IV	257
Appendix V	260
List of References	263

List of Tables

Table 1-1 Species, vectors, and diseases of Leishmaniasis.	22
Table 1-2 Technologies available for single-cell RNA-Sequencing experimental protocols.....	43
Table 5-1 Infected macrophage markers with gene ontologies of interest for <i>L. mexicana</i> infections.....	203

List of Figures

Figure 1-1 Distribution of Leishmaniases.	23
Figure 1-2 Life cycle of the <i>Leishmania</i> parasite	25
Figure 1-3 Life cycle of the <i>Leishmania</i> parasite	45
Figure 1-4 The 10x Genomics Chromium workflow.	47
Figure 2-1 Plasmid map for pPlot-mNG-blast-blast.	64
Figure 2-2 GOO scaffold sgRNA backbone sequence.	66
Figure 3-1 Experimental sample collection and sample information for integrated single cell RNA-Sequencing.	80
Figure 3-2 Cell Ranger summaries of sequenced Rep1 sample.	81
Figure 3-3 Quality control and filtering of transcriptomes in first biological replicate.	84
Figure 3-4 Scatter plot of top 2,000 variable features in Rep1 dataset	86
Figure 3-5 Elbow plot of variance found in principle components observed in Rep1 sample.	87
Figure 3-6 Clustree plot demonstrating Rep1 sample movement with increasing cluster resolution.	88
Figure 3-7 Clustering of <i>L. mexicana</i> transcriptomes across life cycle development from promastigote to axenic amastigote.	90
Figure 3-8 UMAP life cycle marker plots in Rep1 scRNA-seq sample to determine stages defined in clustering analysis.	92
Figure 3-9 Violin plots of selected marker genes in Rep1 scRNA-seq sample.	93
Figure 3-10 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep1 sample.	95
Figure 3-11 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep1.	97
Figure 3-12 Experimental sample collection and sample information for second replicate of promastigote timepoints.	99
Figure 3-13 Clustree plot demonstrating Rep2_Pro sample movement with increasing cluster resolution.	100
Figure 3-14 Clustering of <i>L. mexicana</i> transcriptomes across life cycle development from log phase promastigotes to stationary phase promastigotes.	101
Figure 3-15 UMAP life cycle marker plots in Rep2_Pro scRNA-seq sample to determine stages defined in clustering analysis.	102
Figure 3-16 Violin plots of selected marker genes in Rep2_Pro scRNA-seq sample.	103
Figure 3-17 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep2_Pro.	104
Figure 3-18 Promastigote orthologous labelling from procyclic promastigotes to metacyclic promastigote forms.	106
Figure 3-19 Proportions of labelled promastigote life cycle stages in respectively labelled cell cycle stages in the Rep2_Pro scRNA-seq sample.	107
Figure 3-20 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep2_Pro sample.	109
Figure 3-21 Resolving markers differentially expressed in promastigote life cycle stage clusters in Rep2_Pro.	113
Figure 3-22 Expression of promastigote marker genes across pseudotime trajectories.	116

Figure 3-23 Experimental sample collection and sample information for second replicate of axenic amastigotes timepoints.	118
Figure 3-24 Clustree plot demonstrating Rep2_Axa sample movement with increasing cluster resolution.	119
Figure 3-25 Clustering of <i>L. mexicana</i> transcriptomes across life cycle development in newly developed axenic amastigotes.	120
Figure 3-26 UMAP life cycle marker plots in Rep2_Axa scRNA-seq sample to determine stages defined in clustering analysis.	121
Figure 3-27 Violin plots of selected marker genes in Rep2_Axa scRNA-seq sample.	122
Figure 3-28 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep2_Axa.	123
Figure 3-29 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep2_Axa sample.	125
Figure 3-30 Quality control and filtering of integrated transcriptomes for unique molecular identifiers, feature counts, ribosomal RNA and kinetoplast DNA. ...	127
Figure 3-31 Elbow plot of variance found in principle components observed in integrated Rep1, Rep2_Pro and Rep2_Axa samples.	128
Figure 3-32 Clustree plot demonstrating cell movement in integrated samples with increasing cluster resolution.	129
Figure 3-33 Clustering of integrated <i>L. mexicana</i> transcriptomes across life cycle stage development from promastigote to axenic amastigote in two biological replicates.	130
Figure 3-34 UMAP life cycle marker plots in integrated scRNA-seq samples to determine stages defined in clustering analysis.	132
Figure 3-35 Violin plots of selected marker genes in integrated scRNA-seq samples.	133
Figure 3-36 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in integrated samples.	134
Figure 3-37 GO Term analysis by REVIGO of metacyclic and transitional A cluster.	138
Figure 3-38 Venn diagrams of overlapping markers for life cycle stage clusters between transitional A, promastigote and axenic amastigote clusters.	139
Figure 3-39 Venn diagrams of overlapping markers for life cycle stage clusters between combined scRNA-seq promastigote and axenic amastigote clusters, and bulk RNA-Seq promastigote and axenic amastigote clusters.	141
Figure 3-40 Promastigote labels for transitional A cluster reveals a mixture of promastigote stages with no distinct promastigote life cycle stage across the cluster.	142
Figure 3-41 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in integrated samples.	143
Figure 3-42 Detection of markers differentially expressed in replicating promastigote and axenic amastigote clusters.	145
Figure 3-43 Promastigote labels for promastigote replicating cluster reveals only procyclic and leptomonad promastigote stages.	147
Figure 3-44 Detection of markers differentially expressed in replicating promastigote cluster.	148
Figure 3-45 Detection of markers differentially expressed in replicating axenic amastigote cluster.	150

Figure 3-46 Pseudotime analysis demonstrates two distinct trajectories from Pro 1 cluster ending in Meta and Axa Rep clusters.	152
Figure 3-47 Ordering and grouping of genes differentially expressed through <i>L. mexicana</i> life cycle progression.	154
Figure 3-48 Expression of life cycle marker genes across pseudotime trajectories.	156
Figure 3-49 Top differentiated gene expression profiles across pseudotime, from start of trajectories to branching of trajectories.	158
Figure 4-1 pPLOT-mNeonGreen blast-blast plasmid restriction enzyme digest.	170
Figure 4-2 PCR products for DNA amplification of 3' tagging cassette for pPLOT mNeonGreen plasmid donor fragments for tags of hypothetical and life cycle stage marker proteins.	171
Figure 4-3 PCR products for DNA amplification of downstream guide RNAs for tags of hypothetical and life cycle stage marker proteins.....	172
Figure 4-4 Expression of selected life cycle stage markers across pseudotime and fluorescent tag expression in promastigote and amastigote timepoints.	173
Figure 4-5 Expression of the hypothetical protein LmxM.08.0810 across pseudotime and fluorescent tag expression in promastigote and amastigote timepoints.	175
Figure 4-6 Median fluorescence expression from promastigote and axenic amastigote timepoints for all 96 tagged proteins separated by those matching pseudotime expression at respective life cycle stages and non-matching.	176
Figure 5-1 Experimental sample collection and sample information single cell RNA-sequencing experiment using macrophages infected with <i>L. mexicana</i> . ..	188
Figure 5-2 Quality control and filtering of transcriptomes in macrophage samples.	190
Figure 5-3 Scatter plot of top 3,000 variable features in macrophage datasets.	191
Figure 5-4 Elbow plot of variance found in principle components observed in integrated macrophage samples.	192
Figure 5-5 Clustree plot demonstrating integrated macrophage sample movement with increasing cluster resolution.	193
Figure 5-6 Clustering of human macrophage and <i>L. mexicana</i> transcriptomes across two timepoints post infection.	194
Figure 5-7 Clustering and top differentiated marker analysis by heatmap with dot plot of <i>L. mexicana</i> transcripts in integrated macrophage samples.	195
Figure 5-8 UMAP feature plots and violin plots in integrated macrophage scRNA-seq samples for top 4 human and <i>L. mexicana</i> markers.	196
Figure 5-9 Elbow plot of variance found in principle components observed in uninfected control macrophage sample and Clustree plot demonstrating movement with increasing cluster resolution.	198
Figure 5-10 Clustering and top differentiated marker analysis by heatmap in uninfected control macrophage sample.	199
Figure 5-11 Elbow plot of variance found in principle components observed in infected macrophage sample and Clustree plot demonstrating movement with increasing cluster resolution.	200
Figure 5-12 Clustering and top differentiated marker analysis by heatmap in infected macrophage sample with <i>L. mexicana</i>	201
Figure 5-13 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage sample.....	202
Figure 5-14 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage samples by scRNA-seq.	204

Figure 5-15 Quality control and filtering of transcriptomes in amastigote sample isolated from infected macrophages and top 3,000 variable markers.	205
Figure 5-16 Elbow plot of variance found in principle components observed in <i>L. mexicana</i> amastigote sample and Clustree plot demonstrating movement with increasing cluster resolution.	206
Figure 5-17 Clustering and top differentiated marker analysis by heatmap in <i>L. mexicana</i> amastigote sample.	207
Figure 5-18 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage sample.	208
Figure 5-19 Elbow plot of variance found in principle components observed in all integrated <i>L. mexicana</i> samples and Clustree plot demonstrating movement with increasing cluster resolution.	209
Figure 5-20 Clustering of all integrated <i>L. mexicana</i> samples.	210
Figure 5-21 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in all integrated <i>L. mexicana</i> samples.	212
Figure 5-22 Gene ontology term analysis by REVIGO of Low cluster.	214
Figure 5-23 Gene ontology term analysis by REVIGO of Meta cluster.	216
Figure 5-24 Gene ontology term analysis by REVIGO of Trans A cluster.	217
Figure 0-1 Cell Ranger summaries of sequenced Rep2_Pro sample.	232
Figure 0-2 Quality control and filtering of transcriptomes in second biological replicate containing only promastigote stages.	234
Figure 0-3 Scatter plot of top 2,000 variable features in Rep2_Pro dataset.	236
Figure 0-4 Elbow plot of variance found in principle components observed in Rep2_Pro sample.	237
Figure 0-5 Cell Ranger summaries of sequenced Rep2_Axa sample.	238
Figure 0-6 Quality control and filtering of transcriptomes in second axenic amastigote replicate.	239
Figure 0-7 Scatter plot of top 2,000 variable features in Rep2_Axa dataset ...	241
Figure 0-8 Elbow plot of variance found in principle components observed in Rep2_Axa sample.	242
Figure 0-9 Cell Ranger summaries of sequenced uninfected macrophage control sample.	243
Figure 0-10 Cell Ranger summaries of sequenced infected macrophage sample.	244

Conference Proceedings

Parts of this thesis have been presented at the following national and international conferences and meetings:

- ◆ Kinetoplastid Molecular Cell Biology Meeting, Woods Hole, USA, 2019, [Poster]
- ◆ Neglected Tropical Diseases Network CRISPR-Cas9 Genetic Manipulation Workshop, Kolkata, India, 2019 [Oral Presentation and Poster]
- ◆ Scottish Universities Life Sciences Alliance's Disruptive Technologies in the Life Sciences Conference, Edinburgh, UK, 2019 [Poster]
- ◆ British Society for Parasitology Spring Meeting, Granada, Spain, 2020 [Oral Presentation]
- ◆ Kinetoplastid Single Cell RNA-Sequencing Virtual Seminar, Online, 2020 [Online Oral Presentation]
- ◆ Molecular Parasitology Meeting XXXII, Woods Hole, USA, 2021 [Online Poster]
- ◆ British Society for Parasitology Spring Meeting, York, UK, 2022 [Oral Presentation]
- ◆ World Federation of Parasitologists ICOPA 2022, Copenhagen, Norway [Oral Presentation]

Acknowledgements

This thesis represents the culmination of a decade of study in higher education, something I always thought unobtainable in my youth. I've had a lot of help along the way, and to list every name would form a chapter in and of itself. If I have missed recognising any contributions by name, please know I am eternally thankful.

In chronological order, I would have to first thank Prof. John Kusel, who taught the evening access classes in biology while I worked full-time as a cheesemonger, with a dream of going to university. I'm forever grateful for his inspiration and thankful for his continued communications. Huge gratitude also to the summer lab placements I was given through the Dean's Scholar's List Scheme, which formed the foundation of my love of the laboratory; Prof. Richard Cogdell, who was always so generous with his time and recommendations, and Dr. Andrew Hamilton, who showed me how important curiosity is for a scientist.

Sincerest gratitude also for my time during my work placement MSci at the Paul O'Gorman Leukaemia Research Centre. Prof Tessa Holyoake and Dr. Heather Jorgensen will both always remain an inspiration to me, and my time there will serve as a reminder that perseverance will reap rewards. Special thanks also to Prof. Matthew Dalby and Dr. Virginia Llopis-Hernandez for seeing me through my Honours project with such enthusiasm and encouragement.

For my time during this Ph.D. in Precision Medicine I'd also offer kind gratitude and recognition to the Medical Research Council, and Alexis Merry for all their support, without which this project could not take place. For their professional mentorship I thank my supervisors Prof. Martin Llewellyn, Prof. Richard McCulloch for their support, insights, and guidance through this project, making this a truly exceptional experience. Special thank you to Dr. Emma Briggs also, for all her help and patience with my often-baffling code questions and kind thesis corrections. Without their help this thesis would certainly look very different. Thank you to the Gluenz group for all their support during the CRISPR work described here. Thank you to Dr. Pamela Scott for the opportunity to demonstrate in undergraduate labs, which has often been a highlight of my year.

A huge thank you to Dr. Sara Restuccia for her kind guidance as my thesis mentor, offering stories and encouragement when I thought this was all beyond me. And lastly, to all the members, past and present of Office B617 and those kind folks at the Wellcome Centre for Integrative Parasitology. To quote a hobbit on their 111th birthday, “I don't know half of you half as well as I should like; and I like less than half of you half as well as you deserve”.

This thesis is dedicated to my cat Calcifer, who has been a constant over the past 17 and a half years. Even if he is a grumpy old chap these days. Sincerest appreciation to my friends, especially Kalle Hartwig and Kenny Wilson for all their support, it means a lot, much love. A special thank you to my partner Saveena Ginda, who gave me the best encouragement I could have hoped for during the most stressful of times. This thesis is also dedicated to the cast and crew of Critical Role. Thank you for the escapism on those long walks back from the 16-hour days in the laboratory.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name: Felix Sean Lee Warren

Definitions/Abbreviations

°C Degrees Celsius

µg Microgram

µL Microlitre

µm Micrometre

µM Micromolar

µs Microsecond

AMA - Infected *Leishmania mexicana* scRNA-seq sample reads

BSA - Bovine serum albumin

CFSE - carboxyfluorescein succinimidyl ester

cHOM Complete HOMEM - HOMEM medium (GE Healthcare) + 10% HiFBS

CL Cutaneous leishmaniasis

CO₂ Carbon dioxide

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

cSDM Complete Schneider's Drosophila Medium (Gibco) + 20% HiFBS + mM Haemin, pH 5.5

CTL - Uninfected Human macrophage control scRNA-seq sample reads

DE - Differential expression

DMSO - Dimethyl sulfoxide

DPBS - Dulbecco's phosphate-buffered saline

DNA - Deoxyribonucleic acid

E. coli - *Escherichia coli*

FBS - Foetal Bovine Serum

FC - Fold Change

g gram

GO Gene Ontology

INF - Infected Human macrophage scRNA-seq sample

HiFBS - Heat Inactivated Foetal Bovine Serum (Gibco)

HOMEM - Haemoflagellate minimal essential medium

H. sapiens - *Homo sapiens*

kDNA - Kinetoplast DNA

L Litre

L. braziliensis - *Leishmania braziliensis*

L. mexicana - *Leishmania mexicana*

LB-medium - Lysogeny Broth medium

Log Logarithm

M Molar

mg Milligram

mL Millilitre

mM Millimolar

mm Millimetre

mRNA Messenger RNA

ms - Microsecond

nm - Nanometre

NTC - Nourseothricin Sulfate

PBS Phosphate-buffered saline

PC - Principle Component

PCR - polymerase chain reaction

PSG - PBS supplemented with 1% D-glucose

PI - Propidium iodide

Rep1 - First scRNA-seq replicate

Rep2_Pro - Second replicate of promastigote stages

Rep2_Axa - Second replicate of axenic amastigote stages

scRNA-seq - single-cell RNA-sequencing

T. brucei - *Trypanosoma brucei*

T. congolense - *Trypanosoma congolense*

UTR - Untranslated region

VL Visceral leishmaniasis

Chapter 1 Introduction

1.1 Leishmaniasis

Collectively, the Leishmaniasis are diseases caused by the intracellular protozoan parasite *Leishmania*. The Leishmaniasis are classified as a Neglected Tropical Disease (NTD) by the World Health Organisation (WHO) (2010).

Leishmaniasis affects some of the poorest populations and are no longer limited to subtropical and tropical areas (Black, 2020; Utzinger et al., 2012.).

Disproportionately poor *Leishmania* prognoses affect the most deprived people living in unsuitable housing, often associated with inadequate measures to protect themselves from the insect vector of the parasite, the sandfly (Alvar et al., 2006; WHO., 2010). Passed during the vectors' blood meal, several species of sandfly are responsible for transmission of *Leishmania*. The Old World species is transmitted by the *Phlebotomus* sand fly (Aoun & Bouratbine, 2014), while the New World (Akhoundi et al., 2016) species is typically transmitted by the *Lutzomyia* sand fly (Gramiccia & Gradoni, 2005). *Leishmaniasis* is an ancient disease. Fragments of *Leishmania* kinetoplast DNA found in four Egyptian mummies provide evidence for the parasitisation of humans by *Leishmania* reaches dating back at least 4000 years (Kupfer et al., 2006).

Clinical outcomes and severity of the Leishmaniasis are, to some extent, determined by the *Leishmania* species infecting the mammalian host, however some species cause multiple and variable outcomes. Three forms of the disease are normally considered: Cutaneous Leishmaniasis (CL), Mucocutaneous Leishmaniasis (MCL) and Visceral Leishmaniasis (VL). *L. mexicana*, *L. donovani*, and *L. braziliensis*, the causative agents Leishmaniasis, typically result in CL, VL, and CL/MCL, respectively (Mann et al., 2021.) (See Table 1-1 below).

Additionally, health factors of the host, such as malnutrition or a compromised immune system (e.g., through coinfection with HIV), can influence severity (Burza et al., 2018) These aspects are further convoluted when factors of the *Leishmania* parasite modulate the host immune response, as further discussed below (Séguin & Descoteaux, 2016). Future challenges also await in determining the infectiousness of asymptomatic infections (Singh et al., 2014).

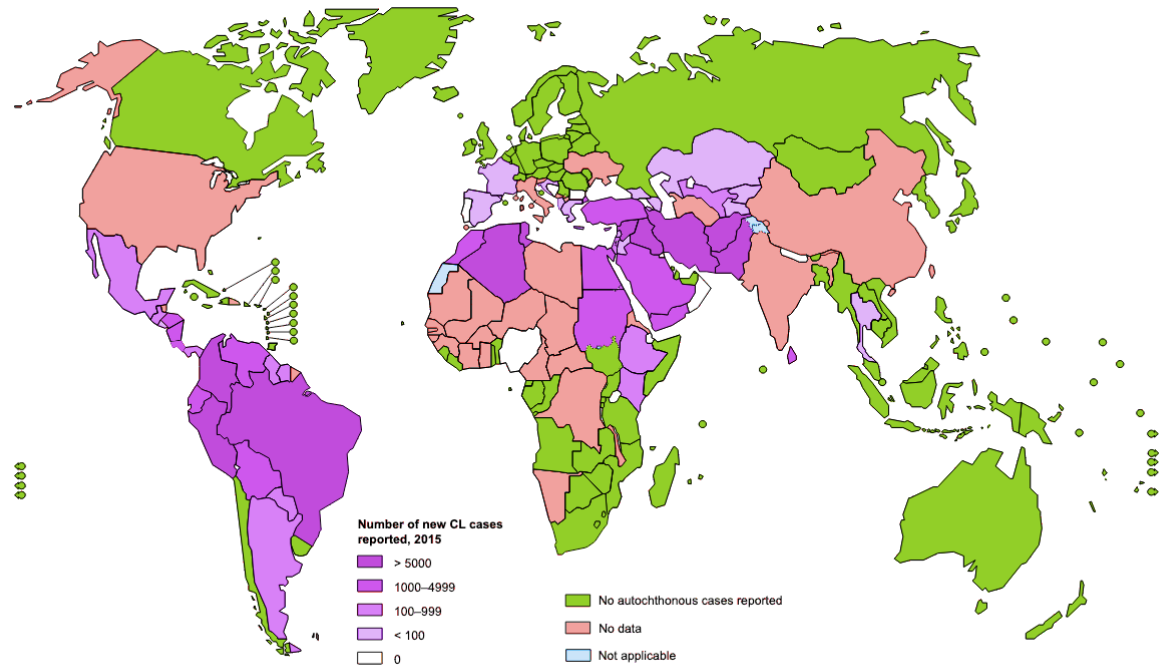
Table 1-1 Species, vectors, and diseases of Leishmaniasis.Examples of a range of different diseases associated with *Leishmania*. Adapted from Bates (2007).

Species	Sand fly vector	Disease category	World designation
<i>L. mexicana</i>	<i>Lutzomyia olmeca olmeca</i>	Cutaneous	New world
<i>L. major</i>	<i>Phlebotomus duboscqi</i> <i>Phlebotomus papatasi</i> <i>Phlebotomus Salehi</i>	Cutaneous	Old world
<i>L. braziliensis</i>	<i>Lutzomyia wellcomei</i> <i>Lutzomyia complexus</i> <i>Lutzomyia carrerae</i>	Cutaneous / Mucocutaneous	New world
<i>L. donovani</i>	<i>Phlebotomus argentipes</i> <i>Phlebotomus orientalis</i> <i>Phlebotomus martini</i>	Visceral	Old world
<i>L. infantum</i>	<i>Phlebotomus ariasi</i> <i>Phlebotomus perniciosus</i> <i>Lutzomyia longipalpis</i>	Visceral	New and old world

The most common form of the Leishmaniasis is CL, diagnosed by skin plaques, ulcers and/or nodules, which left untreated can cause disseminated/diffused CL. While typically self-healing, CL can lead to scarring and severe malformities left from skin lesions caused during the infection. The psychosocial burden and stigma of CL is further reviewed by Bennis et al., (2018). For MCL, if left untreated the mucous membranes of the airways and oral cavities are damaged affecting the larynx, pharynx, and soft palate, also leading to potential scarring. VL has a very high mortality rate if untreated, with symptoms including inflammation of the liver and spleen, weight loss, and anaemia. Also of note is the extensive life-long scarring and disfiguring skin lesions caused by Leishmaniasis can have on mental health and psychosocial wellbeing of patients globally, as reviewed by Pires et al. (2019). Further impacts include increased domestic abuse, divorce and children being excluded from schools due to perceived stigmas, highlighting how the parasite can disproportionately affect the most vulnerable and poorest in societies ((Okwor & Uzonna, 2016). Across the globe, CL has been documented as being endemic in 87 countries, and VL in 75,

according to the 2018 WHO Leishmaniasis update (WHO, 2018a). Global presence is demonstrated below in Figure 1-1.

A



B

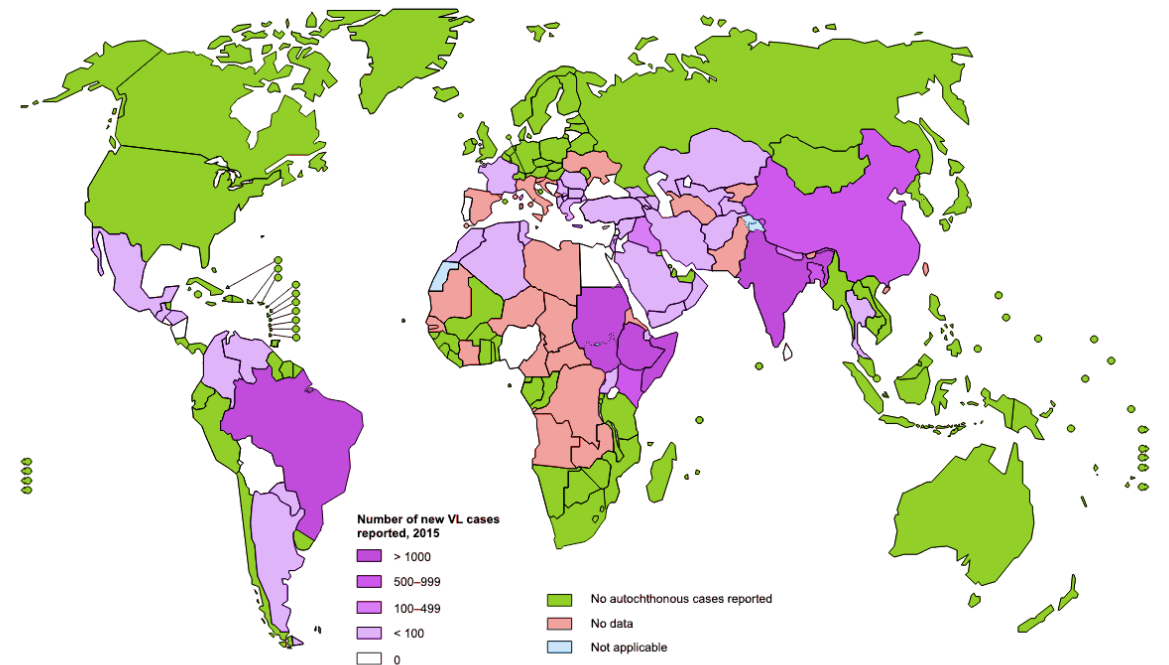


Figure 1-1 Distribution of Leishmaniasis.

Global distributions of (A) Cutaneous Leishmaniasis (CL) and (B) Visceral Leishmaniasis (VL), depicting the number of new cases reported in 2015 (WHO, 2018a).

Reported annual cases of CL in 2016 were numbered over 200,000, while VL was reported as over 22,000 cases by the WHO (WHO, 2018b). Of note, however, is the lack of incident reports for 25 countries for CL and 21 countries for VL, indicating an underestimation of the number of new cases in this report (WHO, 2018). Collectively across all forms of diagnosed Leishmaniasis, the WHO give a combined global estimate of 350 million people at risk of infection, and approximately 1.3 million new cases of Leishmaniasis arising annually (WHO, 2015). Also hampering efforts to estimate global rates of incidence for Leishmaniasis are endeavours to record outbreaks which are undermined by the remoteness of locations, and the difficulty of reporting during conflict. Despite these circumstances, global annual estimates of CL morbidity can range between 200,000 - 400,000 cases with global annual mortality of VL estimated up to 40,000 deaths, when factoring in limitations mentioned above (Alvar et al., 2012).

1.2 The Biology of the *Leishmania* parasite

Leishmania have a digenetic life cycle consisting of an invertebrate host, the sandfly, and numerous wild and domestic mammalian hosts, including cats, rodents, dogs, and humans (Bates, 2007; Roque & Jansen, 2014; Rougeron et al., 2017). The life cycle of *Leishmania* can be split into two distinct developmental stages, each representing survival and parasitic strategies dependant on the current vector or host they inhabit. The flagellated and motile infective stages, the promastigotes, reside within the sandfly. The non-motile stage, the amastigote, typically resides in macrophages, professional phagocytotic cells, and transiently in the blood of the vertebrate hosts when dividing amastigotes have burst forth from vacuoles within the macrophage (Bates, 2007). The amastigotes survive within host cells largely due to their membrane coating making them resistant to host cell defence mechanisms and signalling (Gupta et al., 2013; Shio et al., 2012). For example, mutations of the membrane-bound protein amastin, reduces parasitic viability and survival strategy effectiveness of *L. braziliensis* within the host macrophages (Marcia Cardoso de Paiva et al., 2015). Additionally, amastigotes survive and develop by appropriating nutrients and proteins scavenged from the host cell before dividing by binary fission, via the parasitophorous vacuole membrane at the parasite-host interface (Young & Kima, 2019).

Leishmania have a multifarious transmission cycle, as detailed in Figure 1-2 below and reviewed by Sunter and Gull (2017).

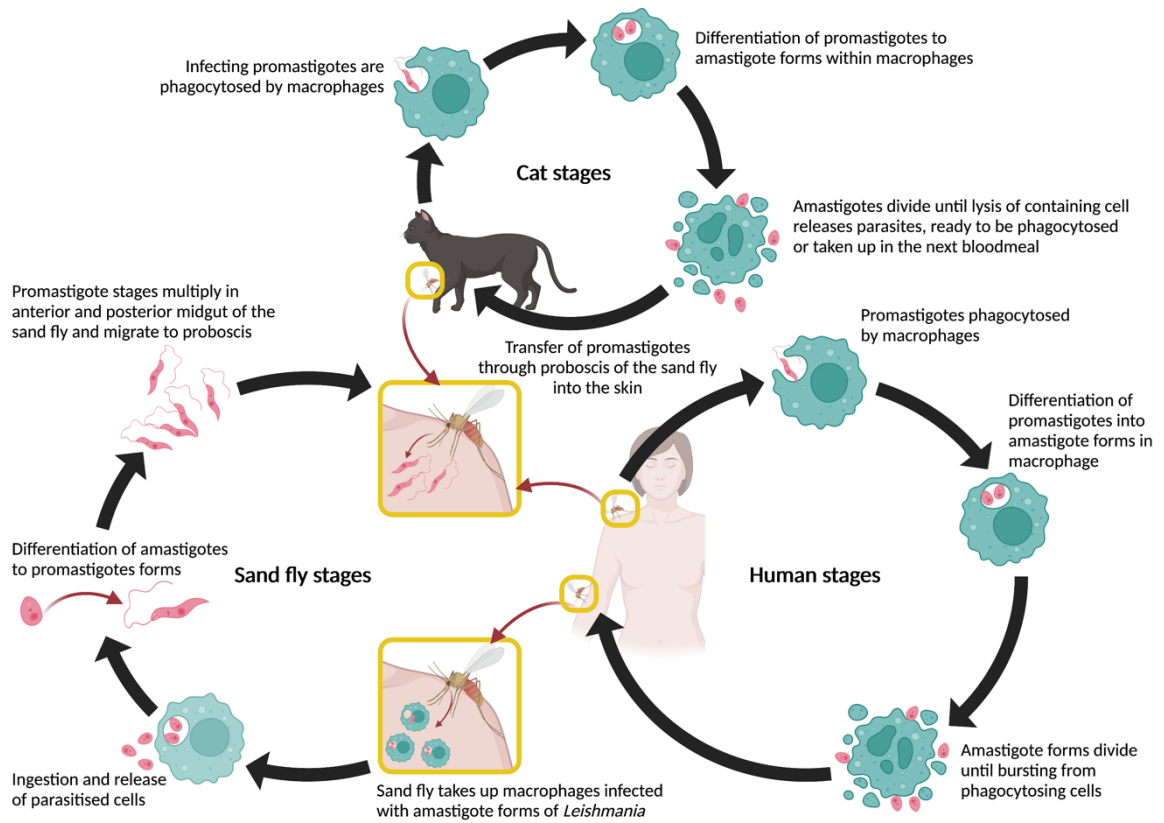


Figure 1-2 Life cycle of the *Leishmania* parasite

Life cycle forms of the *Leishmania* parasite found in the sandfly vector and mammalian hosts. *Leishmania* transfer from mammal to vector via blood meal of the sandfly where differentiation amastigotes takes place for adaptation to the invertebrate vector. Differentiation into promastigotes forms occurs from the posterior to anterior fly midgut as procyclic, nectomonad and leptomonad forms, before migration of the mammal-infective metacyclic promastigote forms to the sandfly proboscis. Metacyclic forms are then transferred into mammalian hosts via the sandfly bite and are phagocytosed by macrophages, where differentiation into the non-motile amastigote forms takes places in vacuoles within the macrophage. These amastigotes multiply in the host cell before lysing into the hosts' bloodstream, establishing the infection and ready to be taken up in the next bloodmeal. Figure adapted from Esch & Petersen, (2013). Created with BioRender.com.

Beginning with the female sandfly, a bloodmeal is ingested from a vertebrate host, such as a rodent, which contains amastigotes circulating in phagocytic cells or in the bloodstream. Once ingested, the majority of amastigotes undergo transformation to the first invertebrate life cycle stage, the replicative procyclic promastigote, within the vertebrate midgut (Gossage et al., 2003). A select population of amastigotes will continue to dividing or non-dividing forms (Alexander et al., 1999; Rogers & Bates, 2007). However, the control and selection process for these populations is still not fully understood. Once in the posterior midgut of the sand fly, the procyclic promastigotes emerge after ingestion (Dostálová & Volf, 2012). Procyclic promastigotes are ovoid in shape

with an emerging flagellum and relatively low motility compared to other promastigote forms. Procyclic forms divide asymmetrically to form the foundation of the infective stages established in the sandfly (Ambit et al., 2011; Wheeler et al., 2011). The next differentiation step is from procyclic to nectomonad promastigote, a non-dividing stage responsible for migration to the anterior midgut of the sandfly vector (Sunter & Gull, 2017). These nectomonad promastigote forms are distinguishable from procyclics by their body-length flagellum, providing them with the increased motility required for migration (Rogers et al., 2002). The non-replicative nectomonad forms then differentiate into replicative leptomonad promastigotes, which initiate a second dividing population in the anterior midgut (Gossage et al., 2003; Wheeler et al., 2011). Nectomonads are notably larger than leptomonads, being 12 - 20 μm in body length while progression to leptomonads sees a reduction in body length to 6 - 8 μm (Rogers et al., 2002; Sasidharan & Prakash Saudagar, 2021). The leptomonad stage is not only important for further establishing a parasitic population within the sandfly anterior midgut, but also for producing the factor promastigote secretory gel (PSG), which prevents the sandfly fully ingesting the bloodmeal, and blocking regurgitation, thereby aiding in transmission of the parasite (Rogers et al., 2002). The next step in the transmission cycle is differentiation from leptomonad to the non-replicative and mammal-infective life cycle stage, the highly motile metacyclic promastigote (Bates, 2007; Giraud et al., 2019). This infective stage travels via the foregut and finally resides in the proboscis of the sandfly, ready for transfer into the next host via another bloodmeal (Sunter & Gull, 2017). Each of these different forms have been described with sequential differentiation from amastigotes to procyclic, nectomonad, leptomonad to metacyclic promastigotes, taking approximately three doublings of the population during replicative stages (Gossage et al., 2003; Rogers et al., 2002)

Following egestion of the metacyclic promastigotes from the sandfly into the bloodstream of the new host, phagocytotic immune cells, typically macrophages, are signalled towards the bite site to ingest the parasites (de Menezes et al., 2016). Once phagocytosed, the parasite goes through structural and morphological shifts of its form, losing the majority, but not all, of its flagellum and forming an ovoid cell 2 to 4 μm in length (Torres-Guerrero et al., 2017). It is still not fully understood if phagocytosis may occur at locations within the host

at a distance from the original bite site, or if further host evasion and tissue transit may occur outwith infected macrophage movement. Furthermore, the polarity in which the parasite is phagocytosed by the immune cells is a matter of debate, with contrary examples indicating flagellum-first entry with others indicating cell body entry or from both polarities (Courret et al., 2002; Forestier et al., 2011). Regardless, a consistent observation is found for the final orientation of the amastigote form of the parasite within the vacuole, in which the truncated flagellum pointing towards the edge of the engulfing cell (Courret et al., 2002; Forestier et al., 2011).

Variations between different species are seen for parasitophorous vacuole occupancy. So far, two discrete parasitophorous vacuole conformations have been found, correlating with *Leishmania* species. Infection of macrophages with *L. major* produces a type I vacuole conformation, whereby small, tight-fitting single-occupancy parasitophorous vacuoles are formed (K.-P. Chang & Dwyer, 1978; Real et al., 2010). In contrast, infections with *L. mexicana* produce large multi-occupancy parasitophorous vacuoles termed type II vacuoles, which contain multiple amastigotes within the same vacuole (K.-P. Chang & Dwyer, 1978; Young & Kima, 2019). Interestingly, in co-infections of *L. amazonensis* and *L. major*, type I and type II vacuole conformations respectively, showed no mixing within vacuoles, with distinct parasitophorous vacuoles between the two *Leishmania* species formed separately in the same macrophage. This observation potentially indicates the parasitophorous vacuole conformation is used to match unknown specific needs of the respective species, or differences in the manner of vacuole generation (Real et al., 2010). Once several rounds of replication of the amastigote forms have occurred, the surrounding cell eventually ruptures, releasing the parasites into the bloodstream or neighbouring cells for further phagocytosis, or to be taken up again via the next sandfly bloodmeal, completing the transmission cycle (Loría-Cervera & Andrade-Narvaez, 2020; Rogers et al., 2002). Another potential avenue for transmission was recently observed by Doehl et al. (2017), whereby a population of *Leishmania*-infected mononuclear cells were found dispersed along the surface of the skin in VL. This transportive population of infected macrophages located in the skin will add additional efficiency to transmission of parasites from host to vector, reducing the need for bite sites located at skin lesions containing *Leishmania*. Further

speculation persists for a hypothetical primed form of amastigotes ready to survive in the sandfly midgut when transferred, akin to the metacyclic promastigote forms or the stumpy transmission form of the African trypanosome (Bates, 2007; Rico et al., 2013). A recent study by Sandoval Pacheco et al. (2021) indicates a polarisation of M1 subtype macrophages (pro-inflammatory subtype with microbicidal properties) over M2 subtypes (anti-inflammatory/regulatory subtype related to inflammation resolution and tissue repair). Studies investigating *Leishmania* interactions with other skin cell types, such as keratinocytes, are currently lacking. Further investigations into interactions with other phagocytotic immune cells, such as neutrophils, and how they are recruited to infection sites and the microenvironment generated there are required (Passelli et al., 2021).

The differentiation from one life cycle stage to another to adapt to changing vector and host environments requires regulation of the expression of developmental and cell cycle genes. In addition, multiple replicative and arrested forms also require molecular checkpoints to ensure one stage is maintained or moved into as and when needed. Next generation sequencing strategies that have examined RNA changes associated with these life and cell cycle features are discussed in section 1.3. However, many examples of regulatory control at the protein level of the infective and host-residing forms have been detailed. One such adaptation for the survival of amastigotes, which differentiate from metacyclics residing in cold-blooded invertebrates to their warmer mammalian hosts, have been heat-shock proteins (Kröber-Boncardo et al., 2020). Heat shock proteins are associated with temperature induced differentiation of *Leishmania*, between vector and host stages (Grünebast & Clos, 2020) and are currently under consideration as druggable targets (Prasanna & Upadhyay, 2021). Additional structural biology adaptations include analysis by Wheeler et al. (2015), who demonstrated rearrangement of the promastigote flagellum, comprising of 9+2 microtubule arrangement, to a 9+0 orientation in amastigotes (Wheeler et al., 2015). However, an important consideration in these multifarious regulatory steps involved in life cycle development is the putatively constitutively expressed genome of the *Leishmania* parasite, since virtually all protein-coding genes are expressed as multigene, polycistronic transcription units that share a single promoter (Clayton, 2019; Worthey et al.,

2003). Such polycistronic pre-mRNA expression is not only found in *Leishmania* but appears common to Trypanosomatida and wider kinetoplastids (Mahmood et al., 1999), where mature mRNAs are generated by coupled trans-splicing of a capped splice-leader RNA and polyadenylation (Curotto De Lafaille et al., 1992; Lamontagne & Papadopoulou, 1999). Of note, is how this polyadenylation of processed, mature mRNA allows the capture of mature mRNA transcripts by selecting for poly-A modifications in transcriptome analysis, such as polyadenylation selection methods for bulk and single-cell RNA-Sequencing (Grünebast & Clos, 2020). Despite this unconventional gene expression, *Leishmania* exhibit some change in levels of mRNA and protein between life cycle stages, adapting to their current host (Leifso et al., 2007a). With mRNA and protein levels being described in the literature as both reducing between promastigote and amastigote stages (Coelho et al., 2012; Shapira et al., 1988). How the non-canonical - i.e., non-promoter-driven - control of gene expression is determined in *Leishmania* was recently reviewed by Grünebast and Clos (2020), where they detail the absence of regulated RNA synthesis, the use of post-transcriptional gene regulation, including RNA stability, and regulated translation. Additionally, aneuploidy demonstrated in *Leishmania* also disrupts levels of mRNA, metabolites, and proteins: such as peptidases, chaperone proteins, and heat-shock proteins which can affect evasion strategies of the parasite (Cuypers et al., 2022).

1.2.1 Strategies of *Leishmania* immune system avoidance, survival, and transmission

1.2.1.1 Promastigote strategies

One strategy for parasitic survival can be found in the glycoprotein PSG, which is discharged by the promastigote forms of *Leishmania* into the midgut of the invertebrate vector. Additional roles to those described above include PSG stimulation of regurgitation for the infective metacyclic promastigote stage of the parasites when taking a bloodmeal, and thereby increasing transmission (Bates, 2007; Rogers et al., 2002). In addition, the promastigote forms secrete an enzyme termed chitinase, which aids in digestion of the polysaccharide chitin produced by most invertebrate *Insecta*, including the sandfly (Schlein et al., 1991). The digestion of the chitin surrounding the midgut of the sandfly is

thought to further encourage migration of the parasite to the proboscis, encouraging passage through bloodmeals (Dostálová & Volf, 2012; Volf et al., 2001). Evidence also exists of PSG playing pivotal roles in immune cell recruitment to bite sites and macrophage activation during mammal infection, having been described in the literature as disrupting L-Arginine metabolism in host macrophages to encourage parasitaemia (Rogers et al., 2009). In addition to these roles, PSG has been shown to influence alternative Type 2 T helper (Th2) immune responses. Th2 activation has previously been shown to aid in nutrient salvaging of the parasite within the alternatively activated macrophages and metabolic processes to encourage amastigote growth, as reviewed by Tomiotto-Pellissier et al. (2018). Further stimulation of the Th2 response can be found in the interactions with cysteine protease B (CPB). Found in all strains and stages of the parasite, CPB is an enzyme that cleaves Cluster of Differentiation 23 (CD23) and Cluster of Differentiation 25 (CD25) transmembrane proteins, which act as receptors to immunoglobulin E (IgE) and Interleukin-2 (IL-2), respectively (Pollock et al., 2003), thereby further encouraging the Th2 macrophage response (Bauxbaum et al., 2003).

Alternatively, CD4⁺ Th1 cells generating Interferon γ (IFN γ) are necessary to induce macrophage mediated Nitric Oxide (NO) attack on the infecting parasites (Scott & Novais, 2016). This Th1 mediated responses can lead to recruitment of further immune response cells, such as neutrophils, Ly6C⁺ inflammatory monocytes, and CD4⁺, CD8⁺ T lymphocytes and Natural Killer (NK) cells (da Silva Santos et al., 2014). The parasite has also been shown to influence phagocytosis. While the classical receptor-mediated recognition and instigation of phagocytosis is well documented for pathogens and their respective phagocytic cells (Flannagan et al., 2012), the process is still undergoing further analysis for *Leishmania*. While various surface proteins on the infective promastigote stages help protect the parasite from complement mediated attack from the immune system, two known complement receptors, CR1 and CR3, as well as fibronectin receptors (FnRs) have been described as involved with the process of phagocytosis (Abu-Dayyeh et al., 2010; Ueno & Wilson, 2012; Wenzel et al., 2012). Variation in receptors responsible for phagocytosis is found across different species of *Leishmania*: CR3 is responsible for *L. mexicana* receptor-mediated phagocytosis (Talamas-Rohana et al., 1990), while CR1 recognises *L.*

major parasites the CR1 mediated phagocytosis having also been recorded in neutrophils (Salei et al., 2017). Evidence also exists of surface metalloproteases, such as GP63, which actively use the complement system to encourage phagocytosis by the host innate immune cells, cleaving complement protein 3 into complement subunits C3b and iC3b, encouraging opsonisation of the parasite through CR3 phagocytosis via Cluster of Differentiation 11b (CD11b) and Fc-gamma receptors (FcγR) (Gurung & Kanneganti, 2015; Salei et al., 2017; Woelbing et al., 2006). Disruption of *N*-ethylmaleimide-sensitive factor attachment receptors (SNAREs) have also been recorded by GP63, cleaving vesicle-associated membrane protein 8 (VAMP8) and thereby inhibiting phagolysosome biogenesis and cross-presentation (Matheoud et al., 2013). Once inside the parasitophous vacuole, the parasite continues to subvert the host cells' attempts to digest the invader by arresting phagosomal maturation, as reviewed by Moradin & Descoteaux (2012). Examples of the processes involved include limiting nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex association with vacuoles containing *Leishmania*, which limits the generation of reactive oxygen species (ROS); and affecting association of vacuolar proton-ATPase (v-ATPase) for newly infected lysosomes, potentially indicating differentiation of promastigotes to amastigote forms may start in an environment less acidic than the final pH in which amastigotes typically reside (Rosenzweig et al., 2018).

In addition to the above, phagosome fusion with lysosomes and the late endosomal system is reduced by the parasite, to further allow it to differentiate into amastigote forms and thereby divide (Scianimanico et al., 1999). Strikingly, evidence has been found to suggest infection with *Leishmania* parasites increases the duration of macrophage survival, through a still unknown mechanism inhibiting apoptosis (Donovan et al., 2009; Kamir et al., 2008; Kolli et al., 2008; K. J. Moore & Matlashewski, 1994). While strategies to silence gene expression for microbicidal pathways in the host cell via DNA methylation have also been recorded, this process also remains poorly understood (Arango Duque & Descoteaux, 2015).

Lipophosphoglycan (LPG) is a major surface glycoconjugate of *Leishmania* parasites, which is involved in several aspects of their pathogenesis, including

host cell attachment and immune evasion. LPG has been shown to play a critical role in the attachment of metacyclic *Leishmania* to macrophages, which is the primary host cell for *Leishmania* parasites (Spath and Beverley, 2001). LPG can inhibit the production of cytokines by macrophages, which allows the parasites to evade immune detection and survive within the host cell (Spath et al., 2003). Additionally, LPG can modulate macrophage signalling pathways to promote parasite survival and replication (Bhardwaj et al., 2010).

Mechanistic target of rapamycin (mTOR) is a conserved serine/threonine kinase that regulates cell growth, proliferation, and metabolism in response to environmental cues, including nutrient availability, energy status, and stress. In *Leishmania* parasites, two distinct mTOR complexes, mTOR2 and mTOR4, have been identified, which have been shown to play a role in the regulation of parasite virulence and survival (van Dam, et al., 2011; Mondragon-Shem et al., 2014). mTOR2 has been shown to regulate the host immune response to *Leishmania* by modulating the production of cytokines by macrophages, whereas mTOR4 is involved in the regulation of parasite autophagy and virulence (Mondragon-Shem et al., 2014; Rashidi et al., 2021).

Overall, the relevance of LPG, mTOR2, mTOR4, and others described above, to macrophage infection highlights the importance of understanding the molecular mechanisms underlying *Leishmania* pathogenesis.

1.2.1.2 Amastigote strategies

A further example of parasitic survival in the microbicidal phagolysosome of innate immune cells can be found in the differentiation of infecting metacyclic promastigotes into amastigotes (Arango Duque & Descoteaux, 2015; Podinovskaia & Descoteaux, 2015). The metalloproteinase GP63, present in metacyclic and amastigote forms, also attenuates breakdown of the parasitic membrane in the phagolysosome, as shown experimentally by Seay et al. (Seay et al., 1996) using liposome vesicles covered with GP63, arresting degradation. Furthermore, metalloproteinases, which salvage nutrients such as glucose from the host cell, are expressed (Burchmore & Hart, 1995). Growth in acidic pH concentrations of 5.5 - 6.0 is facilitated by amastigote-specific proton pumps, which are translated and moved to the plasma membrane of the amastigotes to produce a steep

transmembrane pH between the amastigote and its environment (Burchmore & Barrett, 2001). Additionally, the alarm system of antigen presentation classically associated with phagocytotic innate immune cells, via major histocompatibility complex (MHC) II loading, is subverted by the parasite (de Souza Leao et al., 1995) by MHC class II receptors being inhibited from transportation to the host cell surface. Other examples of endocytosis of antigen presenting molecules have also been found in *L. mexicana*, such as the MHC molecule H-2M in mouse models, helping to catalyse peptide release from the MHC (Antonie et al., 1999). Additionally, evidence exists of enzymatic degradation of MHC receptors by cysteine protease B (CPB) (Mottram et al., 2004a). Interestingly, dendritic cells have also been associated with CD8⁺ T cell priming in CL, although how these cells have their MHC I presentation subverted by *Leishmania* infections is not yet fully explored (Kautz-Neu et al., 2012).

1.2.1.3 Trojan Horse vs Trojan Rabbit

The Trojan horse and Trojan rabbit theories are two models that explain mechanisms of *Leishmania* mammalian infection and evasion.

The Trojan horse theory suggests that neutrophils are the initial cells to reach the site of infection within a few hours. They are followed by inflammatory monocytes, tissue macrophages, and dermal dendritic cells, forming a group of professional phagocytes (Carlsen et al., 2015).

Upon phagocytosis by neutrophils, *Leishmania* parasites can either be eliminated or survive within these cells, even during neutrophil apoptosis.

When macrophages engulf apoptotic neutrophils that are parasitised, viable *Leishmania* can be transferred this way to the phagocytosing macrophage. This invasion process thereby allows the parasite to invade macrophages and continue to multiply within the new host cells (Martínez-López, et al., 2018; Carlsen et al., 2015).

This phenomenon, referred to as the "Trojan horse," facilitates the covert invasion and survival of *Leishmania* within target host cells, where the parasite can rapidly proliferate.

The life cycle of the parasite then advances as the protozoa infect macrophages or dendritic cells in the skin. It is currently believed that these cells play a primary role in spreading the parasites to lymph nodes and other organs such as the liver, spleen, and bone marrow. This facilitates the transportation of the parasite and enables the infection of new phagocytic cells present in these tissues (Carreira & da Silva, 2021).

In contrast, the utilization of *in vivo* imaging further revealed another evasion mechanism known as the 'Trojan rabbit' strategy, wherein parasites evade dying neutrophils to infect macrophages (Ritter et al., 2009, Beattie and Kaye, 2011). This phenomenon allows the parasites to gain entry into macrophages, their preferred host cells, and establish infection. The parasites continue to multiply within the macrophages, leading to the progression of the disease.

The "Trojan rabbit" strategy highlights the ability of *Leishmania* parasites to exploit the immune system and use neutrophils as carriers to reach their target cells, thereby facilitating their survival and multiplication. Where experimentally, the effectiveness of neutrophils as targets for immune evasion depends on both the genetic background of the host and the strain of the parasite employed in the experiments. (Ritter et al., 2009).

Both theories suggest that *Leishmania* manipulate host cells to evade immune detection and destruction, allowing them to establish a successful infection. The exact mechanisms of infections remain a subject of ongoing research.

1.3 Transcriptomics in *Leishmania* biology using bulk RNA-sequencing

Gene regulation in *Leishmania* is a complex process that involves transcriptional control, post-transcriptional modifications, and epigenetic regulation.

Transcriptomics, the study of the entire set of RNA transcripts produced by a cell or organism, has been an important tool for understanding gene regulation in *Leishmania*.

In *Leishmania*, transcriptional regulation is non-canonical, and the Trypanosomatida order lacks gene-specific, regulated transcription by RNA

Polymerase II (Grünebast & Clos, 2020). This absence is due to the absence of both typical gene promoter elements and genes encoding related transcription factors (Clayton et al., 2002). However, several non-canonical transcription factors have been identified in *Leishmania*, including the *Leishmania*-specific transcription factor LmxM.31.0670, which is involved in the regulation of virulence genes HASPB and SHERP (Sádlová et al., 2015). Instead, *Leishmania* use polycistronic transcription units (PTUs) where transcription is started and terminated (Martinez-Calvillo et al., 2004).

The regulation of gene expression in *Leishmania* necessitates precise and coordinated mechanisms that respond to changes in the environment. Kinetoplastid parasites mainly rely on post-transcriptional gene regulation mechanisms due to their constitutive transcription of Pol II-driven polycistronic gene arrays (De Pablos et al., 2016). Consequently, RNA binding proteins (RBPs) are abundant in the proteome of these organisms, reflecting their primary role as gene regulators. These trans-regulatory RBPs form ribonucleoprotein complexes (mRNPs) that dynamically bind to mRNA and regulate their processing and trafficking from synthesis to decay (Gehring et al., 2017). In response to environmental pressures, mRNP localisation, composition, and function undergo rapid changes that accelerate mRNA translation, decay, or sequestration into intracellular granules (Fritz et al., 2015). In addition, RNA-binding proteins, and post-transcriptional modifications, such as alternative splicing and mRNA stability, play important roles in gene regulation in *Leishmania* (De Pablos et al., 2019). Transcriptomics studies in *Leishmania* have focused on identifying differentially expressed genes between different developmental stages of the parasite, as well as between drug-resistant and drug-sensitive strains (Sadlova et al. 2015) used RNA sequencing to compare the transcriptomes of promastigote and amastigote stages of *L. donovani*.

To assess how transcriptome regulation may act in the timing and patterns of developmental control during the multiple life cycle changes, RNA sequencing across life cycle development has been investigated in *L. mexicana* and *L. major* by Fiebig et al. (2017) and Inbar et al., (2015), respectively. Fiebig et al., (2015) applied Illumina sequencing of poly-A selected mRNA to describe and contrast the transcriptomes of *L. mexicana* through promastigote, axenic amastigote, and

intracellular amastigote life cycle stages. Comparative analysis of transcripts across promastigote and intracellular amastigote forms from murine bone-marrow derived macrophages found 3,832 genes that were differentially expressed between promastigotes and intracellular amastigotes (Fiebig et al., 2015). As may be expected, downregulation of transcripts associated with motile flagellum formation were found during differentiation to amastigote forms. In addition, 936 genes were upregulated with functions associated with cell surface membrane-bound proteins, transporters, peptidases. Interestingly, analysis of the genome-wide distribution of differentially expressed genes disclosed a high enrichment for the tetraploid chromosome 30 being upregulated in amastigotes. Examples of the regulation of the morphological alterations in the life cycle forms, allowing *Leishmania* to survive in different hosts, can also be found in research by Inbar et al., (2017), who set out to identify changes in the transcriptome during amastigote to promastigote life cycle development of *L. major* in the sandfly and in culture by using bulk RNA-sequencing. Amongst genes upregulated from amastigotes to procyclic promastigotes include functions associated with cell cycle regulation, and glucose and protein metabolism; amongst genes showing downregulation were expected amastigote markers, such as the complement resistant surface protein amastin. The non-replicative nectomonad promastigote forms exhibited a decrease in expression of cell cycle regulated genes, such as cyclin 6, Ribosome Binding Protein 33 (RBP33), and Cell Division Cycle 20 (CDC20) congruent with cell cycle arrest (Inbar et al., 2017). Metacyclic promastigotes taken from sandflies were found to have an upregulation of amastigote-like transcripts, such as amastin and glutamate dehydrogenase, proposed as an anticipatory expression to allow survival when egested into the mammalian host (Inbar et al., 2017). Reassuringly, when comparing transcriptomes of metacyclic promastigotes taken from sandfly samples with *in vitro* culture-derived metacyclics, little variation was found in their transcriptomes, though *in vivo* derived metacyclics were distinct in their increased upregulation of transcripts associated with nutrient stress (Inbar et al., 2017). These two examples of transcriptome analysis provide insightful results into the dynamic expression of genes and how they may be involved and/or control life cycle progression and host-parasite interactions for pathogenic survival (Patino & Ramírez, 2017). In this context, single cell RNA-sequencing (scRNA-seq) is a transformative and disruptive technology which

seeks to further order and identify patterns and changes in gene expression patterns over time, with enhanced precision of analysing transcriptomic changes at a single cell level (Kulkarni et al., 2019); scRNA-seq is discussed further below (section 1.5). Further transformative examples in the literature for bulk RNA-Seq existing in the study by Dillion et al. (2015), who further deconvolved promastigote and metacyclic stage markers, noting modest 3-fold changes in *L. major*. Thus, demonstrating the possibility for transcriptome profiling of life cycle stages.

Studying the transcriptome of parasites as they progress through different stages in their life cycle and cell cycle, such as through scRNAseq, could bring a focus on the processes involved in *Leishmania* survival mechanisms, and ultimately, inform on how these survival mechanisms may then be targeted and manipulated to aid pathogenic clearance and treat infection.

1.4 The macrophage response to *Leishmania* infection at the transcriptomic level

In addition to analysis of life cycle changes within the parasite, other experiments have asked how host cell gene expression changes in response to being infected, either driven by the parasite or in response to infection. For instance, does the parasite modify host cell signalling, are there factors employed by the parasite to help resist the host cell mechanisms to defend against infection, and what methods employed in sequestering host cell nutrients to aid in parasite proliferation? Such examples can be seen in other parasites, such as *Toxoplasma gondii*, whereby evidence collected by Sibley et al. (1986) has shown modifications to the host cell vacuole making them more permeable and therefore more prone to uptake of nutrients required by the parasite. Alternative methods by *Leishmania* to modify the macrophage translational process have been found through interactions of the protease GP63, which is bound to the cell surface of the parasite and also secreted and upregulated in expression in promastigote forms (Shio et al., 2012). Jaramillo et al. (2011), have previously documented GP63 cleaving the serine/threonine kinase mTOR, which regulates the translational suppressor 4E-BP1 in the macrophage, thereby activating this translational suppressor aiding pathogenic survival and growth.

Macrophages are a mononuclear phagocytic cell, differentiating from myeloid precursors such as monocytes. Interestingly, autophagy, the process of lysosomal degradation of cellular homeostasis, is an integral component of monocyte to macrophage differentiation (Jacquel et al., 2012). A previous study by Zhang, et al., (2015) has shown that chemical and siRNA inhibitors for autophagy would instead result in apoptosis of monocytes already committed to differentiation. Where signalling pathways for differentiation activate Beclin 1, releasing it from B-cell lymphoma 2 (Bcl-2) via c-Jun N-terminal kinases (JNK) activation, resulting in cleaving autophagy-related gene 5 (Atg5), demonstrated combined factors are necessary for autophagy to begin (Zhang et al., 2015). Macrophages are typically viewed as being the main host cell in which *Leishmania* parasites may reside, sequestered within their phagolysosomes (Kaye & Scott, 2011), evidence in the literature exists of other professional phagocytes, such as neutrophils and dendritic cells, also being prone to limited infection by *Leishmania* (Charmoy et al., 2010; Kulp et al., 2008; Ng et al., 2008). Macrophages are often viewed as one of the first lines of defence for the immune system, and while phagocytosis of most pathogens by a macrophage would result in its death, *Leishmania* require to be engulfed by macrophages for their life cycle to progress. Previous research has been undertaken in the transcriptomics of macrophages in bacterial studies (Nalpas et al., 2015; Srikumar et al., 2015; Subbian et al., 2015), demonstrating inflammatory responses in the macrophage when tackling these infections, although these studies are hampered by limited hits of differentially expressed gene markers to produce a detailed infection response in the macrophage. Interestingly. Examples in the literature also exist for *Leishmania* inhibiting signalling pathways for inflammation of the gamma interferon (IFN- γ) responses by attenuating or inhibiting major histocompatibility complex class II molecules (Ray et al., 2000), the Janus kinases/Stat1 axis (Nandan & Reiner, 1995a), and Mitogen-activated protein kinases (MAPKs) (Nandan et al., 1999a).

Due to a previous lack of understanding of the transcriptomic changes associated with macrophages infected with *Leishmania*, Fernandes et al. (2016) undertook bulk RNA-sequencing analysis to postulate how the host-cell and pathogen may interact. While macrophages are typically viewed as being the main host cell in which *Leishmania* parasites may reside, sequestered within their phagolysosomes

(Kaye & Scott, 2011) evidence in the literature does exist of other professional phagocytes, such as neutrophils and dendritic cells, also being prone to limited infection by *Leishmania* (Peters et al., 2008; Ng et al., 2008; Charmoy et al., 2010). To better understand the simultaneous host-pathogen interaction at the transcriptomic level, Fernandes et al., (2016) used bulk RNA-seq dual profiling of both macrophages and infecting *Leishmania* with two CL causing parasitic species, *L. major* and *L. amazonensis*. Strikingly, little variation in expression of orthologous gene sets were found at the transcriptomic level between the two parasitic species, and in the host response at the transcriptional level when infected with the two CL parasites (Fernandes et al., 2016). Variation was, however, seen in parasitic responses between human and murine macrophages: in *L. major*, 1,133 host transcripts were upregulated and 754 down-regulated in human cells, while 862 transcripts were upregulated and 764 down-regulated in murine cells.

The most variation found in *Leishmania* transcripts between human and mouse-infected macrophage experiments was seen in Surface Antigen Protein-2 (PSA-2), GP63 and META domain containing protein (META1), all membrane-bound surface proteins associated with metacyclic promastigotes and their infection of macrophages, where these transcripts were downregulated in humans but upregulated in mice. Conversely, cathepsin L-like proteases and RNA binding protein 5 (RBP5) were upregulated in humans and downregulated in mice. GP63 is known to play a pivotal role in evading complement-mediated lysis of the parasite, being a zinc-dependant metalloprotease, which cleaves C3b to iC3b (Dunkelberger & Song, 2010; McConville et al., 1992). Additionally, GP63 have been shown to activate entry into host cells by complement receptor 1 (CR1) (da Silva et al., 1989) and 3 (CR3) (Mosser & Edelson, 1985), having been previously attributed to intracellular survival. Examples have also been found for GP63 altering host signalling pathways by interacting with Myristoylated Alanine-Rich C-kinase Substrate (MARCKS) related proteins (MRPs) (Corradin et al., 1999), which have been linked to inflammatory responses (Amri et al., 2018). META1 has previously been shown to be associated with infectivity, being upregulated in procyclic to metacyclic promastigote forms and localising to the flagellar pocket (Gonzaga dos Santos et al., 2011). Cathepsin-L-like proteins have been demonstrated to modulate host immune responses in *Leishmania* (Gonzaga dos

Santos et al., 2011; Mottram et al., 2004) and have been considered as potential targets for removing parasitaemia in *L. mexicana* infections (de Sousa et al., 2015). Finally, RBP5 has recently been associated with proliferation in the related kinetoplast *T. brucei* (Gilabert Carbajo et al., 2021). Of note, is the stark variation in these virulence factors between human and murine hosts.

In the human macrophages analysed by Fernandes et al. (2016), several inflammatory response cytokines were found to be significantly upregulated when compared to naïve controls, including interleukin-1 β (IL-1 β), interleukin-6 (IL-6), tumour necrosis factor (TNF), and TNF superfamily members. In addition to these inflammatory bursts identified in the human macrophage infected with *Leishmania* was the transient nature of the transcriptomic response, where differential expression (DE) analysis displayed an increase in such factors 4 h post-infection, while 24 h, 48 h and 72 h timepoints showed little significant difference in transcriptomic expression for these pathways compared to large particle phagocytotic controls, which used latex beads to differentiate from baseline phagocytic responses. This transient effect is potentially unique in the literature for *Leishmania* infections when compared to other pathogens that also inhabit macrophages, such as *Mycobacterium tuberculosis* (P. Chandra et al., 2022; Sasindran et al., 2011), which are thought to induce a prolonged and perhaps continuous inflammatory response.

In a similar study previously undertaken by Dillon et al., (2016) the dual transcriptomes of *L. major* and murine macrophages were investigated. Using enriched Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis in tandem with their DE gene lists, early transcriptomic responses were again found to be related to multiple signal pathways in the murine macrophages. Similar to Fernandes et al. (2016), TNF pathway enrichment was identified, along with MAPK and Jak-STAT pathways. In addition, phosphoinositide-3-kinase-protein kinase B (PI3K-Akt), nuclear factor kappa light chain enhancer of activated B cells (NF-kappa B), Hypoxia-Inducible Factor (HIF-1), and pathways linked to glycolysis, arginine and proline metabolism, and cytokine-cytokine receptor interactions, were identified (Dillon et al., 2016). Conversely, anti-inflammatory, tissue growth, and repair pathways were enriched in 4 h post-infection. Dillon et al., (2016) noted similarities in such responses to *L. major*

infections in murine macrophages and those described by Fleming et al., (2015) using macrophages exposed to LPS and immune complexes. A consistent observation was also made in the transient expression of the transcriptomic immune responses in murine macrophages, being highly expressed 4 h post-infection timepoint and being significantly reduced in following timepoints, with the exception of glycolytic and ATP production pathways, which were also enriched at 24 h post infection, but again reduced at 48 h and 72 h timepoints (Dillon et al., 2016).

More recently, both bulk RNA-Seq and high-resolution scRNA-seq analysis has been undertaken in *in vivo* murine infections of *L. major* by Venugopal et al., (2022). These data include sampling of infections placed intradermally at the ear of the mice. Lesions formed on the infection site were then excised post-infection and separated into single-cell suspensions for analysis of all recruited cell types towards the lesion. In their bulk analysis, response pathways associated with antigen processing, such as cluster of differentiation 4 (Cd4), CD8 antigen beta chain 1 (Cd8b1), and Regulatory Factor X5 (Rfx5), were identified (Venugopal et al., 2022). Alongside such changes were chemokine signalling molecules, such as Chemokine (C-X-C motif) ligand 9 (Cxcl9), and Chemokine (C-C motif) ligand 5 (Ccl5) and cell adhesion molecules, such as Integrin alpha M (Itgam), Selectin P Ligand (Selplg), and Vascular cell adhesion molecule 1 (Vcam1). In contrast, ribosomal pathways were found to be significantly downregulated in their bulk analysis (Venugopal et al., 2022). In their scRNA-seq analysis the heterogeneity of the recruited cells to the lesion sites was further revealed. For example, a large variation in interferon induced GTPases, and antigen presentation molecules was discovered in neutrophils, inflammatory monocytes, and monocyte-derived macrophages being recruited to the infection site (Venugopal et al., 2022). These observations match previous examples of the heterogeneity of cells involved in the host immune response to *Leishmania*, as reviewed by Sacks & Noben-Trauth (2002) and Scott & Novais (2016). Interestingly, Ingenuity Pathway Analysis (IPA) of the scRNA-seq samples found a downregulation of Eukaryotic Initiation Factor 2 (EIF2) signalling, Eukaryotic translation initiation factor 4/ Ribosomal protein S6 kinase beta-1 (eIF4/p70S6k), and mammalian target of rapamycin (mTOR) pathways in various cell types, including macrophages (Sacks & Noben-Trauth, 2002; Scott & Novais,

2016). This represents the first evidence of the potential role EIF2 may have in *L. major* infections of mice.

Of note within the expansive murine cell dataset by Venugopal et al., (2022) is the lack of clearly detectable *L. major* transcripts found in the cells being sequenced. While some transcripts were mapped to the *L. major* genome originating from within multiple murine cell types, a significant population of *Leishmania* was not found or analysed in these samples. In the following sections I shall describe the advantages of applying scRNA-seq to both host-cell and parasite interactions in tandem to further infer possible transcriptomic responses.

1.5 Single cell RNA-sequencing

Single-cell RNA sequencing (scRNA-seq) is a PCR-based approach for capturing the RNA transcriptome at a single cell resolution, where transcriptomes of individual cells can be analysed, captured to allow DE analysis, and compared for similarity with all other cells captured within the sample(s) (Chen et al., 2019; Haque et al., 2017; Hwang et al., 2018; Zheng et al., 2017). Over recent years, scRNA-seq has emerged to become a revolutionary and disruptive technology for approaching key biological questions surrounding cell heterogeneity and biological development, including embryonic cells, cancers, stem cells and, more recently, single-celled parasites (see below, section 1.6). As such, Nature Methods awarded scRNA-seq its 'methods of the year' accolade in 2013 (*Method of the Year 2013, 2014*). This technology was originally developed following a technological breakthrough by Tang et al., (2008), where sequencing of a single blastomere and oocyte transcriptome was achieved. Since these first steps, the technology and experimental procedures have advanced to encompass high-throughput data acquisition and analysis, where some technologies can now sequence hundreds of thousands of cells in multiplexed samples (Stuart & Satija, 2019) (See Table 1-2 below).

Table 1-2 Technologies available for single-cell RNA-Sequencing experimental protocols. A selection for comparison between established scRNA-seq platforms and technologies. Abbreviations; Unique Molecular Identifier (UMI), Fluorescence-Activated Cell Sorting (FACS). Adapted from Jovic et al., (2021).

Platform	Isolation method	Cell No.	UMI	Region	Reference
Smart-Seq 2	FACS	Hundreds of cells	No	Full-length	Picelli et al., (2013)
Fluidigm C1	Microfluidic	Hundreds of cells	No	Full-length	Xin et al., (2016)
10x Genomics	Microdroplets	Thousands of cells	Yes	3' end	Zheng et al., (2017)

Where quantitative RT-PCR was originally applied to analyse the levels of mRNA of individual genes, this has been largely supplanted by more comprehensive and quantitative bulk RNA-sequencing, where sample preparation would require isolation and purification of particular sub-types, and their transcriptomes sequenced to infer gene expression from a grouped population. This process is limited in that gene expression cannot be assigned to any potential sub-population residing within the sample (Thind et al., 2021). By contrast, scRNA-seq technologies allow for mixed samples to be sequenced at the same time and find individual transcriptomes to analyse a mix population without prior separation, allowing for high-resolution analysis of rare sub-populations and developmental changes associated with gene expression changes across time. These sub-populations can be compared and clustered by similarity, separated and compared for differential expression of their genes, categorised by GO and KEGG enrichment, cell cycle scoring assigned per cell, DE performed to find genes that then can be investigated, and developmental trajectories inferred across a virtual timeline, termed pseudotime (Briggs, et al., 2021). Such applications for analysing these datasets have led to a transformation of our understanding of rare cell types, the immune response, infection biology,

developmental biology, tumour heterogeneity, and precision oncology (Zheng et al., 2017; Haque et al., 2017; Hwang et al., 2018; Chen et al., 2019).

The initial stage for producing a sample for scRNA-seq is to generate a single cell suspension (Haque et al., 2017; See et al., 2018). This suspension requires high viability of the cells to reduce dead or dying cells having to be removed from analysis after cell capture, by quality control cut-off steps during the analytical pipeline and workflow. Furthermore, dying cells release free RNA to create noise, requiring removal from samples before further analysis can take place. Once a viable sample has been produced and cell numbers calculated, experimental platforms utilise different methods to isolate cells; for example, Fluidigm C1 uses micro-fluidics, Smart-Seq2 uses FACS for separation of cells in plates, and 10x Genomics separates cells through microdroplets. Each platform has relative advantages and limitations with respect to sample size, transcript lengths, and regions amplified (Jovic et al., 2022). Considerations for scRNA-seq analysis pipelines have been provided below in Figure 1-3, adapted from a review on scRNA-seq applications in kinetoplastids by Briggs et al., (2021).

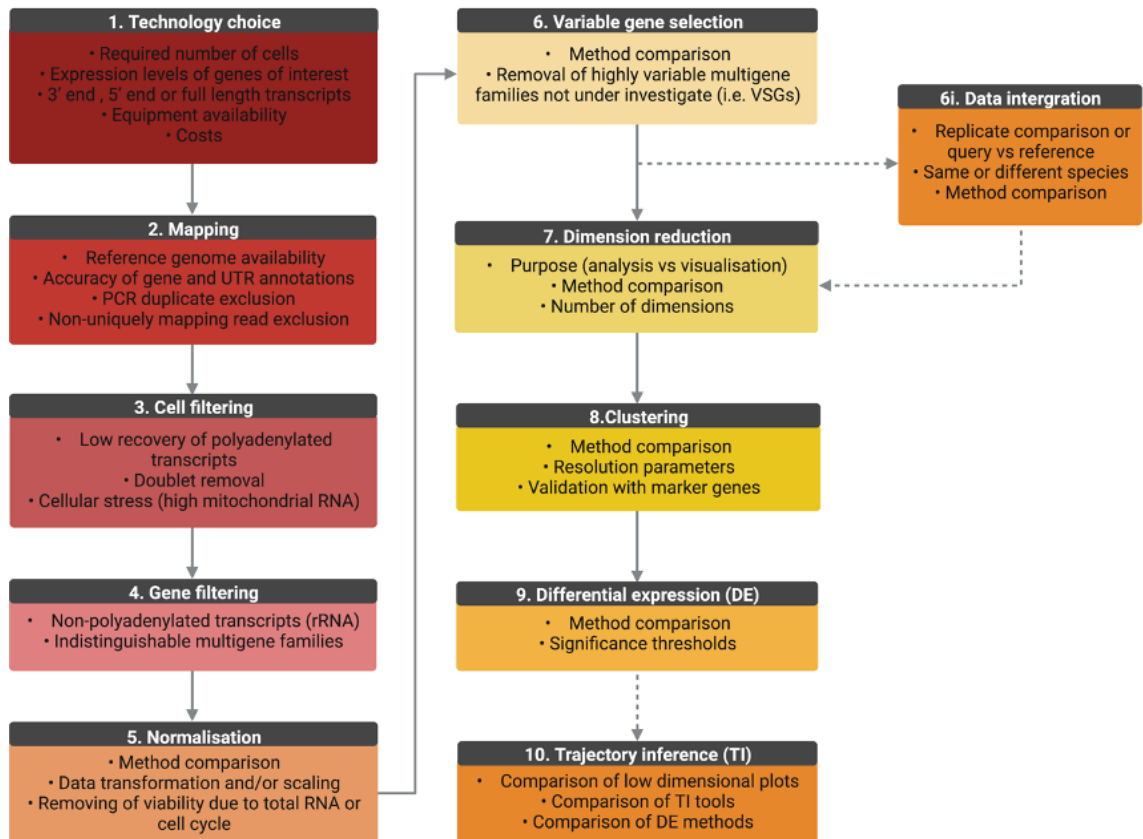


Figure 1-3 Life cycle of the *Leishmania* parasite

Deliberations and steps to consider for scRNA-seq analysis. [1] Experimental circumstances will determine the ideal platform and technology. [2] Once samples have been run on the platform of choice, considerations should be made for mapping counts data to the reference genome of choice. [3] Quality control filtering steps for removal of low-quality cells, such as low count numbers or mitochondrial counts. Considerations should be made for variations in these parameters from cell types to cell type. [4] Any sources of RNA that is typically not polyadenylated should be removed, such as most rRNA. [5] Dataset dependant normalisation with transformation and/or scaling is required. [6] Top variable genes selected for further analysis, cut-offs again dependant on the cell type being analysed or if samples have been integrated. [6i] Integrated datasets mapped onto reference datasets, such as cell atlases with various methods compared. [7+8] Dimension reduction determined by experimental set-up and clustering to provide biologically relevant distinctions between clusters. [9] Differential expression analysis to distinguish similarities or differences between developmental stages or treated versus naive samples. [10] Trajectory inference tools for ordering gene expression in given datasets across a virtual timeline, termed pseudotime analysis. Examples of such analysis include the identification of potential transient transcripts that may control and/or affect development. (Adapted from Briggs et al., 2021). Created with BioRender.com

In this thesis, the 10x Genomics Chromium platform was the system utilised for all single-cell sample capture and analysis (See Figure 1-4 below).

1.5.1 10x Chromium platform for single cell RNA-sequencing

For the 10x Genomics Chromium platform, the second stage of sample processing is to separate and compartmentalise individual cells so that RNA can be released after lysis of individual cells and labelled with a 16-nucleotide oligonucleotide barcode, specifying cellular identity to the captured RNAs (See

et al., 2018). Within the molecular barcoded oligonucleotides are individual 12 base oligonucleotides, termed unique molecular identifiers (UMIs), which allow identification of identical transcripts that have been replicated during PCR to remove the PCR bias in the data, and a 30-nucleotide poly[dT] to capture polyadenylated RNA. The Chromium platform is formed by a Chromium Controller unit and a Chromium next GEM chip, consisting of two separate inputs: a channel containing aqueous gel beads and a second containing suspended cells at the desired concentration (Figure 1-4, B). Suspended cell samples are loaded into a well on the Chromium Next GEM Chip, along with a master mix and reverse-transcriptase reagents and enzymes necessary for gel beads in emulsion (GEMs) to form in the final step of the separation process (Briggs, et al., 2021a). Individual cells are separated into microdroplets with barcoded gel beads, which are added via another well on the Chromium Chip. Captured individual cells and barcoded gel beads are separated by partitioning oil in the central stream to ensure one cell is associated with one barcoded bead (Figure 1-4, C). The Chromium Chip is then loaded into the 10x Chromium Controller, where droplet-based encapsulation produces the final nanodroplet GEMs containing one cell and one barcoded bead (Briggs, et al., 2021a).

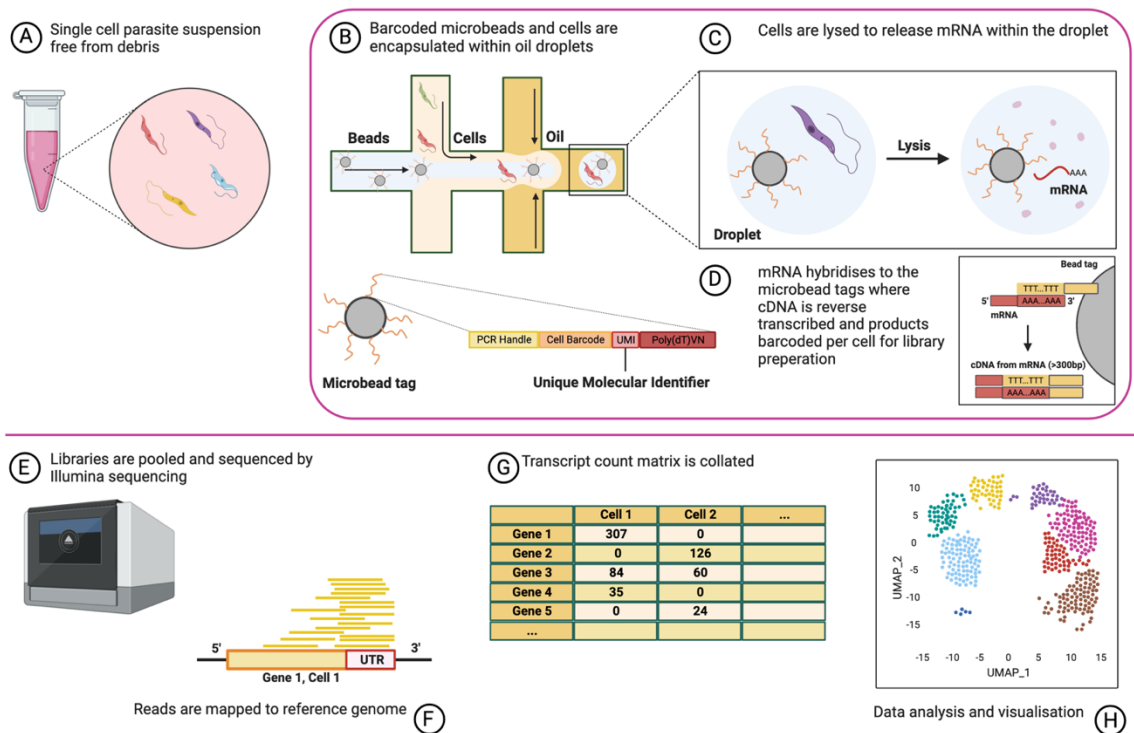


Figure 1-4 The 10x Genomics Chromium workflow.

[A] A single cell suspension is filtered to remove any debris and ensure a few cells associated with one another in suspension as possible. Counts and required density attained ready for sample to be loaded onto the 10x Chromium Chip. [B] Barcoded beads are added via another channel and nanodroplets separated by separation oil so that each droplet contains a single cell and one barcoded bead. [C] Once gel beads in emulsion (GEMs) occurs to compartmentalise cells and beads together, cells are lysed within. [D] Reverse-transcription takes place within GEMs and cDNA is generated from polyadenylated RNA by reagents included in the individual droplets. [E] Barcoded cDNA is then released from GEMs and pooled together for next-generation Illumina sequencing. [F] Reads are mapped to the reference genome of choice. [G] A transcript count matrix is assembled, and quality control steps exclude cells with low quality transcripts. [H] Data is plotted on low-dimensionality plots, such as Uniform Manifold Approximation and Projection (UMAP) plots, where each dot represents the captured and filtered transcriptome of an individual cell which are clustered together by likeness. (Adapted from Briggs et al., 2021a). Created with BioRender.com.

Individual GEMs undergo lysis, then RT-PCR in a thermocycler to produce cDNA tagged with UMIs and barcodes from their gel beads, so that each transcript can be associated with the cell of origin captured in the GEM (Figure 1-4, D). These tagged transcripts can then be pooled from all the GEMs into one suspension, run through a clean-up phase, and primers added for cDNA amplification. Following amplification, cDNA is then sequenced by next-generation Illumina sequencing (Figure 1-4, E). After sequencing, cDNA can be aligned to the reference genome of the organism(s) in the sample, and counts allocated from UMIs for captured transcripts and mapped to the individual cell of origin using bioinformatic tools (Figure 1-4, F). Further processing of the data is made by standard quality control cut-offs to exclude any low-quality cells or multiplets captured.

1.6 Single cell RNA-sequencing in kinetoplastids

As previously stated, one of the more powerful aspects of the application of scRNA-seq is the dissection of a heterogeneous population into discrete cell types, which suits very well the examination of variation seen in life cycles across parasites. Examples of this approach include *in vivo* transcriptional analysis of mice infected with *L. major*, giving distinct transcriptomic profiles of both parasite and infected murine myeloid cells (Venugopal et al., 2022). Using scRNA-seq Louradour et al., (2022) focused on *in vitro* hybridisation of *L. tropica* promastigote forms, highlighting the effects of stress responses in the *Leishmania* sexual cycle. Other examples of the power of scRNA-seq in Trypanosomatida include works in *T. brucei* by Briggs et al. (2021b), where asynchronous developmental stages between replicative slender form and transmissible stumpy bloodstream forms were elucidated (Briggs, et al., 2021b). The same study also identified the cell cycle stage of proliferating cells, revealing that short stumpy forms exit the cell cycle at early G1, and examined the influence of a known developmental factor, ZC3H20, by performing via scRNA-seq of a null mutant, revealing the change in developmental progress of the parasite (Briggs et al., 2021b). Further examples in the literature in *T. brucei* include a study by Howick et al., (2022), which unveiled sexual stages of the parasite in the tsetse and identified a role for HAP2, a gamete fusion protein, similar to *in vitro* work in *Leishmania* by Louradour et al., (2022). Further examples and applications in the literature have been considered in a review by Briggs et al. (2021b), such as a study by Vigneron et al., (2020) which explored further heterogeneity of *T. brucei* in tsetse fly salivary glands, being able to identify and label sub-populations of epimastigotes, early stage metacyclics and last stage metacyclics. Further examples include scRNA-seq used by Hutchinson et al., (2021) to elucidate variant surface glycoprotein monoallelic expression in *T. brucei* derived from tsetse fly salivary glands. Further studies in *T. brucei* using Smart-Seq scRNA-seq based technology by Muller et al. (2018) demonstrating that mutations of histone variants amplified antigen-gene clustering, DNA availability at loci associated with antigen expression, and interchanging of the translated antigens, by homologous recombination.

As such, these studies show the power and application of scRNA-seq to investigate heterogeneous populations, with future considerations for applications including the further elucidation of the various life cycle stages in *Leishmania* promastigotes (Serafim et al., 2018), and how infected cells may sub-categorise into varying response signal pathways, and indeed how the parasite may manipulate these response mechanisms. scRNA-seq has also been extensively applied to examine the immune response over various infection biology backdrops (Buchholz et al., 2016; Buchholz & Flossdorf, 2018; Chattopadhyay et al., 2014, 2015). Potential examples of heterogeneity that could be explored using scRNA-seq include African trypanosomes in tissue reservoirs located across the mammalian host, such as adipose tissue and within the skin (Girard et al., 2021; Trindade et al., 2016). Similar to Vigneron et al., (2020), experimental samples could also explore *T. cruzi* life cycle stages within their insect vector, the triatome bug, where trypomastigotes, spheromastigotes and epimastigotes life cycle stages reside within the gut of the invertebrate (Castro et al., 2007). Sexual life cycle stages in *T. cruzi* have also been recorded but are, as yet, unexplored via scRNA-seq (Schwabl et al., 2019).

1.6.1 Trajectory inference

Though less widely applied in kinetoplasts as yet, scRNA-seq allows understanding and dissection of dynamic processes, such as the cell cycle, cell activation, and differentiation. Such processes can be modelled from scRNA-seq data with computational tools, termed trajectory inference (TI) methods, as reviewed by Saelens et al., (2019). This method of analysis orders individual cells in a given sample by their captured transcriptome across a virtual timeline, called pseudotime, which are then ordered by likeness of their expression pattern. Examples of 45 pseudotime analysis tools that performed well in benchmark tests of 110 real and 229 artificial datasets were compared by Saelens et al., (2019), which found Slingshot and Monocle to be reliable tools for ordering cells across pseudotime branches and topology scoring, respectively. Although, as noted by the authors, variation in accuracy of differing pseudotime analysis tools depends on the complexity of the datasets. This method of ordering cell differentiation by TI was employed by Briggs et al., (2021b) on their asynchronous bloodstream samples containing slender and stumpy forms in *T. brucei*, where transient markers potentially holding integral differentiation

roles in the parasite were identified. Similarly in *T. brucei*, Howick et al. (2021) used TI to infer developmental markers between prospective sexual forms, using the gamete fusion protein HAP2 to identify clusters, through metacyclic forms expressing antigenic surface genes. TI has been used in numerous diseases and infection backgrounds related to macrophage infection dynamics and differentiation (Lantz et al., 2020; Rizzo et al., 2020; Wang et al., 2020; Wauters et al., 2021). However, no TI has been applied previously in *Leishmania* analysis at all, or in dual-transcriptomic approaches using scRNA-seq with macrophages infected with *Leishmania*.

1.7 High-throughput gene tagging using CRISPR-Cas9

Once a selection of genes of interest (GOI) have been identified for differing expression using scRNA-seq analysis methods, such as TI, a high-throughput method for validating and analysing the GOI should be selected. Recently, huge bounds have been made by the Gluenz group (Beneke et al., 2017; Beneke & Gluenz, 2019) using Clustered regularly interspaced short palindromic repeats/CRISPR-associated gene 9 (CRISPR-Cas9) systems (Doudna & Charpentier, 2014). The CRISPR-Cas9 system was originally isolated from *Streptococcus pyogenes* (Doudna & Charpentier, 2014). Where other *Trypanosoma* species, such as *T. brucei*, have used RNA interference systems to examine gene function at a genome-scale, many *Leishmania* species have been delayed in such approaches due to the absence of RNAi components (1990). The first example of gene replacements or deletion in *Leishmania* were demonstrated by Cruz & Beverley (1990), using homologous recombination. Genome-scale screens using this approach can be hampered in *Leishmania* by the plasticity found in chromosome copy numbers (Rogers et al., 2011). Additionally, a further impediment is the requirement of *Leishmania* species for longer lengths of sequence homology arms to target cassettes for homologous integration (Dean et al., 2015). To address these limitations for GOI tagging and knock-out, Beneke et al., (2017) developed a CRISPR-Cas9 cassette method for generating sequence-specific double-strand DNA breaks by the Cas9 ribonucleoprotein complex. This accurate and programmable sequence specificity is provided by designing single-guide RNAs (sgRNAs) for the GOI in cell lines of *L. mexicana*, *L. major* and *T. brucei* constitutively expressing stable T7 TNA polymerase, selected for with antibiotic resistance genes. Building on pioneering works using episomal plasmids which

expressed the Cas9 cassettes as performed in *T. cruzi* (Lander et al., 2015), *L. major* (Sollelis et al., 2015), and *L. donovani* (Zhang & Matlashewski, 2015) using various methods of *in vivo* transcription of sgRNAs which had been transfected with a sequence of donor DNA for homology directed repair of the knock-out or tag. These sgRNAs require only 124 nucleotide sequences to be transfected, with overlapping nucleotides containing the T7 promoter sequence, which can be readily ordered from manufacturers and amplified by standard PCR (Jones et al., 2018). Previous to the high-throughput method described in Beneke and Gluenz, (2019), an estimated 200 out of approximately 9000 *Leishmania* genes had been successfully knocked-out (Jones et al., 2018). Now, whole cohorts of genes have been knocked-out and analysed; including the flagellum proteome (Beneke & Gluenz, 2019), kinome (Baker et al., 2021), and deubiquitination proteomes (Damianouid et al., 2020). A current project from the Gluenz lab also aims to knock-out all genes in the *L. mexicana* genome using their high-throughput CRISPR/Cas9 method. Beneke & Gluenz utilise the specific mechanisms of double strand breaks (DSBs) found in *Leishmania*, and other kinetoplastids, whereby DSBs are re-joined via microhomology-mediated end joining (MMEJ) (W. W. Zhang & Matlashewski, 2015b) instead of the typical repair pathway and mechanisms of nonhomologous end joining (NHEJ) utilised in mammals (Chang et al., 2017). As MMEJ only requires recognition of condensed nucleotide sequences on both sides of DSBs, homology flanks of repair templates need only be 24 nucleotides long (Beneke et al., 2017), providing accuracy in tagging and knockouts of the GOI. Additionally, sgRNA design tools have been provided via an online app, named after the method, www.LeishGEdit.net. Plasmids can be used to provide and amplify donor DNA constructs, consisting of homology flanks of just 30 nucleotides matching the GOI, and antibiotic drug-selection genes for selection of cell lines correctly modified by the CRISPR-Cas9 system. Due to skipping any required gene-specific cloning steps, this method proves to be highly scalable to 96-well plates transfections for mNeonGreen tags of a suite of selected GOI (Beneke and Gluenz, 2019).

1.8 Thesis summary and aims

For consideration in this thesis is the application of scRNA-seq technology in *L. mexicana* to further deconvolve the timing and patterns of gene expression as

the parasite develops through life cycle forms. Questions addressed in the following chapters include:

Can scRNA-seq elucidate further the life cycle and cell cycle of *L. mexicana*?

In Chapter 3, we utilised scRNA-seq data generated with the 10x Genomics Chromium platform to capture the transcriptomes of discrete *L. mexicana* parasites as they differentiated from promastigote to axenic amastigote stages *in vitro*. Samples are separated and analysed initially by individual replicates, before being analysed as an integrated dataset in Chapter 4. This highly intricate data will then be analysed for differential expression of genes between life cycle stages, and cell cycle stages assigned to sub-populations using bulk RNA-Seq orthologous gene markers found and regulated in related kinetoplastids, so that differential gene expression changes could be investigated between the cell cycle of the different parasite life cycle stages. Pseudotime analysis of gene expression changes over a virtual timeline will be employed and cells ordered into modules of genes that can then be associated with developments through cell cycle transitions. It is hypothesised that discrete life cycle and cell cycle clusters will be identified using classic stage specific markers, with potentially hundreds of new additional markers identified for each.

Do we have a full and complete understanding of the timing and patterns of gene expression as *L. mexicana* progresses through its life and cell cycle?

In Chapter 3, as discussed above, individual replicate datasets are integrated for analysis of life cycle clusters and cell cycle labels applied to identified replicating clusters. TI with pseudotime analysis is applied for integrated datasets and modules of transient genes potentially associated with developmental progression identified. It is hypothesised that TI will help identify and order transient expression patterns of transcripts that may be involved in controlling and activation of the progress between life cycle and cell cycle stages.

Can we validate hypothetical proteins identified from the transcriptomic level at the protein level?

In Chapter 4, following on from the above analyses, 96 GOI identified by scRNA-seq analysis are selected for mNeonGreen tagging at the C-terminus using the high-throughput CRISPR-Cas9 methodology described above in this chapter. 6 known genes associated with life cycle stages and the synthesis-phase in the cell cycle are chosen as positive controls for expected expression patterns. While 90 hypothetical proteins, not previously studied tagged in the literature, are selected to validate if their protein expression might match their transient expression at the mRNA level. It is hypothesised that the majority of tagged hypothetical proteins will match their transient transcriptomic expressions and thus be applicable to inferring possible functional roles in life cycle and cell cycle progression.

Can the dual transcriptomes of human macrophages and infecting *L. mexicana* cells be analysed by scRNA-seq to identify additional factors associated with infection?

In Chapter 5 human macrophages grown in culture, containing infecting *L. mexicana* that have differentiated into amastigotes forms within them, will be analysed using scRNA-seq. The integrated dataset formulated in Chapter 4 will be used as a reference for aligning *L. mexicana* that have infected macrophages, to compare and contrast transcriptomic differences. Thereby potentially identify factors involved in infection of human macrophages and simultaneously identify the host response to infection by *L. mexicana*. Additionally, a new TI analysis will be performed with these combined datasets to compare the timing and expression patterns of amastigotes infecting human macrophages against axenically cultured amastigotes. It is hypothesised that additional factors, not previously identified in axenic amastigotes, will be discovered. These additional transcripts may provide new targets to be considered for future chemotherapeutics.

Chapter 2 Methods & materials

2.1 Methods and lab procedures

2.1.1 *Leishmania in vitro* culture and differentiation

The species and strain of *Leishmania* used in this study was *L. mexicana* M379 for all experiments (Lainson & Strangways-Dixon, 1964), excluding those used for CRISPR tagging experiments. Promastigote parasites were recovered, and cultures were expanded in haemoflagellate minimal essential medium (HOMEM) (GE Healthcare) supplemented with 10% heat-inactivated foetal bovine serum (HiFBS), named complete HOMEM (cHOM), at 25 °C in non-vented flasks and maintained in logarithmic phase culture by routine passage every 2 - 3 days. Axenic amastigotes were cultured in Schneider's *Drosophila* Medium (Gibco) adjusted to pH 5.5 supplemented with 20% FBS and 3mL of 2.5mg/mL haemin in 50mM NaOH (cSDM). Amastigotes were transformed from promastigote cultures by placing 1×10^6 cells/mL late log/stationary phase promastigotes in cSDM and incubating at 32 °C with 5% CO₂ in vented flasks and were maintained in culture by weekly passage.

For the CRISPR tagging experiments a modified strain of M379, LmexCas9T7, was kindly donated by the Gluenz lab (Beneke et al., 2017). For tagging experiments, promastigote forms of LmexCAS9T7 and their genetically modified derivatives were grown at 27 °C in M199 medium (Life Technologies) supplemented with 2.2 g/L NaHCO₃, 0.005% haemin, 40 mM 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid pH 7.4 and 10% FBS (Dean et al., 2015). Relevant selection drugs (InvivoGen) were added to the medium, when required, at the following concentrations: 32 µg/mL Hygromycin B Gold, 5 µg/mL Blastidicin S Hydrochloride, 50 µg/mL Nourseothricin Sulfate (NTC). Following transfections and drug selection for two weeks, tagged cell lines were then grown and adapted to cHOM without antibiotic selection, as per all other *L. mexicana* M379 cultures used here. Both promastigote and axenic amastigote cultures underwent a maximum of 15 passages before a new cell line was retrieved.

2.1.2 Growth curves

Parasite growth and viability was monitored by diluting 20 µL of culture 1:1 in Trypan Blue with 2% formaldehyde, and then placing 10 µL of the 1:1 fixed cells

on a Neubauer haemocytometer for the Trypan Blue exclusion counting method (Cuervo et al., 2009). Following cell counting, a parasite growth curve could be formulated by plotting cell number vs. culture time.

$$\text{Doubling time} = \frac{\text{duration} * \log(2)}{\log(\text{FinalConcentration}) - \log(\text{InitialConcentration})}$$

Data was graphed in GraphPad Prism (<https://www.graphpad.com/scientific-software/prism/>).

2.1.3 Stabilate preparation and retrieval

For storage, 1×10^6 cells/mL mid-log phase promastigotes in culture were cryopreserved 1:1 in 70% HiFBS (Life Technologies) with 30% glycerol into 1.8 mL cryovials. Stabilates were stored in a freezer container partially filled with isopropanol, to allow slow cooling, and placed at -80°C . Parasites were kept at -80°C for short term storage and transferred to liquid nitrogen for longer storage. To retrieve cells, stabilates were defrosted at room temperature and added to 9 mL of appropriate drug-free media. After 24 h cells were passaged as a 1:1,000 dilution into media containing any required selective drugs as normal.

2.1.4 Suspended sample single cell RNA-sequencing

2.1.4.1 Sample retention and viability optimisation for single cell RNA-sequencing

To optimise sample retention and viability for low cell numbers required for single cell RNA sequencing (scRNA-seq) experiments, spinning and filtering steps for 10x Genomics single cell suspension protocols are optimised with DNA LoBind Tubes, 2.0 mL (Eppendorf). Initial harvest spins of 8250 cells in a 200 μL suspension of cHOM are resuspended in 1 mL PBS with 0.04% BSA. Ranges of centrifuge speeds and spin times were then assessed for optimal cell number retention before and after filtering with a 40 μm Flowmi™ Tip Strainer (Merck). The following buffers were assessed for washing steps, following the 10x Genomics single cell suspension protocols: HOMEM (GE Healthcare) only, 1x PBS + 0.04% BSA, 1x Dulbecco's phosphate-buffered saline (DPBS) + 0.04% BSA, and 1x PBS with 1% D-glucose (PSG) + 0.04% BSA were tested using the optimised

spinning conditions found. Viability was tested with Trypan Blue exclusion counting, as stated in section 2.1.2, with final viability of all samples found to be above 90% before running samples.

2.1.4.2 Suspended sample preparation and collection

For each suspended scRNA-seq sample, five staggered cultures were set up over two weeks; three cultures for the promastigote forms at 24, 48 and 144h for the promastigote samples, containing logarithmic growth phase and stationary phase promastigote forms, and 192 and 240h timepoints for two axenic amastigote culture forms as described in section 2.1.1. Time points were selected to best capture prospective changes across the logarithmic to stationary promastigote forms and the differentiation of promastigote forms in axenic amastigotes.

Equal numbers of wild type parasites from each of the five staggered cultures were then combined to generate one pooled sample. Samples were spun and washed following optimised conditions as described in section 2.1.4.1. 1.5 mL of the pooled culture was centrifuged, and the pelleted cells washed twice with ice-cold 1 mL PSG and 0.04% bovine serum albumin (BSA). Cells were then resuspended in approximately 500 μ L PSG plus 0.04% BSA, filtered with 40 μ m Flowmi™ Tip Strainer (Merck) and adjusted to 1500 cells/ μ L. In all steps, cells were centrifuged at 400 \times g for 10 min. In total, 15,000 cells (15 μ L) from the mixed sample were loaded into the 10x Chromium Chip by Julie Galbraith at the Glasgow Polyomics facility along with gel beads coated with unique molecular barcodes to capture individual cells and their poly-adenylated transcripts, partitioning gel and amplification master mix. The Chromium Chip was then loaded onto the 10x Chromium for amplification and library preparation by Glasgow Polyomics before next-gen sequencing. Libraries were prepared using the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 (10x Genomics). Pair-end sequencing was carried out by Glasgow Polyomics at 28x130bp and was performed with the Illumina NextSeq™ 500 platform to a depth of approximately 50,000 reads per cell. For the first WT replicate experiment, *L. mexicana* parasites were mixed 1:1 with *Trypanosoma brucei* (*T. brucei*) prepared by the same method so the heterogenous doublet rate of 8.6% could be calculated. The heterogenous doublet rate being the rate of doublets containing cells from both samples. The second replicate was split between two different sample

preparations, named Rep2_Pro for a replicate only with promastigotes, and Rep2_Ama for a replicate containing only amastigotes. For the second, Rep2_Pro, sample only 3 promastigote samples were cultured and mixed 1:1 again with *T. brucei* for a total of 9,000 cells, with a doublet rate of 6.5%. A following, Rep2_Ama, sample was run containing only axenic amastigotes timepoints, mixed 1:1 with *Trypanosoma congolense* (*T. congolense*), for a total of 6,000 cells, with a doublet rate of 4%. All samples loaded into the Chromium 10x system by Julie Galbraith of Glasgow Polyomics.

2.1.4.3 *Leishmania* read mapping and transcript counting

For the reference genome, compiling was run with Cell Ranger v3.0.2, to combine the *L. mexicana* MHOM/GT/2001/U1103 (release 55, TritrypDB) nuclear reference genome (Rogers et al., 2011) and *Leishmania braziliensis* (*L. braziliensis*) maxicircle kDNA sequence (GenBank: LR697134) (Camacho et al., 2019), as part of Briggs (2021b). For the reference genome, 3'UTR annotations were extended to increase the proportion of reads correctly assigned to annotated transcripts as described by Briggs et al., (2021b). Briefly, directly downstream of each stop codon a 2500 bp extension was added and consigned as the 3'UTR of each protein-coding gene, except where the existing 3'UTR was longer than 2500 UTR, where then the full length was kept. However, if the new assigned 3'UTR overlapped with other coding or non-coding genome features then the UTR was abbreviated to remove any potential overlap. As *L. mexicana* samples were also multiplexed with *T. brucei* parasites, a customised *T. brucei* transcriptome was generated in the same way by Dr. Emma Briggs (Rogers et al., 2011, Briggs et al., 2021b) and mapping was performed to both *L. mexicana* and *T. brucei* genomes by Dr. Emma Briggs. Using the Cell Ranger count function reads were mapped and unique reads aligned to each annotated gene were counted and matched to an individual cell barcode. Dual-species multiplets were also identified by Cell Ranger count. Multiplets (beads containing more than one cell), *T. brucei* cells and all *T. brucei* transcripts were consequently removed in individual replicates and before integrating replicates. Cell Ranger v3.0.2 (<http://software.10xgenomics.com/single-cell/overview/welcome>) was used with all default settings.

2.1.4.4 Data processing and integration of suspension samples

Data processing as Briggs et al., (2021b). Count data for individual samples Rep1 (5 combined cultures from promastigotes to axenic amastigotes), Rep2_Pro (3 combined promastigote cultures only) and Rep2_Axa (2 combined axenic amastigote cultures only) were processed separately prior to integration using the Seurat v4.0.3 (Stuart & Satija., 2019) with R v4.1.0. The percentage of transcripts encoded on the maxicircle kDNA was calculated as a percentage per cell, as a substitute for mitochondrial DNA, due to cells with an excess proportion of mitochondrial transcripts are likely to be poor quality caused by premature lysis (Ilicic et al., 2016). The percentage of transcripts per cell encoding ribosomal RNA was also calculated, as high levels of rRNA potentially indicate poor capture of polyadenylated transcripts.

For individual replicate analysis, low-quality cells in Rep1 were removed for low feature counts (<50), with potential doublet cut-offs at >3000 and UMIs >5000, high proportion of kDNA (>0.5%) and high proportion rRNA (>50%). For Rep2_Pro cells these quality control cut-offs were applied for low total RNA (<50), with potential doublet cut-offs at >4000 and UMIs >8000, high proportion of kDNA (>0.5%) and high proportion rRNA (>50%). For Rep2_Axa cells these quality control cut-offs were applied for low total RNA (<50), with potential doublet cut-offs at >4000 and UMIs >8000, high proportion of kDNA (>0.5%) and high proportion rRNA (>50%). Of note was the higher percentage of rRNA transcripts across life cycle stages, being higher in amastigote forms. These quality control (QC) plots are displayed for Rep1 in chapter 3, with following QC plots for Rep2_Pro and Rep2_Axa in Appendix I.

For integrated analysis, low-quality cells were removed by filtering for low total features (<50), high proportion of kDNA (>0.5%) and high proportion of rRNA (>50%). As for individual replicate analysis, potential doublets in integrated analysis were removed by filtering for high total RNA (>10,000) and high total unique transcripts counts (>5000). These QC plots are displayed for integrated samples in Appendix I.

Individual samples and integrated data were analysed using methods based on Briggs et al., (2021b). Briefly, each filtered sample was log₂ normalised

individually using scater v1.20.1 and Scran v1.20.1 (Lun et al., 2016). Scaled normalisation is first applied to cells so that zero counts were accounted for in raw data. Then cells are deconvolved for size factors (Lun et al., 2016). Principle components then used variable features selected by dual methods (Yip et al., 2018). Scran was used for log₂ normalisation of counts, preceded by Seurat (Stuart & Satija., 2019) which used raw counts. The top 2,000 features were used for individual sample analysis with 3,000 selected for integrated analysis. This left 1906, 1492 and 1149 for Rep1, Rep2_Pro and Rep2_Axa samples, respectively. Seurat v4.0.3 was employed for batch correction and integration (Stuart & Satija., 2019). During integration, variable features were identified, and integration anchors applied before integration. Subsequently, data was then scaled, followed by principle components identified by means of common variable features. Principle components were calculated individually by Elbow plot, and clustering resolution determined by Clustree (Zappia & Oshlack, 2018). Total RNA regressed whilst scaling data.

2.1.4.5 Cluster analysis and differential marker gene identification

Cluster and differential marker analysis as Briggs et al., (2021b). Briefly, clustering and marker gene analysis using Seurat v4.0.3 package was engaged (Stuart & Satija., 2019). With cells plotted as Uniform Manifold Approximation and Projection (UMAP) for dimension reductions (McInnes et al., 2020) and nearest neighbours were identified using 7 dimensions. Multifarious iterations were deliberated, and individual resolutions determined, based on biological expectations in samples, and resolutions that provided significant markers for every cluster, determined by Heatmap. MAST v1.18.0 (Finak et al., 2015) was selected for identifying markers over other normalisation packages, such as SCTransform (Hafemeister & Satija, 2019), however no discernible variation in top markers were delineated. Expressed marker cut-offs were chosen for >25%, with a log₂FC threshold of >0.25.

Gene ontology (GO) terms found for biological processes by TriTrypDB searches (Amos et al., 2022), with terms displaying $p < 0.05$, redundant terms banished with Reduce & Visualise Gene Ontology (REVIGO) (Supek et al., 2011) (allowed similarity = 0.5) and dot plots generated.

2.1.4.6 Scoring and labelling of life cycle and cell cycle stages

Scoring and labelling of cell cycle stages were determined as Briggs et al., (2021b). Briefly, using bulk RNA-Seq derived markers for cell cycle stages found in *T. brucei* by Archer et al. (2011), orthologues for *L. mexicana* genes were derived from TriTrypDB using BLASTp searches (Amos et al., 2021). Markers for each cell cycle stage were identified in cells with variable minimum percentages, ranging from 10-15% depending on samples. Cells found to have minimum identifiers were labelled with the highest cell cycle phase score. Scores were queried with the Meta feature function in Seurat v4.0. Subsequently, fold change for apiece phase scores divided by the mean phase score throughout each cell was next computed. Highest fold change scores were then labelled to each cell as the cell cycle phase most likely to be in by cell cycle markers. If all fold changes were found to be under 1.1 ratio of expression, they were allocated as being “non-cycling”. Using the same methodology, life cycle markers for promastigote stages were also determined, as above, using promastigote life cycle stage markers determined by bulk RNA-Sequencing in *L. infantum* (Coutinho-Abreu et al., 2020). Orthologs from the Coutinho-Abreu dataset were found for *L. mexicana* using BLASTp searches through TriTrypDB (Amos et al., 2022). Using ortholog group assignments as predicted by OrthoMCL (Li et al., 2003; Chen et al., 2006) with only one-to-one orthologues kept for further marker labelling and analysis.

2.1.4.7 Pseudotime analysis and trajectory inference

Pseudotime and trajectory analysis as in Briggs et al., (2021b). Briefly, trajectory inference was determined by means of PhateR v1.0.7 (Moon et al., 2019). The same common variable features were used as for clustering analysis for individual or integrated samples. Trajectories were determined using slingshot v2.0.0 (Street et al., 2018), with no chosen starting point for individual promastigote sample analyses (Rep2_Pro) and for integrated samples, Pro 1 cluster was chosen for a starting cluster. Cell cycle trajectory inference used a principle curve with princurve v2.1.6 (Weingessel, 2015), as a circular trajectory is expected for cell cycle. Expression patterns for genes related with advancement of the trajectory were pinpointed using the tradeSeq package v1.6.0 (Van den Berge et al., 2020), which uses a Generalised Additive Model

(GAM). Number of knots, representing segmentation of drawn trajectories, were selected by the `evaluateK()` function, where the largest number of knots providing the highest number of genes without overfitting was selected. Integrated samples used 9 knots, and individual promastigote trajectory used 7 knots. TradeSeq was also employed for differential expression analysis using the `associationTest()` function, as tradeSeq vignette (van den Berge et al., 2020). As per default analysis TradeSeq uses Wald tests, for a null hypothesis that all smoother coefficients are equivalent throughout the trajectory and calculates a p-value for genes, respectively, based on X -squared asymptotic null distribution of Wald statistics (Van den Berge et al., 2020). Clustering was used with genes found to have p-value < 0.05 and mean fold change > 2 over 100 points in inferred trajectories. Merge cut-offs were used as the default 0.95, and subset to 0.90 when testing different merging thresholds. Following gene clustering, modules were determined per sample. With trajectories containing multiple lineages differentially expressed markers for early trajectory were found with the `earlyDETest()` function in TradeSeq, in integrated sample analysis from promastigote to metacyclic and promastigote to axenic amastigote differentiation trajectories. And again, between fixed knots 3 and 4, selected due to branching of trajectories. Again, the same Wald test for the null hypothesis as described above was used (Van den Berge et al., 2020).

2.1.5 Structural bioinformatics of hypothetical proteins

Hypothetical proteins identified by pseudotime analysis were selected due to transient expression profiles along trajectories. Once identified, structural bioinformatics used sequences derived from TriTrypDB (Amos et al. 2021) (<http://tritrypdb.org/tritrypdb/>) and submitted to AlphaFold for structural predictions to be obtained (<https://alphafold.ebi.ac.uk/>) (Jumper et al., 2021). Model confidence and predicted aligned error were produced.

2.1.6 Flow cytometry

To assess cell cycle stages in promastigote forms of the scRNA-seq sample Rep2_Pro, 4×10^6 parasites were taken for each of the three promastigote timepoints. These were washed once in 1 mL of phosphate-buffered saline (PBS) and pellets resuspended in 150 μ l PBS with 350 μ l ice-cold methanol, added

dropwise, to fix cell cycle progression. Following fixing, samples were washed and resuspended in 400 μ l PBS containing 10 μ g/mL propidium iodide (PI) (ThermoFischer Scientific) and 10 μ g/mL RNase A (Sigma Aldrich). Samples were then transferred to FACS tubes through 35 μ m nylon cell strainer caps and left for 30 minutes before analysis by BD Biosciences LSR II Flow Cytometer in the PE-Cy5-Lin channel. Data was analysed on FlowJo_V10™ software (FlowJo, LLC).

2.1.7 Transforming *E. coli* for plasmid synthesis

Template pPLOT plasmid, pPLOT-mNG-blast-blast (Figure 2-1 Plasmid map for pPlot-mNG-blast-blast.) was kindly donated by the Gluenz lab for transformation and synthesis (Dean et al., 2017)., and used for tagging the proteins of interest with mNeonGreen.

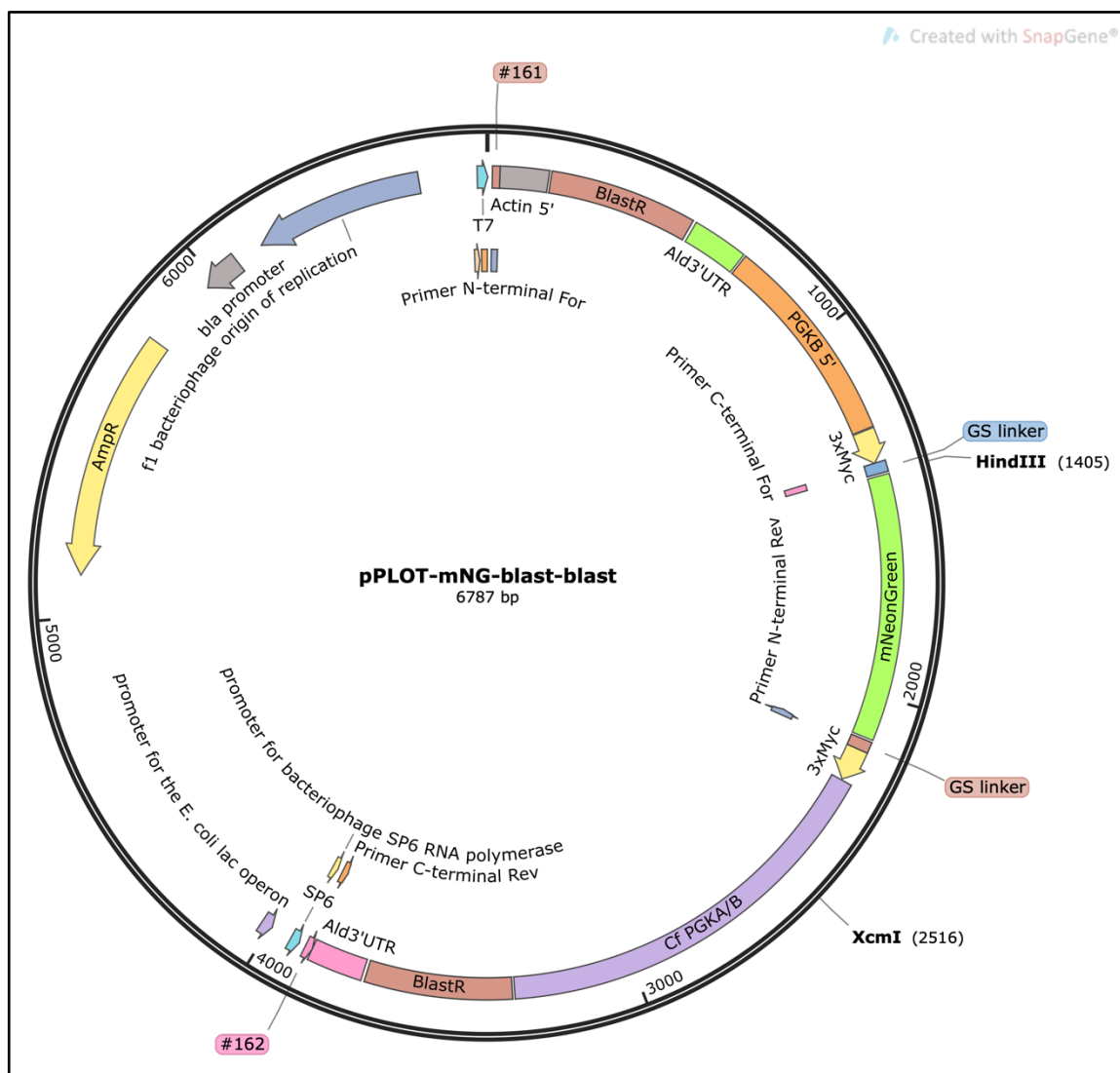


Figure 2-1 Plasmid map for pPlot-mNG-blast-blast.

Plasmid map for pPLOT-mNG-blast-blast denoting ampicillin resistance gene for plasmid selection during DH5- α *E. coli* plasmid synthesis (AmpR), Blastidicin S Hydrochloride resistance genes (BlastR), mNeonGreen tagging gene sequence, and restriction digest enzyme locations for HindIII and XcmI producing a 1111 bp sequence between cuts. Figure created with SnapGene.

To synthesize sufficient plasmid amounts for further tagging experiments the pPLOT plasmid was transformed into *Escherichia coli* (*E. coli*). 50 μ L of competent *E. coli*, MAX Efficiency® DH5- α ™ (ThermoFischer Scientific), were thawed on wet ice and 10 ng of pPLOT-mNG-blast-blast plasmid was added with 10 μ L ligation mix per 10 ng of plasmid. Cells and plasmids were then incubated on ice for 20 mins. Solutions were heated for exactly 1 minute in a 42 °C water bath, then returned to ice for 5 mins. 1 mL of sterile room temperature Luria-Bertani broth (LB-medium) was added (Gibco), and cultures were left at 37 °C with shaking at 250 rpm for 60 mins. Cells were resuspended gently, and 50 μ L placed in universal containers holding LB-medium with 100 μ g/mL ampicillin.

Universal containers were then placed in a 37 °C incubator and left overnight. Plasmids were purified by DNeasy Blood and Tissue kit (Qiagen) as per manufacture's protocol to extract DNA. Once extracted, concentrations of plasmid DNA were quantified using a NanoDrop spectrophotometer (ND1000, Thermo Scientific) with the absorbance measured at 260 nm. Samples were eluted in 50 µL of Buffer AE and stored at 4 °C. Excess bacterial stocks kept at -80 °C in glycerol for long-term storage.

2.1.8 Plasmid validation by restriction enzyme digest

Restriction enzyme digests were performed on purified plasmid samples to ensure validity. 100 ng of pPLOT-mNG-blast-blast plasmid DNA was incubated with 1 µL each of HindIII and XcmI restriction enzymes, and 10x rCutSmart buffer (New England Biolabs) for double and single cut digests then brought up to 20 µL with nuclease free water. Samples were analysed by agarose gel electrophoresis as in section 2.1.9

2.1.9 Agarose gel electrophoresis

DNA fragments were separated with 1% or 2% (w/v) UltraPure™ Agarose (Life Technologies) in gels made with 1x TAE buffer (40 mM Tris base, 19 mM acetic acid, 1mM EDTA) and SYBR® Safe DNA Gel Stain (Life Technologies), added at a 1:10,000 dilution. Loading dye was added to DNA samples before loading alongside 7 µL 1 Kb Plus ladder (Invitrogen™, Life Technologies). Gels were run for 30 - 60 min in 1x TAE buffer at 100 - 120 v.

2.1.10 High-throughput CRISPR-Cas9 tagging

CRISPR-Cas9 mediated tagging of LmexCAS9T7 cell lines was performed using the high-throughput 96-well plate-based transfection assay as outlined by Beneke and Gluenz (2019), and described in this following section (2.1.10). Of the 96 proteins selected for tagging, 6 were life cycle stage markers and the remaining 90 were hypothetical proteins selected for their transient expression across pseudotime analysis of inferred trajectories, and those which were found to have only a single 3' sgRNA target during CRISPR-Cas9 tagging primer design.

2.1.10.1 Hypothetical protein primer design for mNeonGreen C-terminal tagging

Selected proteins of interest had Gene IDs from TritypDB.org entered into www.leishgedit.net and primers were designed for pPLOT plasmids with the tagging gene edit strategy selected. Primers designed via www.leishedit.net contain a T7 promoter sequence, 20 nt sgRNA target site which must be next to a protospacer adjacent motif (PAM) “NGG” at the target locus and a sequence complementary to the sgRNA backbone (G00 primer) (Figure 2-2 G00 scaffold sgRNA backbone sequence.).

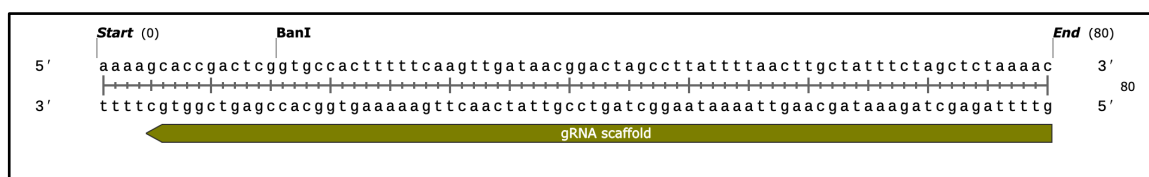


Figure 2-2 G00 scaffold sgRNA backbone sequence.

The G00 scaffold sgRNA backbone sequence which is an oligonucleotide used as template DNA for the amplification of all sgRNA in these experiments found in this thesis. Figure created with SnapGene.

Design of donor DNA primers via www.leishedit.net include a 30 nucleotide sequence with homology flanks at either end of the donor DNA, aligning to the target locus sequence next to the sgRNA target and sequences binding to the pPLOT template plasmid (Figure 2-1 Plasmid map for pPlot-mNG-blast-blast.).

Once validated, the primer list was exported and ordered via plate-based oligo manufacturing (ThermoFisher Scientific) (See Appendix II for primer list).

2.1.10.2 Polymerase chain reaction amplification of sgRNA templates

For gene tagging using the CRISPR-Cas9 high-throughput system (Beneke and Gluenz, 2019) preparation of one polymerase chain reaction (PCR) sgRNA template product for a C-terminal tag, placed the double-stranded break at the 3' end, was required per tagged cell line. 100 μ M sgRNA primer stocks were diluted to 4 μ M by placing 4 μ L of each 100 μ M primer stock and 96 μ L ddH₂O into a certified thin wall 96 x 0.2mL PCR plates (StarLab). 4 μ M primer dilutions were mixed well and 10 μ L of this dilution added to a new 96-well plate. The plate was briefly spun down to ensure contents were collected at the bottom of wells. The plate was then frozen at -80°C for 30 minutes in a UltraCruz® Freeze Block (Santa Cruz).

In parallel with the freezing process, a PCR master mix was made as follows: For each reaction 0.4 μL of 100 μM sgRNA backbone G00 primer, 0.4 μL of 10 mM dNTP mix, 2 μL of 10x reaction buffer supplemented with 15 mM MgCl_2 and 7 μL of ddH₂O was added. With 0.2 μL ExpandTM High Fidelity polymerase (Merck) added last. PCR was carried out with the following thermocycler programme: 98 °C for 30 seconds, 35 cycles of denaturation at 98 °C for 10 seconds, annealing at 60 °C for 30 seconds and elongation at 72 °C for 15 seconds and ending with a final extension for 10 minutes at 72 °C. Thermocycler used was a AC196 (PCRMax).

As soon as 98 °C was reached, the program was paused, and primer dilutions taken from the freezer. The PCR master mix was added on top of primer dilutions by pipetting the entire master mix into a reagent reservoir and adding 10 μL of the PCR master mix from the reservoir to each primer dilution by using a multichannel pipette, working quickly before reactions reach room temperature. The plate was briefly centrifuged a final time, sealed with ThermalSeal ATM sealing films (Merck), and placed in the hot PCR block before restarting the PCR program. Following completion of the PCR programme, reactions were held at 4 °C until removed from thermocycler.

To ensure all PCR reactions had worked, each product was visualised on an agarose gel, separated by electrophoreses and imaged as section 2.1.9.

2.1.10.3 PCR amplification of donor DNA

Gene tagging required preparation of one PCR product encoding the mNeonGreen tag and Blastocidin drug resistance marker. Dilution of 100 μM stocks of donor PCR primers to 10 μM by adding 5 μL of 100 μM forward primer stock, 5 μL of corresponding 100 μM reverse primer stock and 40 μL ddH₂O to the corresponding wells of a thin walled 96 x 0.2mL PCR plates (StarLab). 10 μM primer dilutions were mixed well and 8 μL of this dilution pipetted into a new thin walled 96-well plate and spun to collect primers at the bottom of wells. The plate was frozen at -80 °C in a UltraCruz® Freeze Block (Santa Cruz). During the freezing process, the required volume of PCR master mix is produced. For each reaction 0.5 μL of 30 ng/ μL plasmid template, 0.8 μL of 10 mM dNTP mix, 1.2 μL of 100% (v/v) DMSO, 3 μL of 25 mM MgCl_2 solution, 4 μL of 10x reaction buffer

supplemented with 15 mM MgCl₂ and 22.1 µL of ddH₂O is mixed well. 0.4 µL High-fidelity polymerase (Merck) was added last. The following program on a PCR thermocycler was ran: denaturation 94 °C for 5 minutes and then 40x repeated cycles of denaturation at 94 °C 30 seconds, annealing 65 °C 30 seconds, elongation 72 °C 2 minutes and 15 seconds with a final extension of step of 72 °C for 7 minutes.

When the initial 94 °C was reached, the program is paused, and primer dilutions removed from the -80 °C freezer. The PCR master mix was added on top of primer dilutions by pipetting the entire master mix into a reagent reservoir and adding 32 µL of the PCR master mix from the reservoir to each primer dilution by using a multichannel pipette, working quickly before reactions reach room temperature. The plate was then briefly centrifuged a final time, sealed with ThermalSeal A™ sealing films (Merck), and placed in the hot PCR block before restarting the PCR program. Following completion of the PCR programme, reactions were held at 4 °C until removed from thermocycler.

To ensure all PCR reactions had worked, each reaction is run on an agarose gel, separated by electrophoreses and imaged as section 2.1.9.

2.1.10.4 “Magic” M199 media for transfection recovery

To aid in transfection recovery efficiency in 24-well plates following the forthcoming electroporation a supplemented M199 media, coined “Magic” M199, was used. For 500 mL, 100 mL 5× stock M199 including 1.10 g NaHCO₃, was mixed with 100 mL FBS, 20 mL 1 M HEPES.HCl at pH 7.4, 10 mL 5 mM Adenine hemisulphate in water, 0.5 mL 2.5 mg/mL Haemin in water, 2 mL 0.3 mg/mL Biopterin in DMSO, and brought up to 500 mL with 267.5 mL dH₂O. “Magic” M199 media was sterilised and tested for contamination before conducting transfections, as below.

2.1.10.5 “Magic” M199 preparation and testing

4 days prior to the transfection day, fresh “Magic” M199 medium was tested as follows: Firstly, to check for contamination a 5 mL aliquot of “Magic” M199 was transferred to a vented cell culture flask in sterile conditions, without any antibiotics, and incubated at 27 °C, 5% CO₂. After 2 - 3 days the media was

examined under microscope to for bacterial or fungal contamination. Secondly, for *Leishmania* growth. Again, in sterile conditions, a 5 mL aliquot of “Magic” M199 was transferred to a vented cell culture flask without any antibiotics and warmed to 27° C with 5% CO₂. A recently defrosted culture (<6 passages) of *L. mexicana* Cas9 T7 M was seeded to the media, without any antibiotics, at a density of 2x10⁵ cells/mL. Cell density was counted by Trypan Blue exclusion every 24 h up to at least 72 h, as detailed in section 2.1.1. Cell growth was plotted and the fresh “Magic” M199 media was only used if the cells showed the expected doubling time of 6.5-7.5 h. Data was graphed in GraphPad Prism (<https://www.graphpad.com/scientific-software/prism/>).

2.1.10.6 Transfection of *Leishmania mexicana* promastigotes

In preparation for transfections, an exponentially growing culture of the *Leishmania mexicana* promastigote cell line LmexCas9T7 expressing Cas9 and T7 RNAP was cultivated. This parental cell line was previously cultured with selection drugs for Cas9 and T7 RNAP construct over several passages until the day of transfection, as described in section 2.1.1. For each transfection 1 x 10⁷ cells were required.

2.1.10.7 Pooling of PCR reactions

For each new cell line generated, sgRNA PCR reactions were pooled together with the corresponding donor PCR reactions to combine all PCR reactions for the same target gene in a single well of a plate via multichannel pipette. For tagging transfections the yield is approximately 50 µL total PCR reaction, being 40 µL donor DNA PCR plus 20 µL sgRNA DNA template PCR, accounting for volume loaded for gel electrophoresis and sample lost to evaporation or handling. Combined PCR products were pooled and sealed with adhesive sealing sheets (ThermoFisher). Pooled PCR products were then heat sterilized at 94° C for 5 minutes in preparation for transfection protocols.

2.1.10.8 96-well plate transfection

1 x 10⁹ LmexCas9T7 cells expressing Cas9 and T7 RNAP (at least 1 x 10⁷ cells per reaction) were cultivated and centrifuged at 800g for 15 min. Whilst centrifugation was running, four 24-well plates were labelled and each well

filled with 1 mL of warm “Magic” M199 medium from section 2.1.10.4. Fresh transfection buffer was prepared as follows: 2 mL CaCl₂, 6.5 mL modified 3x Tb-BSF and 6.5 mL ddH₂O and sterilised by passaging through a 0.22 µm filter.

Following centrifugation of cells, supernatant was removed, and cells resuspend in 3 mL transfection buffer, then centrifuged again as above. Whilst in the second spin, the BTX ECM 830 Electroporation System was programmed with the following settings: 1500 V, 24 pulses, 2 counted pulses, 500 ms interval, unipolar, 100 µs. Pooled and heat-sterilized PCR products (from section 2.1.10.7) were transferred into 96-Well disposable electroporation plates with 4 mm gap and 200 µL wells (BTX). Following the second spin any remaining supernatant was removed and cells resuspend in 7.8 mL transfection buffer. This cell suspension with transfection buffer was then transferred into a sterile reservoir and 150 µL of cell suspension pipetted into each well containing the pooled PCR products. Without pipetting up and down, the cell suspension was added to wells with the pipette tip touching the wall on the bottom corner. The plates were sealed with adhesive sealing sheets and transferred to the BTX unit.

Electroporation pulses were applied using the settings prepared above. After electroporation, cells were transferred via multichannel pipette to 24-well plates containing warmed “Magic” M199 media. Remaining wells were rinsed with medium to flush out any remaining cells. Plates were then incubated in a 27°C, 5% CO₂ incubator and left for 8 - 16 h, then the required selection drugs for the repair cassettes were added and cells incubated until drug resistant populations emerge, taking approximately 2 weeks. Following selection, tagged cell lines were then grown in cHOM at 25°C in sealed plates for comparison to expression levels of the WT *L. mexicana* scRNA-seq analysis, removing potential artifacts due to variations in growth conditions.

2.1.11 Fluorescence microscopy

L. mexicana samples were prepared by using transfected cell lines expressing mNeonGreen fluorescent fusion proteins and were imaged live, adhered to glass slides. Parasites were harvested from culture by centrifugation at 800g for 5 min and washed three times in PBS with Hoechst 33342 at 10 µg/mL in the first wash. The cells were resuspended in 10 µL PBS, placed on a poly-lysine coated microscope slide, then a coverslip was applied, and the cells were immediately

imaged with a Leica DMI8 S platform live cell fluorescence microscope (Leica) 20x/0.75 objective and a 63x/1.40 oil objective at the ambient temperature of 25 - 28°C. Data was graphed in GraphPad Prism (<https://www.graphpad.com/scientific-software/prism/>).

2.1.12 THP-1 culture and differentiation

The human acute leukaemia monocyte cell line (THP-1) (Tsuchiya et al., 1980), was cultivated in RPMI-1640 media with 10% HiFBS (Life Technologies), 25 mM HEPES, 1% L-glutamine at 37°C with 5% CO₂ in vented flasks. Routine passages were conducted every 2 - 3 days, maintaining cell at a density of 1 - 3 x10⁵ cells/mL to prevent cell count from exceeding 1 x 10⁶ cell/mL. Cells were kept for a maximum of 15 subculture dilution cycles.

THP-1 cells were differentiated into macrophages with 100 ng/mL of phorbol 12-myristate 13-acetate (Sigma) for 48 h at 37°C, 5% CO₂. Differentiated macrophages cells are adherent and were seeded in Corning™ Costar™ 24-well Clear TC-treated Multiple Well Plates (Thermo Fisher Scientific) with a round 12 mm cover slip placed in the bottom, at a confluence of 3 x 10⁵ cells per well.

2.1.13 CFSE infection assays

To access infection rates using fluorescence microscopy, stationary promastigote cells were stained with the cytosolic fluorescent stain carboxyfluorescein succinimidyl ester (CFSE) cell division tracking dye (ThermoFisher) by first washing in PBS with 2% FBS before staining with CFSE at 1µM concentration and incubated at 37°C for 10 minutes exactly. Staining was quenched using ice cold PBS with 20% FBS. *Leishmania* were then washed and counted to give the 10:1 promastigote:macrophage ratio.

L. mexicana promastigotes at stationary phase were added to the plates containing coverslips (10:1 promastigotes:macrophage ratio) for 4 h. Cultures were then washed several times until the majority of extracellular promastigotes have been removed. Infections were incubated for 1 day and 3 days at 37°C in a 5% CO₂ environment. Before microscopy, assays were washed three times in PBS with Hoechst 33342 at 10 µg/mL in the first wash. Infected macrophages were

examined using a Leica DMI8 S platform live cell fluorescence microscope (Leica) with a 40x/0.65 objective and a 63x/1.40 oil objective. The number of parasites per 100 macrophages was determined by counting 300 cells in each of the triplicate experiments per round of infection. Image analysis was carried out with Fiji ImageJ v2.

2.1.14 Adherent sample single cell RNA-sequencing

Adherent macrophage samples infected with *L. mexicana* at 4h and 24 h timepoints and uninfected macrophage samples were prepared and analysed as suspended *L. mexicana* samples in section 2.1.4 with the following exceptions described below.

2.1.14.1 Adherent single cell RNA-sequencing sample preparation and collection

Two adherent macrophage samples were cultivated, differentiated, and staggered to give 4 h and 24 h timepoints as in section 2.1.12. Differentiated macrophages were grown in 100 mm Petri dishes and infected with stationary promastigotes as section 2.1.13, excluding CFSE staining steps, and seeded at a 10:1 promastigotes:macrophage ratio as determined by infection assays. Macrophages infected with *L. mexicana* and an uninfected macrophage control were prepared for scRNA-seq as in section 2.1.4.2 with the following additions: Adherent macrophages had culture medium removed and 2 mL warm 0.25% trypsin-EDTA solution washed over the surface to remove excess media that may inhibit the enzymatic action of trypsin. Once washed, plates were incubated with 5 mL warmed 0.25% Trypsin-EDTA solution to cover the cell layer and incubated for 10 - 15 minutes at 37°C. Cell dissociation was periodically checked and when sufficient detachment was reached, 5 mL of culture media was added to stop digestion. The cell containing solution was then pipetted gently 5 - 10 times to break up potential clumps of cells. Once the cell suspension was transferred to a 50 mL conical tube all washing steps were then followed sample preparation as per the 10x single cell suspension protocol for scRNA-seq.

2.1.14.2 Human macrophage read mapping and transcript counting

Mapping for human and leishmania samples were compiled as section 2.1.4.3 using the same *L. mexicana* MHOM/GT/2001/U1103 (release 55, TritrypDB) nuclear reference genome combined with maxicircle kDNA sequence (GenBank: LR697134) (Camacho et al., 2019) and merged with the standard *Homo sapien* (*H. sapien*) reference genome (GRCh38) from Cell Ranger v3.0.2. Mapping was performed by, Dr. Pawel Herzyk of Glasgow Polyomics, to both *H. sapien* and *L. mexicana* genomes and reads mapped to unique reads aligned to each annotated gene were counted and assigned to a cell barcode with the Cell Ranger count function. Cell Ranger v3.0.2 (<http://software.10xgenomics.com/single-cell/overview/welcome>) was again used with all default settings.

2.1.14.3 Data processing and integration of adherent samples

Count data for the collective infected macrophage with *L. mexicana* sample (INF) and uninfected macrophage control (CTL) were processed separately prior to integration of only the INF *L. mexicana* reads (AMA) to the integrated object of Rep1, Rep2_Pro and WT3. For quality control, standard mitochondrial cut-offs were used for human samples while the same quality control parameters were used for infecting *L. mexicana* as section 2.1.4.4.

For individual sample analysis, low-quality cells in AMA were removed for low total features (<50), high proportion of kDNA (>0.5%) and high proportion rRNA (>50%). With a filtering cut-off for high total RNA of >10,000 and high total unique transcripts counts of >4000. For *H. sapien* INF (INFhs) cells were removed for low total RNA (<100), low unique transcripts (<100), high proportion of mitochondrial DNA (>15%) (D. Osorio & Cai, 2021). Filtering for multiplets used cut-offs for high total RNA of >80000 and high total unique transcripts counts of >7500. For the uninfected *H. sapien* control sample (CTL) cells were removed for low total RNA (<100), low unique transcripts (<100), high proportion of mitochondrial DNA (>15%) (Osorio et al., 2021). With a filtering cut-off for high total RNA to remove potential doublets of >100000 and high total unique transcripts counts of >9000.

For integrated analysis of INF and CTL, low-quality cells were removed by filtering for low total RNA (<50), high proportion of kDNA (>0.5%), high proportion of mitochondrial DNA (>15%), and high proportion of rRNA (>50%). As for individual replicate analysis, potential doublets in integrated analysis were removed by filtering for high total RNA (>100000) and high total unique transcripts counts (>9000). For detailed QC plots, see Appendix I.

For integrated analysis of AMA to the integrated object containing *L. mexicana* samples Rep1, Rep2_Pro, and WT3, low-quality cells were removed by filtering for low total RNA (<50), high proportion of kDNA (>0.5%) and high proportion of rRNA (>50%). As for individual replicate analysis, potential doublets in integrated analysis were removed by filtering for high total RNA (>3000) and high total unique transcripts counts (>2000). For detailed QC plots, see Appendix I.

Integrated sample analysis follows normalisation as section 2.1.4.4 with AMA genes selected by combined Scran (Lun et al., 2016) and (Seurat Stuart & Satija., 2019) normalisation, totalling 1021 genes.

For integration and batch-correction of WT replicate samples with AMA, the Seurat v4.0.3 package was used (Stuart & Satija., 2019). Common variable features and integration anchors were identified, data for all genes integrated and scaled before the PCs were calculated using the common variable features. The first 8 PC dimensions each contributed >0.1% of additional variance and were used to select anchors and integrate data. The effect of total RNA per cell was regressed when scaling data.

2.1.14.4 Cluster analysis and differential marker gene identification

Clustering and marker gene analysis follows the same pipeline as suspension *L. mexicana* sample cell analysis, as detailed in section 2.1.4.5.

**Chapter 3 Life and cell cycle progression
analysis by single cell RNA-sequencing in
*Leishmania mexicana***

3.1 Introduction

For a vector-borne pathogen, survival depends upon adapting to changing conditions in the host and vector and surviving host immune attack. Currently, methodologies establishing the drivers of transcriptomic dynamics for kinetoplastid parasites are still in their infancy, although recent advances in *Trypanosoma brucei* have provided great insights using single-cell RNA-Sequencing (scRNA-seq) (Briggs et al., 2021b). Additionally, in *Leishmania*, as discussed previously, recent studies have employed scRNA-seq to examine hybridisation *in vitro* following stress conditions (Louradour et al., 2022), and further studies examined *in vivo* transcriptional analysis of mice infected with *Leishmania* by Venugopal et al. (2022) and in pre-print by Karagiannis et al. (preprint). Given the complexity of the *Leishmania* lifecycle, with potential for numerous cryptic stages in host and vector, and concomitant potential for targeted chemotherapy, there is also a need to leverage scRNA-seq to undertake a more detailed appraisal. Currently, no scRNA-seq investigations have been made into life cycle or cell cycle progression in *Leishmania* species, and nor have any such studies examined *Leishmania mexicana* (*L. mexicana*). Here, approaches were developed for transcriptomic analysis using scRNA-seq with a view to linking transcriptional changes and developmental changes seen in life cycle and cell cycle progression. Following changes in growth conditions required for the progression of the life cycle of *L. mexicana* here, transcriptomes were captured across five timepoints to identify the transcriptional changes associated with these life and cell cycle developmental processes.

3.1.1 Life and cell cycle progression analysis in *Leishmania* using single cell RNA-sequencing

Previous methods for analysis of changes that occur during differentiation and progression through the life cycle stages in *Leishmania* have relied on comparative life cycle bulk transcriptomics, metabolomics, phosphoproteomics and proteomics. Altogether, this preceding body of work has highlighted prominent changes as the parasite progresses through *in vitro* life cycle stages and provided several markers for each life cycle stage (Tsigankov et al., 2013; Fiebig et al., 2015; Inbar et al., 2017; Kaur et al., 2021). However, despite these

investigations, the full picture remains unclear as to the activation, control, and the timing of these transitional events required for life cycle progression. A benefit to the study of *Leishmania* life cycle progression is that large proportions of its life cycle are readily reproducible *in vitro* (Bates et al., 1992). Vector borne promastigote stages are readily cultured, with duration and density of growth able to reproduce differentiation from replicative procyclic promastigote forms to infective stationary phase metacyclic promastigotes (Gossage et al., 2003). Indeed, several morphologically and transcriptionally distinct promastigote forms have been described *in vitro*, such as replicating leptomonad promastigote forms and non-cycling nectomonad promastigote forms, but how these distinct life cycle stages are related and compare in their timing of gene expression is still unclear (Coutinho-Abreu *et al.*, 2020). Additionally, modification of culture conditions (e.g., increasing temperature and acidity) also leads to axenic amastigote forms, representative of the stage found within professional phagocytotic cells in mammalian hosts (Kima, 2007). One of the more powerful tools of scRNA-seq analysis is the clustering of similar cells within a sample to discover rare cell populations and heterogeneity within a sample not achievable by bulk RNA-sequencing analysis (Reid et al., 2018; Yeo et al., 2020).

By adding to the current knowledge of transcriptomic developments in these parasites by being the first to employ scRNA-seq analysis of life and cell cycle progression in *Leishmania* we believe scRNA-seq is the ideal approach to examine the above questions: i.e. it is capable of discriminating distinct cell populations in a mixture, such as the different promastigote forms and different life cycle forms, and can infer a differentiation trajectory from unsynchronised cells, such as life and cell cycle progression.

3.2 Aims

- To use single-cell transcriptomic data to cluster individual parasites into distinct life cycle forms and identify genetic markers characteristic of different life cycle stages.
- To employ trajectory inference techniques to order cell types by transcriptional change, allowing genes with differential expression patterns to be identified in both life cycle and cell cycle progression.

- To discover genes directing life cycle and cell cycle progression of *L. mexicana* from scRNA-Seq analysis.

3.3 Results

3.3.1 Single cell RNA-sequencing sample culture and preparation

To reveal the timing and patterns associated with developmental gene expression changes, *L. mexicana* were grown from 1×10^6 cells per mL concentration for 7 days in non-vented cap flasks at 25 °C and pH7.5 with samples taken at 24 h, 48 h and 144 h, representing log phase (promastigote forms) and stationary phase (containing metacyclic) populations (Figure 3-1). In parallel, axenic amastigotes were induced from stationary phase cultures at a concentration of 1×10^6 cells per mL cultured with supplemented Schneider's *Drosophila* medium (Bates *et al.*, 1992), pH adjusted to 5.5, and grown at 34 °C in vented cap flasks and samples taken at 24 h and 72 h timepoints (Figure 3-1). Samples were taken from the start of axenic culture to attempt to capture cells in the process of differentiation from metacyclic to axenic amastigotes, and the following sample taken to capture replicating axenic amastigote forms. 3,000 cells were taken at each time point and were mixed together; with a predicted cell loss of around 50% through the 10X Chromium system, we expected to recover 1,500 cells for each timepoint and 7,500 cells in total across the five timepoints. These samples were run in two replicates, the first replicate (Rep1) containing all five timepoints (*L. mexicana* spanning promastigotes to axenic amastigotes), and the second replicate analysed as two separate datasets, one containing only the three promastigote timepoints (Rep2_Pro) and the other only the two axenic amastigote timepoints (Rep2_Axa). Replicates Rep1 and Rep2_Pro were mixed with *T. brucei* samples at a one-to-one ratio, so that estimation of multiplets in quality control pipelines further downstream could be made (Briggs *et al.*, 2021b). Rep2_Axa was mixed with a *T. congolense* sample (provided by Dr. Emma Briggs).

All samples were loaded into the 10X Chromium microfluidic chip by Julie Galbraith of Glasgow Polyomics, who also ran the Illumina NextSeq 500 sequencing (Haque *et al.*, 2017). The reads generated were mapped to the *Leishmania mexicana*_MHOM_GT_2001_U1103 reference genome generated by

Rogers et al., (2011) and *Leishmania braziliensis* (*L. braziliensis*) maxicircle kDNA sequence (GenBank: LR697134) (Camacho et al., 2019) by Dr. Emma Briggs, as described in section 2.1.4.3. (Briggs et al., 2021b).

Each replicate of scRNA-seq will be explored here individually to assess quality control parameters, cell population clustering, markers present in life cycle stages, and cell cycle labelling in each replicate. Replicates are then integrated for further life and cell cycle analysis, and pseudotime analysis is employed to order transcriptional events through life cycle progression, as explained in Chapter 1.

3.3.2 Single cell RNA-sequencing analysis for first *L. mexicana* replicate

Following mapping of reads generated for the Rep1 sample, reads are assessed by Cell Ranger outputs, run and generated by Dr. Emma Briggs. Sample data with timepoints shown for each combined sample is shown below in Figure 3-1.

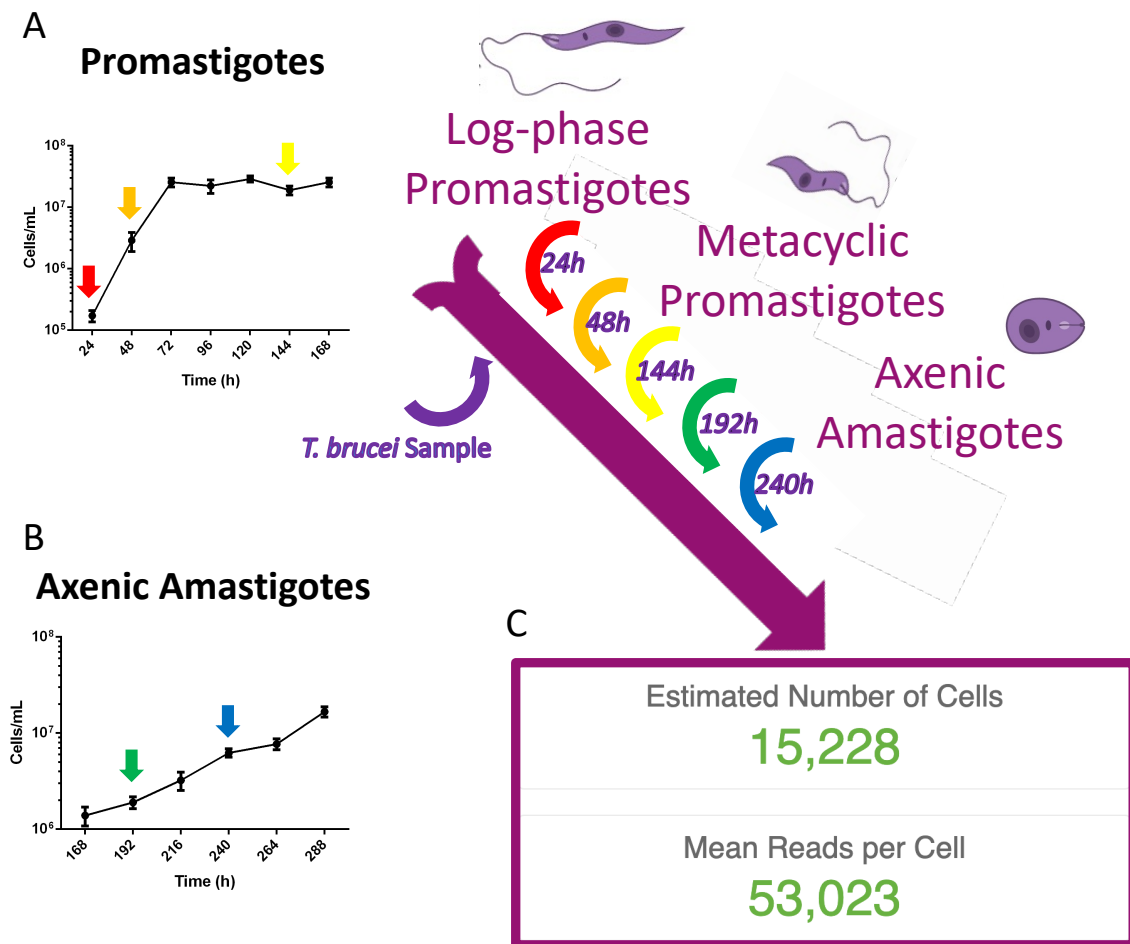


Figure 3-1 Experimental sample collection and sample information for integrated single cell RNA-Sequencing.

Experimental protocol for first replicate *Leishmania* sample run in single cell RNA-Sequencing over 5 timepoints, at 24, 48, 144, 192, and 240h, across promastigote to axenic amastigote life cycle progression. Replicate Rep1 contained all 5 timepoint samples with 3000 cells taken from each timepoint containing a mixture of **A**) log-phase and stationary phase promastigotes, the latter to enrich for metacyclic forms, and **B**) axenic amastigotes. *Leishmania mexicana* samples were then mixed with *Trypanosoma brucei* cells to estimate multiplerts in sequencing sample. **C**) In total, 15,228 cells were sequenced with mean reads per cell of 53,023.

Here, the combined samples of *L. mexicana* and *T. brucei* contained an estimated 15,228 cells with a combined mean read per cell of 53,023.

3.3.2.1 Quality control filtering of single cell RNA-sequencing for first replicate

Following the generation of reads, quality control of the samples is utilised to limit the impact of noise generated due to biological and technological variations (Kolodziejczyk et al., 2015). Unique reads were mapped and aligned to gene annotations in the custom reference nuclear genome and maxi-circle kinetoplast genome. Counts were then assigned to cell barcodes within single-cell Gel Beads-in-emulsion (GEMs) using the Cell Ranger count function. Cell Ranger v3.0.2 (<http://software.10xgenomics.com/single->

cell/overview/welcome) was run with default settings. Dual-species multipliers were identified by Cell Ranger count, as shown in Figure 3-2 A).

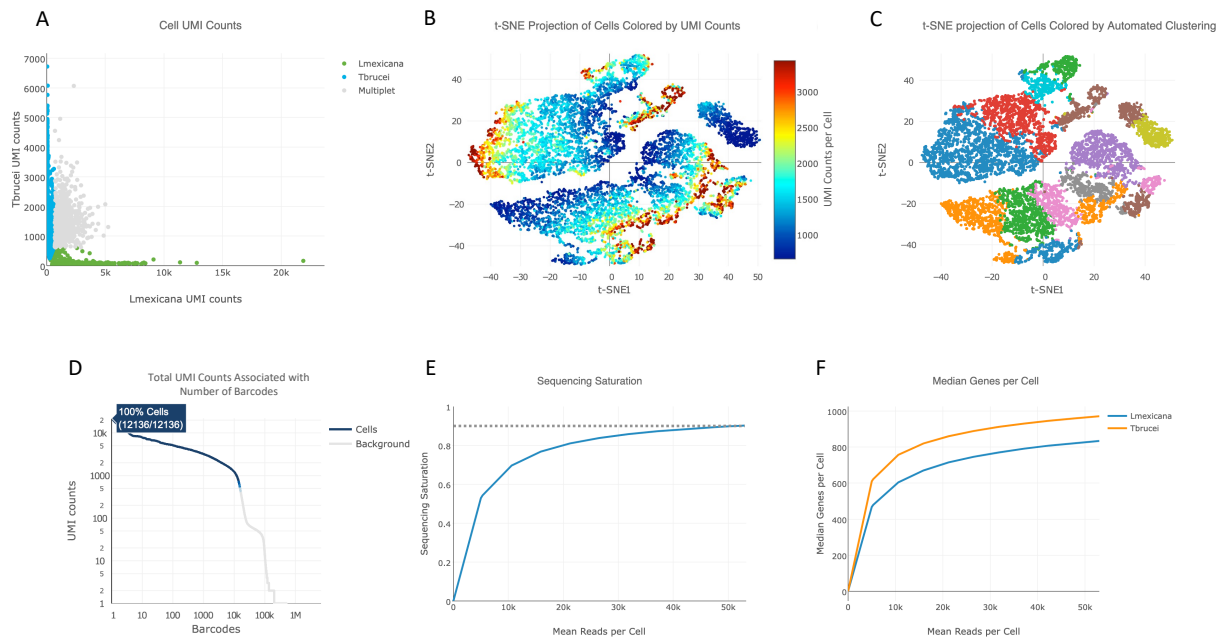


Figure 3-2 Cell Ranger summaries of sequenced Rep1 sample.

Web summary plots of sequenced Rep1 sample containing *Leishmania mexicana* (*L. mexicana*) and *Trypanosoma brucei* (*T. brucei*). Samples run by Julie Galbraith and plots generated by Dr. Emma Briggs in Cell Ranger (v3.0.2). **A)** Cell Unique Molecular Identifier (UMI) counts for each sample, *L. mexicana* in green, *T. brucei* in blue and multipliers in grey. **B)** t-distributed Stochastic Neighbour Embedding (t-SNE) plot of random 10,000 subset of cells in sample coloured by UMI counts. **C)** t-SNE plot of automated cell clustering in Cell Ranger clustered by similar expression profiles. **D)** Total UMI counts in mixed sample against barcoded counts of individual transcripts in mixed sample. **E)** Sequencing saturation of downsampled sequencing depth against mean reads per cell, with reasonably approximated saturation value point indicated by dotted line. **F)** Plot for median genes per cell against mean reads per cell for each sample mapped to their respective transcriptome for each species. Mapping run and Cell Ranger plots generated by Dr. Emma Briggs.

In Figure 3-2 A), dots are coloured by the identity of inferred cells in the GEM related with each barcode: green for *L. mexicana*, blue for *T. brucei*, and multipliers in grey. Each dot represents a cell-barcode, where the x-axis represents total Unique Molecular Identifier (UMI) counts for *L. mexicana* barcodes mapped to the *L. mexicana* transcriptome, and the y-axis the same for *T. brucei*. UMIs being molecular barcoded oligonucleotides with individual 12 base oligonucleotides, as discussed in Section 1.5.0. A multiplier is defined as a GEM inferred to have encapsulated >1 cell or a barcode sequence that was shared by another or multiple other single-cell GEMs. 15,228 GEMs were present in this sample. 8% were inter-species multipliers with more than one barcode, and a mean UMI count purity of 95.8%.

Figure 3-2 B shows the total UMI counts for each cell-barcode. It is inferred that cells with higher UMI counts likely have greater RNA content than cells with fewer UMI counts. The axes represent 2-dimensional embedding produced by the t-SNE algorithm. And cells with similar gene expression profiles are placed together, with cells having larger differences in gene expression profiles placed further away from one another. These plots are limited to a random subset of 10,000 cells by Cell Ranger and confirm captured UMIs for further analysis.

In Figure 3-2 C the same t-SNE projection is coloured by Cell Ranger automated clustering algorithm, where cells are clustered by colour to those with similar expression profiles. This plot is also limited in k-means up to $k=20$, where k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

In Figure 3-2 D barcodes in the mixed sample are plotted against total UMIs. Here, cells are represented by the blue line on the curve, which peaked around 15,000 cells, 8,469 of which belonged to *L. mexicana* cells. Fraction of reads for the *L. mexicana* sample was estimated at 42%. This shows a suitable number of UMIs are captured per number of barcodes.

In Figure 3-2 E the sequencing saturation is shown the number of UMI detected as sequencing depth is increased. Saturation is an indicator of the library complexity and indicates when increased sequencing depth is unlikely to identify further UMIs in the sample. Here increasing sequencing over 40,000 reads per cells results in fewer novel UMIs. Once sequencing to $> 50,000$ reads was completed, 90% of estimated UMIs in the sample were recovered.

In Figure 3-2 F the median gene count per cell for each mapped transcriptome to their relative species were plotted, representing the median gene count per cell as a function of the sequencing depth in mean reads per cell, up to the observed sequencing depth. The slope of each curve indicates where the upper boundary for benefit in increasing sequencing beyond 50,000 reads as been reached. Median genes per cell for the *L. mexicana* sample was 834, with total genes detected across all cells being 8,303 and median UMI counts per cell was 1,359.

Output data from the Cell Ranger pipeline were then read using the `Read10X()` function, giving an output of a matrix for UMI and counts. This matrix represents the number of genes, termed features in 10X analysis, in each row in each cell in the columns. This count matrix is used to generate a Seurat object (Satija et al., 2015), which contains count data in the matrix and metadata for each cell, such as sample or origin and total number of UMI. Seurat being a R package intended for quality control, analysis, and investigation of scRNA-seq data (Satija et al., 2015). Multiplets, individual *T. brucei* cells, and all *T. brucei* transcripts were then removed from this Seurat object before further pre-processing quality control analysis. The steps taken to filter cells based on quality control metrics are described below, in Figure 3-3. The quality control filters considered were: number of unique features, UMIs, kDNA (as a replacement for mitochondrial DNA typically used in human and other samples), and ribosomal RNA (rRNA) contaminates per cell.

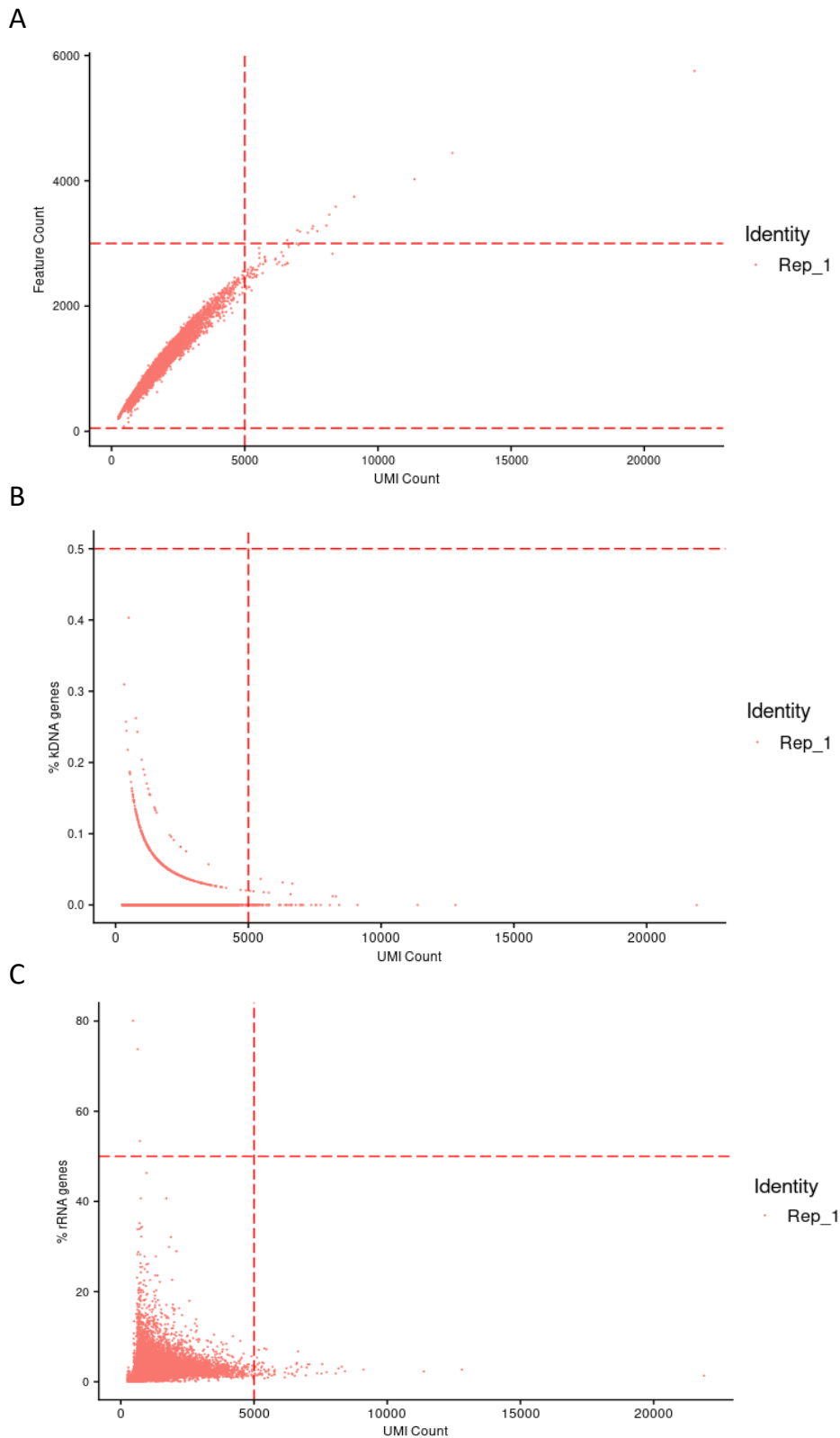


Figure 3-3 Quality control and filtering of transcriptomes in first biological replicate.

Scatter plots of quality control and filtering cut-offs. Each dot representing one captured transcriptome. Unique Molecular Identifiers (UMI) are plotted against **A**) gene counts (features) with UMI counts <5000, feature counts >50 and <3000 to exclude for multipliants, **B**) percentage of kDNA maxi circle genome included with mapping reads < 0.5% to exclude lysed or apoptotic cells and **C**) percentage of ribosomal RNA (rRNA) <50% contamination in sample. Red dashed lines indicate cut-offs used to filter cells per experiment.

In Figure 3-3 A the number of unique features was set at cut-offs of 3,000 (upper limit) and 50 (lower limit), to exclude cell multiplets with unusually high gene counts and empty droplets, respectively. This lower limit is lower than typically used to account for possible stationary phase growth typical of some life cycle stages in *Trypanosomatida* (Christiano et al., 2017; Coutinho-Abreu et al., 2020). In B kDNA metrics were calculated with the `PercentageFeatureSet()` function which calculates the percentage of UMIs mapped to the kDNA maxicircle sequence per cell; as a cut-off, cells with more than 0.5% kDNA transcripts in their transcriptomes were excluded, as cells with low-quality features, or in the process of dying, display higher mitochondrial contamination (Satija et al, 2015). Similarly, in C rRNA contaminates were set at cells with more than 50% rRNA, and then all rRNA transcripts removed to avoid clustering or markers associated with these counts.

Once quality control cut-offs had been determined and applied to the Seurat object to remove unwanted cells, normalisation using the SCRAN package was applied to reduce biological variation originating from technical noise (Lun et al., 2016). The Scran package has been demonstrated as performing better in normalisation tests than other normalisation methods by Lytal et al., (Lytal et al., 2020). Scran was used to log₂ normalise feature counts, which was then multiplied by a scale factor and log-transformation. Next a subset of highly variable features from cell to cell were selected for calculating the dimensions, which then allows clustering and visualisation plots to be generated. This is so only what is variable in the dataset is used in downstream analysis, selecting for the top 2,000 variable features. The top ten variable features are labelled in the below scatter plot in Figure 3-4.

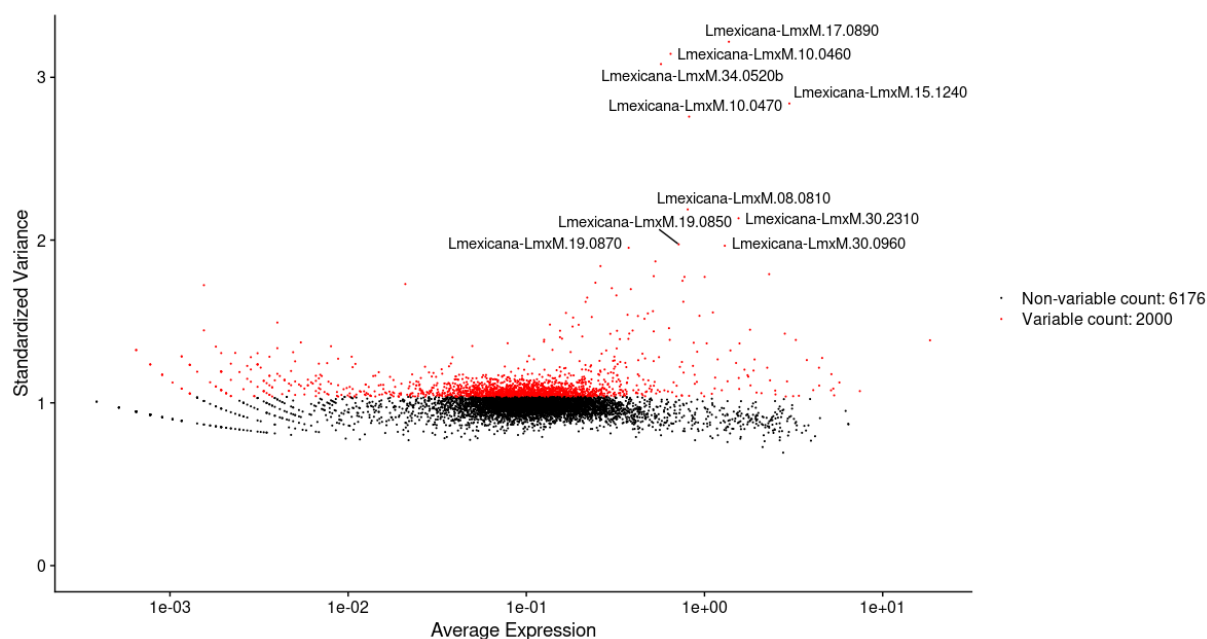


Figure 3-4 Scatter plot of top 2,000 variable features in Rep1 dataset

Scatter plot of top 2,000 variable features (red) selected for downstream analysis in sample Rep1, plotting average expression of feature against standard variance. Top ten variable features labelled.

In performing this analysis, known markers associated with different life cycle stages within *Leishmania* were revealed as variable between cells in the dataset. Prominent amongst these genes were two forms of GP63; LmxM.10.0460 and LmxM.10.0470, which is a membrane glycoprotein associated with the promastigote stages and has been demonstrated as being essential for increased parasitaemia and regulating the host immune system response (Isnard et al., 2012; Olivier et al., 2012). Additional genes associated with promastigote stages were seen in the top ten: the META domain containing protein; LmxM.17.0890 (Ramos et al., 2004), nucleoside transporter 1 (NT1.2); LmxM.15.1240 (Fiebig et al., 2015), 3'-nucleotidase/nuclease; LmxM.30.2310 (Sopwith et al., 2002), and two forms of ATG8/AUT7/APG8/PAZ2; LmxM.19.0850 and LmxM.19.0870. ATG8 is a protein recently studied by Giri & Shaha (2019), where over-expression in *L. donovani* resulted in increased resistance to stress conditions and increased infection rate *in vitro*. Disruption of the gene reduced infection and resulted in the parasites failing to differentiate into axenic amastigotes and stopping infection of both macrophages *in vitro* and in *in vivo* mouse models. Also of note, was prominently variables genes encoding for hypothetical proteins, LmxM.08.0810, LmxM.30.0960 and LmxM.34.0520b, the latter being a pseudogene fragment.

Next, 100 Principle Components (PCs), or “dims” were calculated using the 2,000 highly variable genes selected above following data scaling. To remove PCs which only explain a small proportion of variability in data, and likely contain only noise, the PCs were ranked based on the percentage of variance explained by each (as Briggs et al., 2021b), in Figure 3-5 below.

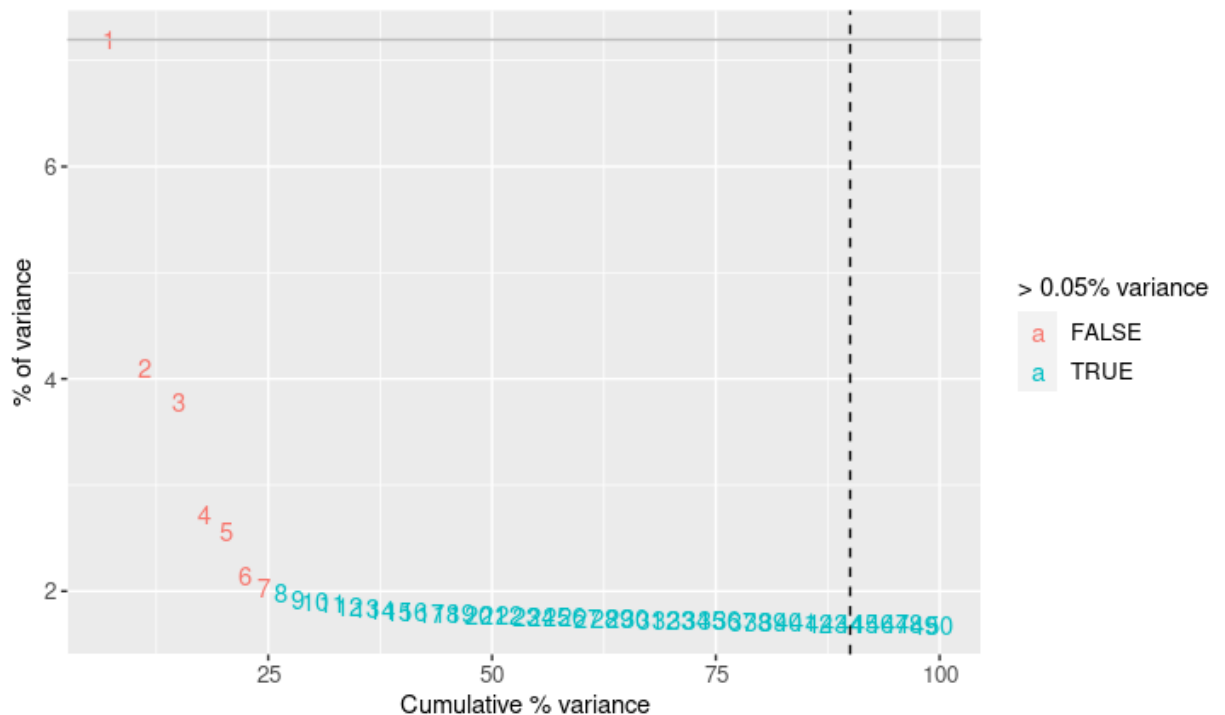


Figure 3-5 Elbow plot of variance found in principle components observed in Rep1 sample. Elbow plot to rank principle components (PC) found in Rep1 sample assessing the percentage of variance explained by each PC. In the Rep1 sample, an “elbow” is seen around PC 7, suggesting that the majority of variance is captured in the first 7 PCs.

In Figure 3-5 an ‘elbow’ is present around PCs 6 -7, which suggests most of the significant variance was captured in the first 7 PCs containing 2% of the variance. PC 7 is the last to contribute > 0.05% of the variable in the data, therefore the first 7 PCs were chosen for downstream analysis.

The top 7 PCs were used for clustering cells by likeness based on approaches first developed by Macosko et al (2015), Xu & Su (2015) and Levine et al. (2015). Briefly, cells are placed in a clustered graphical structure using K-nearest neighbours (KNN) with edges between clusters determined by similarity in feature expression patterns from one cell to another. Using the FindClusters() function, which stipulates a resolution parameter that determines how similar cells must be for them to be grouped into the same cluster. Where larger resolutions values lead to more clusters being defined. An appropriate resolution

value is able to separate cells into biologically meaningful clusters, each with a distinct set of marker genes and so must be user-defined. To help determine the best resolution for Rep_1, the results of different clustering resolutions (0.1 -1) were compared and the Clustree() function, produced by Zappia & Oshlack (2018), was used to determine how cells are clustered as resolution increases, as shown in Figure 3-6.

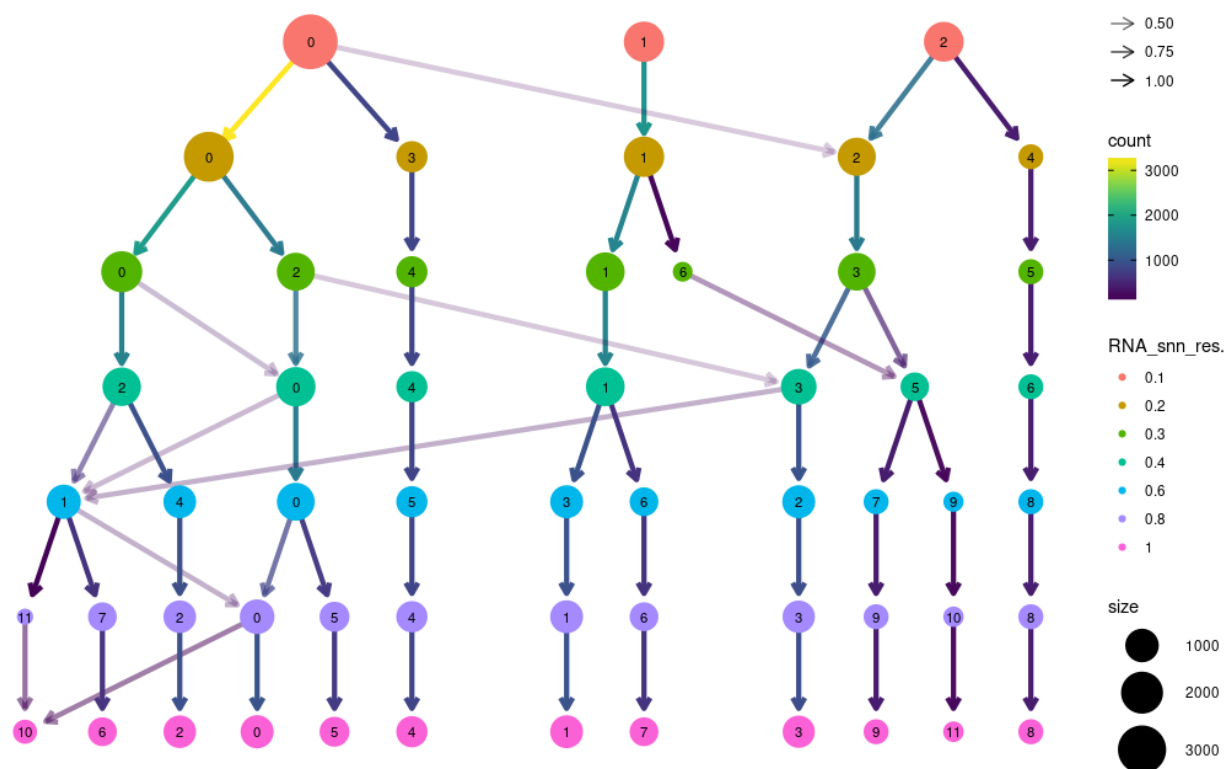


Figure 3-6 Clustree plot demonstrating Rep1 sample movement with increasing cluster resolution.

Clustree plot for Rep1 sample showing changes in cell cluster assignment as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells partitioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.2 (coloured in ochre) is chosen resolving into 5 clusters for further clustering analysis.

As shown in Figure 3-6 A, a resolution of 0.2 (coloured in ochre) partitioned the dataset into 5 distinct clusters and was selected for further marker analysis.

Potentially these clusters represent the 5 timepoints combined in the sample.

3.3.2.2 Clustering analysis of single cell RNA-sequencing for first replicate

To visualise the clustering of this dataset Uniform Manifold Approximation and Projection (UMAP) was used to plot the cells in low dimension space (combining

the 7 PCs), where cells are positioned based on similarity of expression patterns. Cells were coloured by clusters identified above, in Figure 3-6.

UMAP and t-SNE are two commonly used dimensionality reduction techniques used for visualizing high-dimensional data, such as scRNA-Seq data.

The primary difference between UMAP and t-SNE is that UMAP tries to preserve both local and global structures of the data while t-SNE focuses on preserving the local structure of the data. Essentially, UMAP tries to maintain the relationships between all points in the data (McInnes et al., 2018), while t-SNE only preserves the relationships between nearby points (Van der Maaten and Hinton, 2008).

When interpreting UMAP and t-SNE plots, it's important to note that the plots are just representations of the data in two or three dimensions. The position of each point in the plot reflects its distance from other points in the high-dimensional space. Therefore, points that are close to each other in the plot (representative of individual cells here) are likely to have similar expression profiles.

UMAP has become increasingly popular in recent years for visualizing single-cell RNA-Seq data because it is faster and more flexible than t-SNE. Additionally, UMAP has been shown to better preserve the global structure of the data, allowing for more accurate identification of clusters and cell types, and more accurate trajectory inference in following analysis (Hu et al., 2019; Song et al., 2019).

However, t-SNE may still be useful in certain situations where preserving the local structure of the data is more important, such as when analysing highly variable genes or when the data contains distinct subpopulations that are difficult to separate using UMAP.

Both t-SNE and UMAP are methods used in scRNA-seq analysis for reducing dimensionality in a non-linear, graph-based manner. They are primarily used for data visualisation. The general approach for both t-SNE and UMAP is to create a high-dimensional graph and then reconstruct it in a lower dimensional space while preserving its structure. However, the methods differ in that t-SNE moves

points individually from the high-dimensional graph to the lower dimensional space, while UMAP compresses the graph as a whole. However, several evaluations by Betch et al. (2019), Kobak et al. (2019) and Xiang et al. (2021), have demonstrated that considerations of crucial parameters for a given dataset will result in similar conclusions obtained by both methods, validating that both approaches are generally comparable. In these analyses, UMAP was chosen to best represent the larger datasets and for further analysis of trajectory progression, as detailed in Sections 3.3.3.5, 3.3.5.5. and 3.3.5.6..

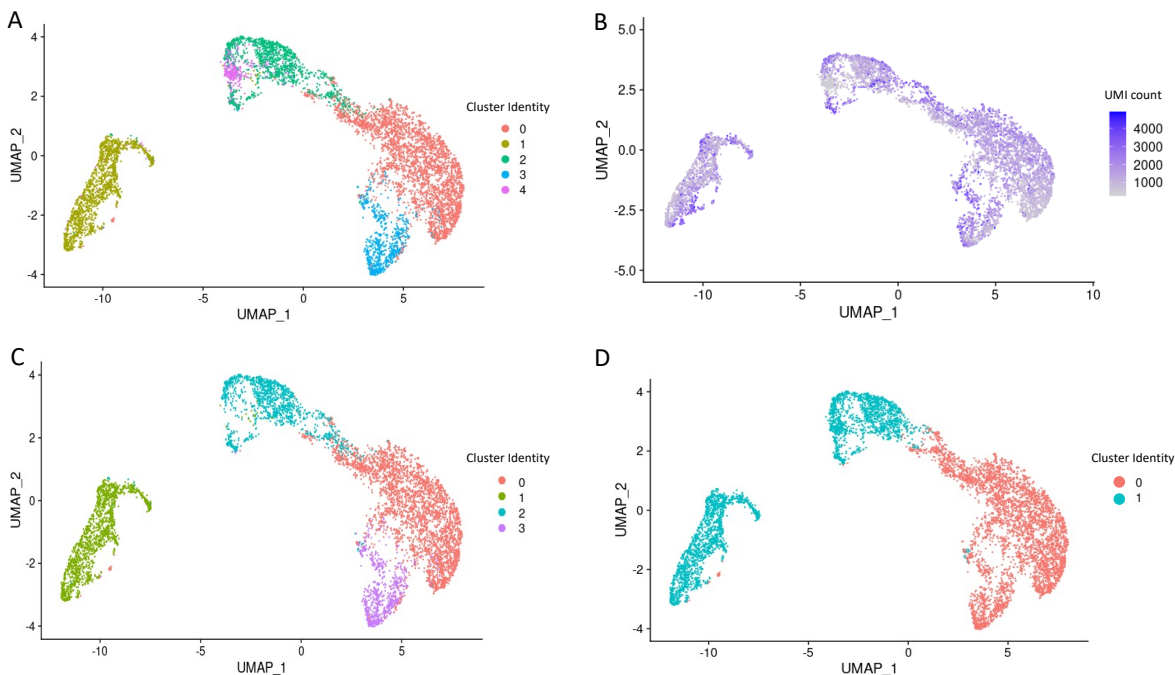


Figure 3-7 Clustering of *L. mexicana* transcriptomes across life cycle development from promastigote to axenic amastigote.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in Rep1 sample to visualise relative relationships between individual transcriptomes. **A)** UMAP of Rep1 sample at a resolution of 0.2 reveals 5 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** UMAP of Rep1 as in **A)** with low transcript cluster (cluster 4 in purple) removed from further marker analysis. **D)** Low resolution (0.03) UMAP of Rep1 sample to give only two clusters for determining similarity of central cluster to clusters placed either side at higher resolutions, demonstrating central cluster is more like clusters placed on the left.

Resulting clusters in Figure 3-7 **A**, using resolution 0.2 selected above, revealed 5 distinct clusters, where every dot represents the transcriptome of a single cell. In **B)** the total UMI count, indicating total transcripts captured, is displayed for each cell, coloured in purple for cells with higher expression counts. Of note is the very low counts in cluster 4 (coloured in purple in **A**). Following differential expression (DE) analysis of these clusters, no biologically relevant markers could be found in cluster 4. As this observation suggests cluster 4 does therefore, not represent a biologically distinct set of cells and has lower than average RNA

content, and so this cluster was removed from the dataset in the downstream marker analysis, giving a UMAP with 4 clusters, as shown in C. Shown in D is the UMAP with dimensionality reduced to 0.03 so that cells are grouped into only two clusters. Here, cluster 2 from plot C (coloured in turquoise) is combined with cluster 1 in plot C (coloured in apple green) despite the closer proximity of cluster 2 to clusters 0 (red) and 3 (purple). This initial result warranted further marker analysis of this cluster, to determine the potential similarity of the cluster to the clusters placed either side.

Next, cluster markers were investigated to assess the potential life cycle stage, and possible cell cycle stage of cells in each cluster. Expression of 12 markers are shown in Figure 3-8 by UMAP expression counts. Promastigote-like markers selected in this study were variants of the surface membrane glycoprotein proteinase, GP63. Forms of GP63 constitute 0.5-1% of all protein produced by the parasite (Wright & Roufaie, 1989). GP63 covers the whole surface of the promastigote cell, including the flagellar pocket, and has been demonstrated as a free moving membrane-anchored protein through antibody tagging (Lieke et al., 2008). Here, GP63 encoded by LmxM.10.0460 was selected for a promastigote stage marker (Medina et al., 2016) as well as another form of GP63, here named GP63-2 (LmxM.10.0470), which was seen in these data as a potential indicator for the metacyclic promastigote stage in these data. The markers Paraflagellar Rod 2 (PFR2; LmxM.16.1430) (Moore et al., 1996; Mishra et al., 2003), Glucose Transporter 2 (GT2; LmxM.36.6290) (Burchmore & Landfear, 1998), Promastigote Surface Antigen 38S (PSA; LmxM.12.0980) (Devault & Bäuls, 2008) and Major Surface Protease gp63 (MSP; LmxM.28.0570.) (Hallé et al., 2009; Yao et al., 2003) were also used as markers of promastigote specific stages. For amastigote stage markers Tuzin-like protein (LmxM.33.1970) (Lakshmi et al., 2014), Amastin (LmxM.30.0450) (Sopwith et al., 2002), A600-4 (LmxM.33.3645) (Bellatin et al., 2002; Murray et al., 2007) and Cathepsin L-like Protease (LmxM.08.1070) (Emiliano Da Silva et al., 2019; Mundodi et al., 2005) were selected. The amastigote-associated protein Cathepsin L, LmxM.08.1080, is a cysteine peptidase which is associated with stage specific expression in amastigote forms. This cystine peptidase plays a critical role in inhibiting immunomodulatory functions to increase amastigote infections *in vitro* using human macrophages and is expressed mainly in amastigotes (Buxbaum et al.,

2003; Kuru et al., 2007). Finally, two markers for cell cycle progression and cytokinesis were included: Histone-lysine N-methyltransferase (DOT1A) (LmxM.07.0025) (Gassen et al., 2012; Uzcanga et al., 2016) and Tip Of Extending FAZ protein 1 (TOEFAZ1) described in *T. brucei* (LmxM.31.2610) (Hilton et al., 2018; Sinclair-Davis et al., 2017).

Expression of the above marker genes in each cell, in order to distinguish which clusters may be attributed to specific life stages within this sample, in Figure 3-8.

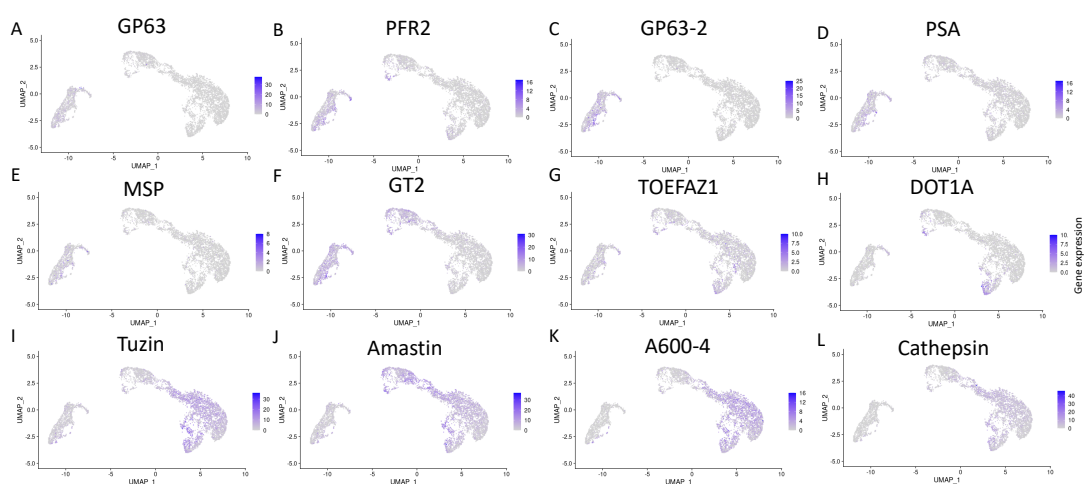


Figure 3-8 UMAP life cycle marker plots in Rep1 scRNA-seq sample to determine stages defined in clustering analysis.

Uniform Manifold Approximation and Projection (UMAP) of Rep1 sample coloured by selected life cycle stage specific markers. Markers **A** - **E**) (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E**) indicative of the metacyclic promastigote stages. **F**) GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** - **H**) (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L**) (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

In Figure 3-8 promastigote markers shown in **A** to **F** shows enrichment in the left most cluster 1, coloured by expression in purple. Cell cycle progression indicators in **G** - **H** were enriched in bottom edges of the central and bottom right clusters 2 and 3 (Figure 3-7). Amastigote markers **I** - **L** were found to be distinctly enriched in the rightmost clusters 0 and 3. Therefore, cell populations have been clustered by life cycle stages. The left cluster represent promastigotes, the right cluster axenic amastigotes with a separate replication amastigote cluster being distinct from the amastigote cluster above it. Interestingly, the cluster positioned between the promastigote and amastigote clusters (Figure 3-7, **C**, in turquoise) shows overlap of both the promastigote and

amastigote markers GT2 (F) and Amastin (J), as well as some enrichment for cell cycle indicating markers towards the bottom edge of the cluster. Thus, this cluster may represent a transitional cell type between the two stages present on either side.

To evaluate these data in another representation, Violin plots were produced, giving expression distributions across clusters, as shown below in Figure 3-9.

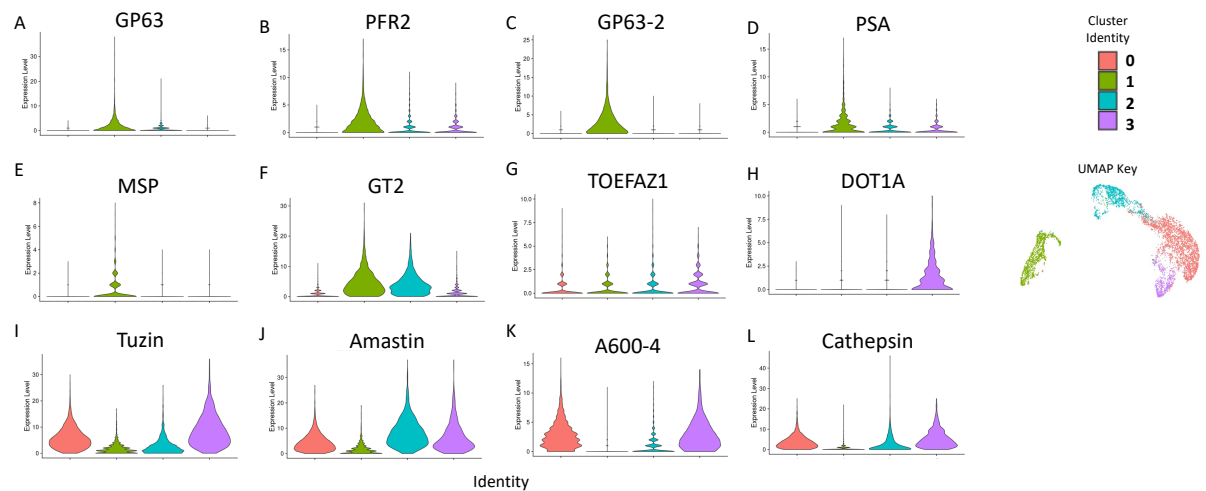


Figure 3-9 Violin plots of selected marker genes in Rep1 scRNA-seq sample.

Violin plots of Rep1 sample divided and coloured by clusters. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F** GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** – **H** (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L** (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

As shown in Figure 3-9, expression trends followed those demonstrated in Figure 3-8, where promastigote markers **A** - **F** were expressed only in cluster 1 (coloured in green), and conversely amastigote markers **I** - **L** were expressed in clusters 0 and 3 (coloured red and purple, respectively). Again, cluster 2 (coloured in turquoise), placed in between the promastigote and amastigote clusters, showed overlap in expression for promastigote marker GT2 (**F**) and Amastin (**J**). Replicating markers **G** and **H** were most seen in cluster 3 (coloured in purple) which also sees expression of amastigote markers, further indicating a potentially replicating amastigote population clustered separately from the larger amastigote cluster 0. Potential cell cycling populations in this sample are further explored in Section 3.3.2.4.

3.3.2.3 Marker analysis of first replicate

Following cluster identity investigation using known marker genes, a UMAP was replotted (Figure 3-10 A) with life cycle stages assigned, named Pro for the promastigote cluster, Axa for the axenic amastigote cluster, Axa Rep for the proliferating axenic amastigote cluster, and Trans A for the putative transitional cluster between promastigotes and axenic amastigotes. The proportion of cells assigned to in each cluster as shown in Figure 3-10 C, demonstrating a larger proportion of cells in the amastigote clusters, than promastigote. DE between clusters was calculated to find genes over expressed in one cluster, compared to all the others.

The top 10 DE markers for each cluster were plotted in a Heatmap in Figure 3-10 B and in a Dot plot in D. Clusters on the y-axis placed in order of life cycle progression starting from promastigote, and promastigote to amastigote life cycle markers, as described above, ordered left to right on the x-axis. Markers indicating potential proliferative cells placed in the centre of the x-axis, shown Figure 3-10 D.

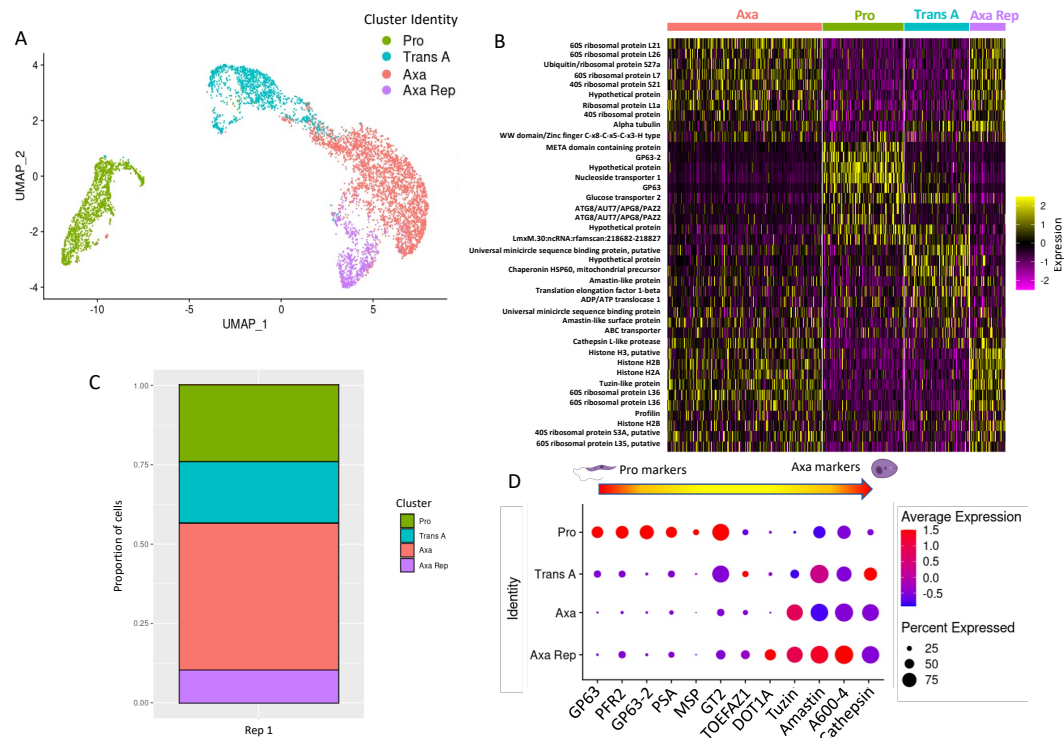


Figure 3-10 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep1 sample.

A) labelled clusters by life cycle stage as determined by expression patterns of life cycle stage specific markers. **B)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Dot plot of marker genes for life cycle specific stages, as Figure 3-8, ordered from left to right for promastigote to amastigote, with life cycle progression for clusters ordered from top to bottom. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot.

Cells were grouped by cluster in Figure 3-10 **B)**, with clusters ordered by size and expression per cell is plotted for each marker gene, showing distinct expression profiles for each cluster, with some overlap for Axa and Axa Rep clusters, consistent with a closer relationship than with the other clusters. Interestingly, expression profiles for Trans A showed most overlap with the Pro cluster. In total, 1,311 differentially expressed markers with a p-value <0.05 were found in this dataset: 156 differentially expressed markers were found for the Axa cluster, 16 of them being hypothetical proteins; 331 markers for the Pro cluster, with 91 hypothetical proteins; 618 for Trans A, with 131 being hypothetical proteins; and 206 markers for the Axa Rep cluster, with 34 being hypothetical proteins. Seven of the top ten markers for the Axa cluster were ribosomal protein encoding genes, a hypothetical protein, a slight decrease of alpha tubulin compared to Trans A and Axa Rep clusters, and a WW domain Zinc finger of the C-x8-C-x5-Cx3-H variety named WW for the two conserved tryptophan

residues, which is a protein domain that interacts with protein ligands (Kramer et al., 2010; Sudol et al., 1995). Seven of the top ten markers (META, GP63-2, NT1, GP63, GT2, and two copies of ATG8/AUT7/APG8/PAZ2) for the Pro cluster were well described promastigote markers as represented in the literature (Fiebig et al., 2015; Coelho et al., 2012; Coutinho-Abreu, et al., 2020), with several being used previously for life cycle cluster identification previously in Figure 3-8 and Figure 3-9 in this study. Interestingly, two hypothetical proteins and a non-coding (nc) RNA were also seen amongst the top ten promastigote cluster markers. Two of the top Trans A markers are copies of the universal minicircle sequence binding protein (Tzfati et al., 1995). Also seen were translation elongation factor 1-beta, ADP/Translocase 1 and an ABC transporter, the former being described as providing immune response protection in mouse models (Santos et al., 2021). Three of the top ten markers for Trans A were markers associated with the amastigote stage: two amastin-like proteins, and the cathepsin L-like protease, further suggestive of a possible transitional cell type between the two life cycle stages and not previously identified. For the Axa Rep cluster four histones and four ribosomal protein genes comprise 8 of the top ten markers, the other two markers being profilin, which is associated with cell development and cytokinesis (Krishnan & Moens, 2009), and amastigote marker tuzin-like protein, again suggesting this is an amastigote-like cluster. And lastly, the Dot plot, shown in **D** shows a general trend between life cycle labelled clusters and life cycle markers, based on average and percent of RNA abundance, from top left to bottom right, indicating a sequential expression of life cycle markers associated with life cycle progression.

3.3.2.4 Cell Cycle Analysis of First Replicate

Currently, it is not fully understood to what extent the timing and expression patterns of genes that modulate the development of life cycle stages is influenced by cell cycle-based gene synchronisation (Wheeler et al., 2011). To investigate this further, we wished to try and disentangle life and cell cycle-dependent gene expression changes as the determinants of the above clustering, and to see if it was possible to reveal the relationship between life cycle progression and cell cycle phase.

Here, we attempted to apply a method of cell cycle stage labelling developed by Briggs *et al.* (2021b), in order to see if it was possible to discern cell cycle-dependent changes in gene expression within your data and clusters. Cell cycle phase-specific markers derived from bulk RNA-sequencing by Archer *et al.* (2011) in *T. brucei* were used by Briggs *et al.* to label individual cells in a given population; cells with no assigned cell cycle phase marker expression above a given threshold, were labelled as non-cycling. Since no equivalent experiment has been performed in *Leishmania* to directly assign cell cycle markers, identified, orthologs from the Archer *T. brucei* dataset (2011) were identified for *L. mexicana* using BLASTp searches through TriTrypDB (Amos *et al.*, 2022). The lists of orthologs can be found in Appendix III, showing 67 orthologs found from *T. brucei* markers in *L. mexicana* for the Early G1 phase, 228 markers for Late G1, 107 for S Phase and 108 for G2M. Using these cell cycle phase markers, a ratio of each cell cycle phase score of 1.2 compared to the average expression is calculated. Shown in Figure 3-11, A - D, is the ratio of each cell over all other cells for the respective cell cycle phase.

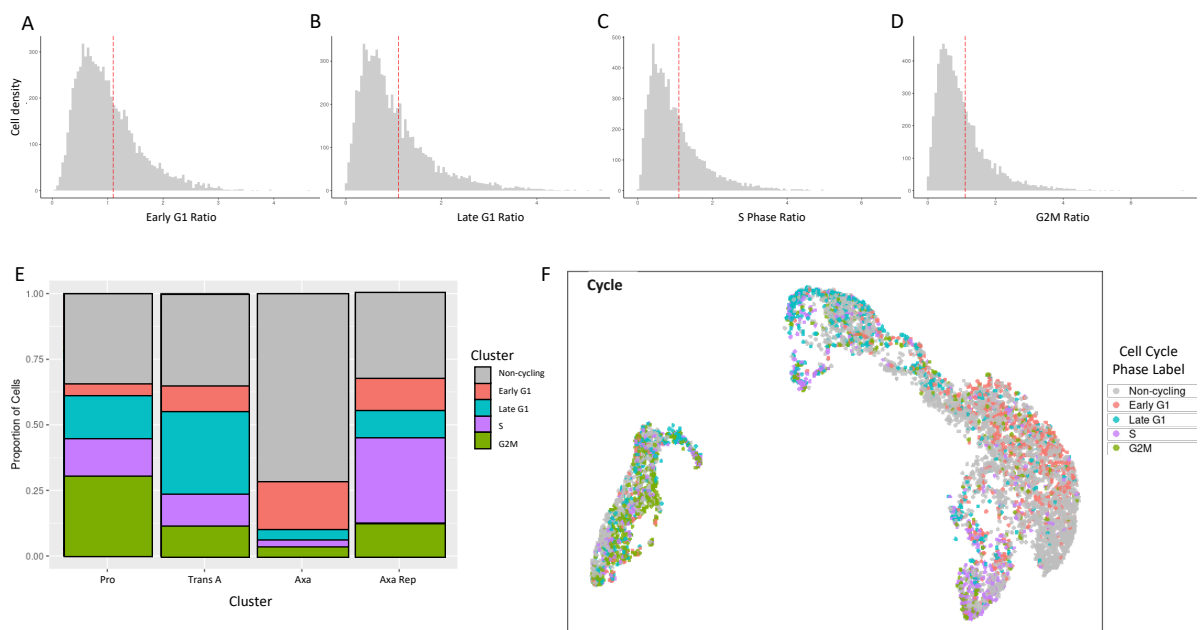


Figure 3-11 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep1.

A) – D) ratio of orthologous cell cycle progression markers from *T. brucei* with no cell cycle stage specific marker determined by those not over-expressing any cell cycle stage specific markers below a 10% ratio cut-off, indicated by the red dashed line. **E)** Proportion of cells labelled by cell cycle phase orthologous markers across clusters, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **F)** UMAP of cells in Rep1 sample assigned and coloured by cell cycle stage, as in **E)**.

Cells were assigned a cell cycle phase, or non-cycling, as shown in Figure 3-11 F revealing distinct groups of cells. This approach reveals differences between the

clusters in the amount of cells predicted to be in a specific cell cycle phase, or to be non-cycling. As shown in E three out of the four life cycle clusters have approximately 40% of cells labelled as non-cycling as they don't express markers associated with any cell cycle phase, with the Axa life cycle cluster being an exception, with 71.7% labelled as non-cycling. Despite this similarity in three of the clusters, there was considerable variation in proportions of cells predicted to be in specific cell cycle phases.

Most of the predicted cycling Axa cells were labelled as early G1 phase (coloured in red). In contrast, the most abundant cell cycle phase predicted for the Pro cluster was G2M (26.6%; coloured in green), with a small grouping in late G1 (14.1%; coloured in turquoise), located in the leading edge of the cluster closest to the Trans A cluster shown in F. Most cells in the Trans A cluster were labelled in late G1 (27.1%). Finally, in the Axa Rep cluster, approximately 26.9% of cells were labelled with S phase markers, consistent with the possibility that many of these cells are S phase enriched, and potentially proliferative. For consideration, is that markers may miss some cycling cells, as they are taken from orthologues from *T. brucei* (Archer et al., 2011).

3.3.3 Single cell RNA-sequencing analysis for second promastigote only replicate

To further explore gene expression during the life and cell cycle of *L. mexicana*, the promastigote samples were examined next (Rep2_Pro), as described above in Section 3.3.1. Sample data with timepoints shown for each combined sample is shown below in Figure 3-12.

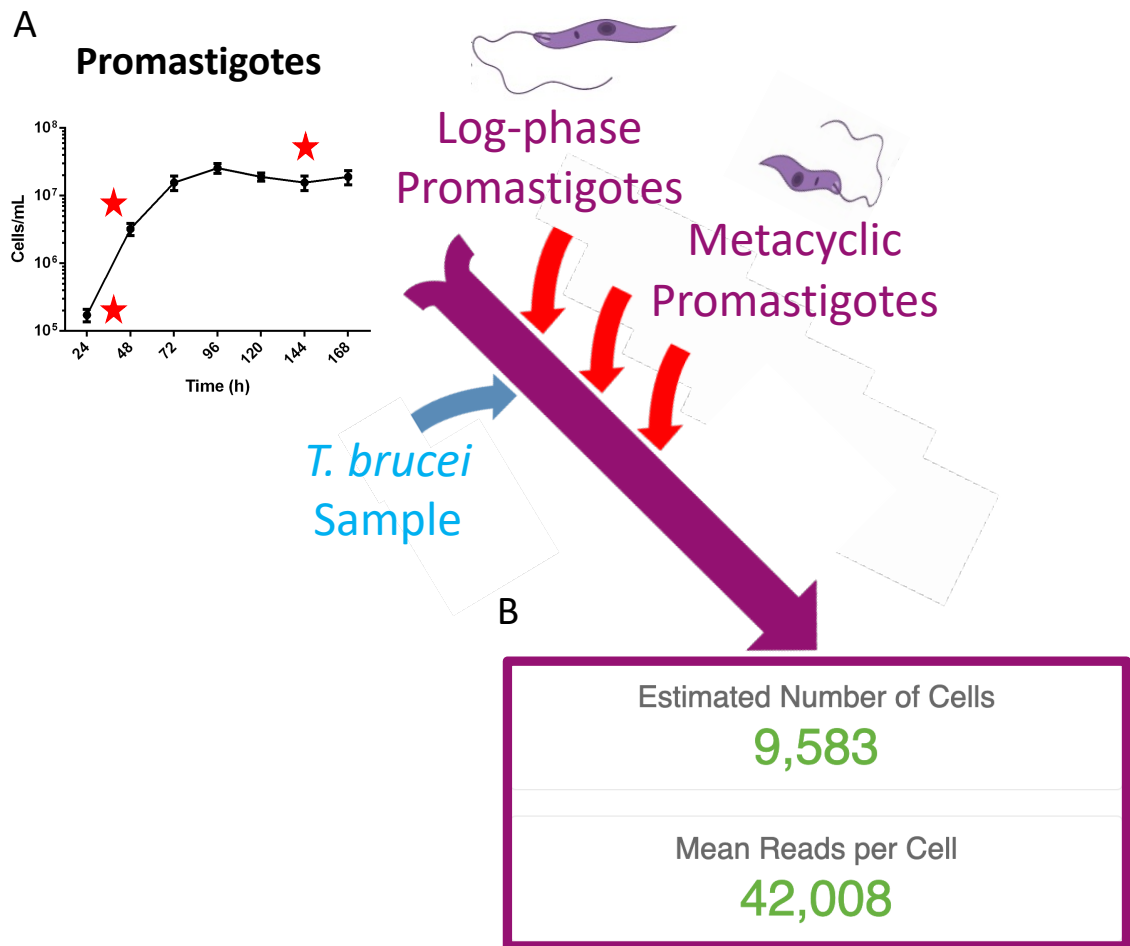


Figure 3-12 Experimental sample collection and sample information for second replicate of promastigote timepoints.

Experimental set-up for second replicate *Leishmania* sample run in single cell RNA-Sequencing **A**) over 3 timepoints (red stars) across log-phase to stationary phase promastigotes, the latter to enrich for metacyclic forms. Replicate Rep2_Pro containing *Leishmania mexicana* samples with 3000 cells from each timepoint and were mixed with *Trypanosoma brucei* cells to estimate multiplets in sequencing sample. **B**) In total, 9,583 cells were sequenced with mean reads per cell of 42,008.

Here, the combined samples of *L. mexicana* and *T. brucei* contained an estimated 9,583 cells with a combined mean read per cell of 42,008. As explained in Section 3.3.1, following the generation of reads, quality control of the samples is applied to limit the impact of noise generated due to biological and technological variations (Kolodziejczyk et al., 2015), as described above. Quality Control plots for sample Rep2_Pro can be seen in Appendix I.

The top nine PCs were used for clustering cells by likeness as described previously for Rep1, shown below in Figure 3-13.

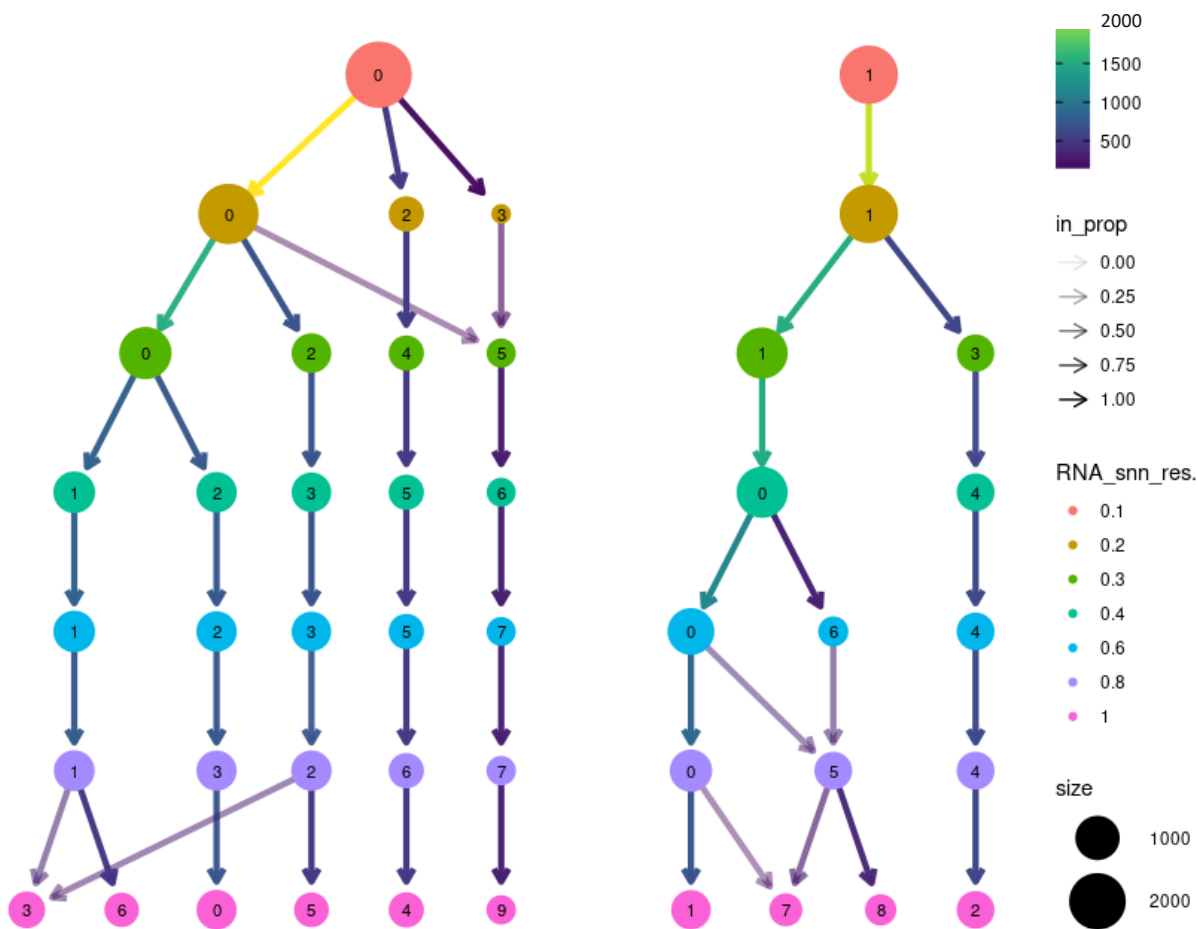


Figure 3-13 Clustree plot demonstrating Rep2_Pro sample movement with increasing cluster resolution.

Clustree plot for Rep2_Pro sample showing changes in sample movement as resolution increases (RNA_snn_res). A resolution of 0.2 is chosen resolving into 4 clusters for further clustering analysis. Proportion of cells moving from between clusters as resolution increases (from top to bottom) depicted by colour of arrows, and size of cells within clusters indicated by the relative size of depicted circles.

A resolution of 0.2 (coloured in ochre) partitioned the dataset into 4 distinct clusters. As this potentially represents the four life cycle stages associated with promastigote differentiation, a resolution of 0.2 was selected for further clustering analysis, explored below in Section 3.3.3.1.

3.3.3.1 Individual clustering analysis of single cell RNA-sequencing for second promastigote only replicate

The clustering of this dataset are plotted as UMAP in Figure 3-14.

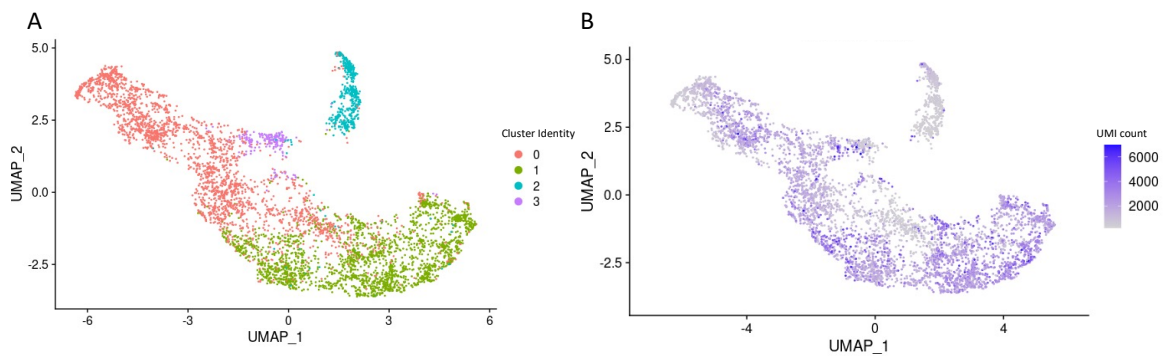


Figure 3-14 Clustering of *L. mexicana* transcriptomes across life cycle development from log phase promastigotes to stationary phase promastigotes.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in Rep2_Pro sample to visualise relative relationships between individual transcriptomes. **A)** UMAP of Rep2_Pro sample at a resolution of 0.2 reveals 4 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell.

Preliminary clustering Figure 3-14 revealed 4 distinct clusters, with cluster 2 (turquoise) being slightly aside from the other three clusters (cluster 0 in red, 2 in green, and 3 in purple). In **B** the total RNA count is displayed for each cell, coloured in purple for cells with higher expression counts. Of note is the lower counts in cluster 2 (turquoise in **A**). However, reiterations of analysis of these clusters did still reveal biologically relevant markers, so all clusters were kept in this dataset for downstream marker analysis. Also for consideration are the non-replicative promastigote stages, nectomonad promastigotes and metacyclic promastigotes which could yield fewer transcripts in comparison to the replicating procyclic promastigote and leptomonad promastigote forms (Gossage et al., 2003; Coutinho-Abreu et al., 2020).

After clustering each cluster was assigned an identity using known cell type markers. Here, life cycle stage markers were used following literature surveys for each of the life cycle stages and clusters possibly going through cell cycle to identify any replicating populations. Twelve markers are shown to determine general promastigote stages, to identify specifically metacyclic promastigotes, and replicating clusters.

The same marker analysis used in Rep1 as described in Section 3.3.2.2 is used for comparison below in Figure 3-15.

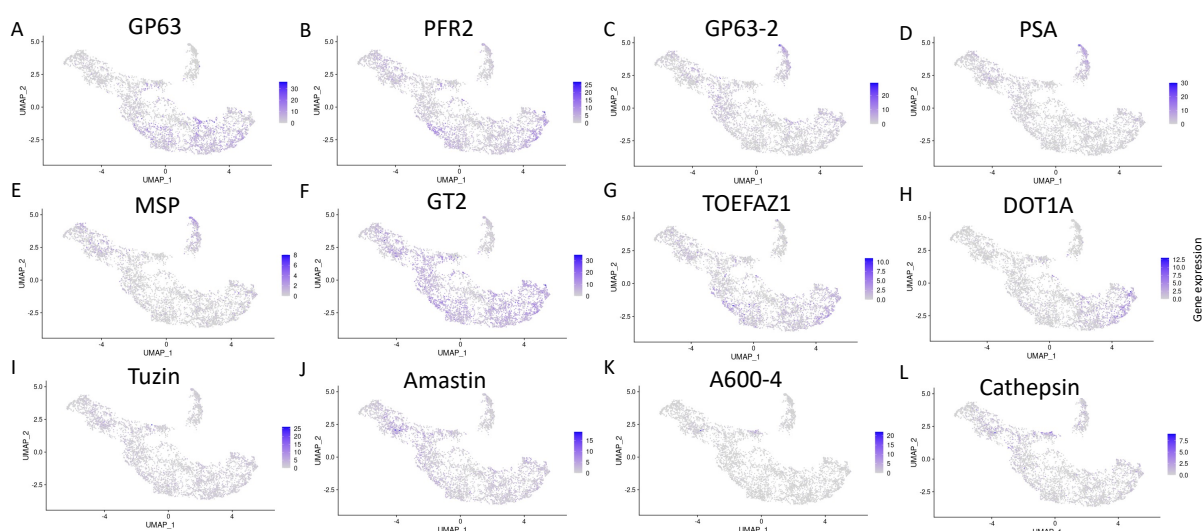


Figure 3-15 UMAP life cycle marker plots in Rep2_Pro scRNA-seq sample to determine stages defined in clustering analysis.

UMAP of Rep2_Pro sample coloured by selected life cycle stage specific markers. Markers **A** - **E**) (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F**) GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** – **H**) (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L**) (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

In Figure 3-15 promastigote markers from **A** to **F** showed enrichment across the sample, coloured by expression in purple. Cell cycle markers in **G** - **H** were enriched in the bottom right cluster, suggesting a potentially proliferative population. Amastigote markers **I** - **L** were not found to be distinctly enriched in any clusters, confirming the absence of amastigotes in this promastigote-only sample. Interestingly, the cluster positioned at the top of the UMAP (cluster 2 in Figure 3-14 **A**) had an enrichment for metacyclic promastigote markers, as shown in **C**, **D** and **E**.

Violin plots of marker gene expression data are shown in Figure 3-16.

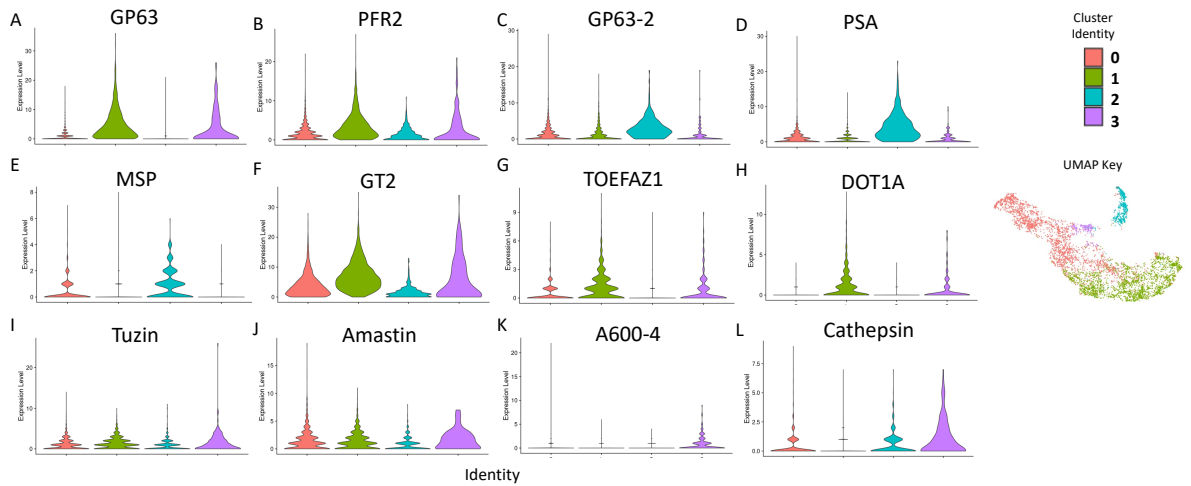


Figure 3-16 Violin plots of selected marker genes in Rep2_Pro scRNA-seq sample.

Violin plots of Rep2_Pro sample divided and coloured by clusters. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F** GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** - **H** (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L** (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

As shown in In Figure 3-16, expression trends follow those demonstrated in Figure 3-15, where promastigote markers **A** - **F** were expressed across the sample with an enrichment of metacyclic promastigote markers in **C**, **D** and **E** most highly expressed in cluster 2 (coloured in turquoise). Conversely, amastigote markers **I** - **L** were comparatively under-represented, albeit with a slight enrichment in the smallest cluster, 3 (coloured in purple). Proliferating markers **G** and **H** were most expressed in cluster 1 (coloured in green), which also showed expression of promastigote markers not associated with metacyclic forms (shown in **A** and **B**), further indicating a replicating promastigote population clustered separately from the smaller metacyclic-like cluster 2.

3.3.3.2 Cell cycle analysis of second replicate in promastigote stages

To further elucidate possible cell cycle-based regulation of life cycle development in *L. mexicana* using this scRNA-seq data, cell cycle stage labelling developed by Briggs *et al.* (2021b) was recapitulated and applied again here for the Rep2_Pro sample (Figure 3-12), as described above in Section 3.3.2.4.

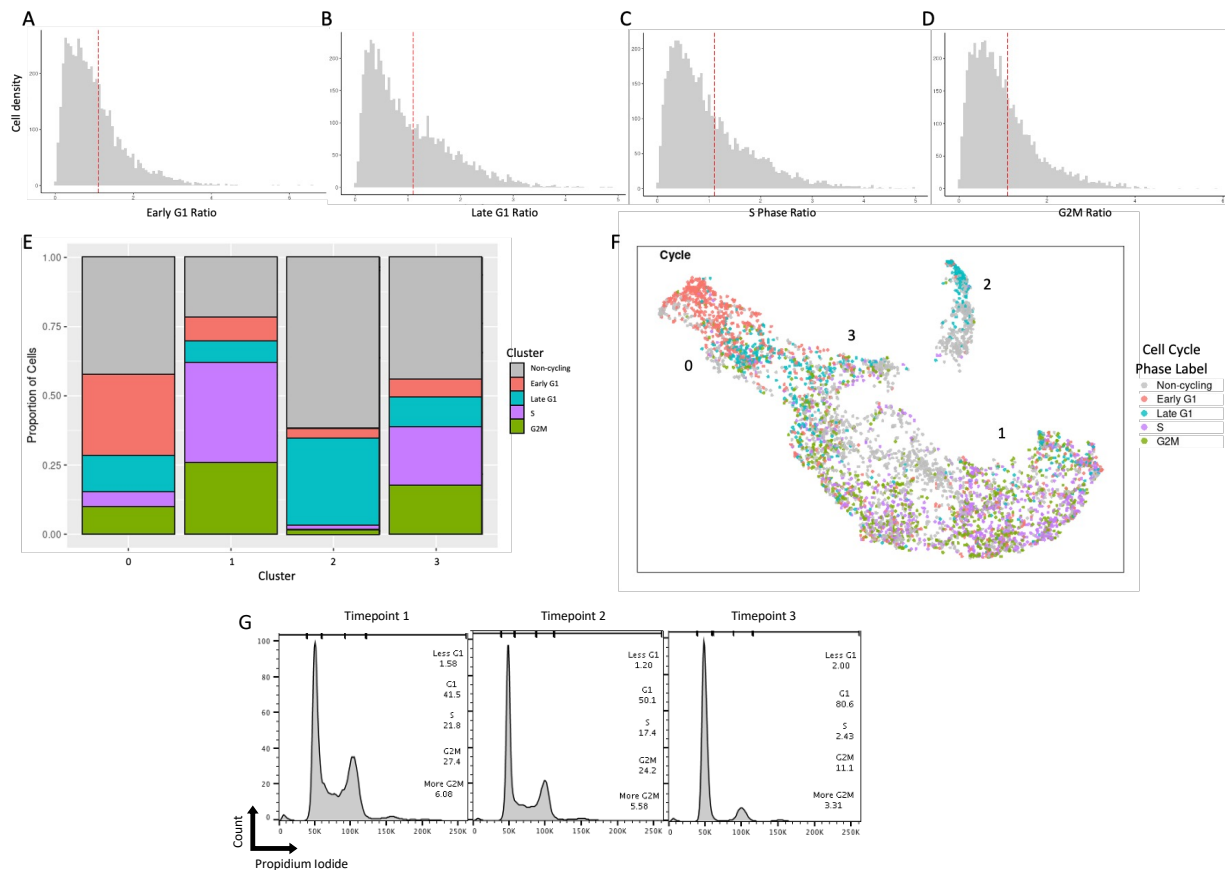


Figure 3-17 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep2_Pro.

A) – D) ratio of orthologous cell cycle progression markers from *T. brucei* with no cell cycle stage specific marker determined by those not over-expressing any cell cycle stage specific markers below a 10% ratio cut-off, indicated by the red dashed line. **E)** Proportion of cells labelled by cell cycle phase orthologous markers across clusters, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **F)** UMAP of cells in Rep2_Pro sample assigned and coloured by cell cycle stage, as in **E)**. **G)** Samples from each timepoint in promastigote life cycle progression were analysed for DNA content by flow cytometry using propidium iodide staining. Proportions in cell cycle stages gated and quantified within plots.

Assigning cell cycle stages to cells over-expressing cell cycle markers and labelling each phase by colour, as shown in Figure 3-17 F, again revealed distinct groups of cells enriched for different cell cycle phases. As shown in E Cluster 1 showed the largest proportion of cells assigned cell cycle phases (77.9%), almost 36.1% of these being in the S phase (purple), indicating a very actively proliferating cluster potentially associated with the logarithmic growth timepoint take for this sample. Clusters 0 and 3 displayed approximately 50% labelling of cell cycle phases, with cluster 0 seeing a large enrichment of early G1 phase (coloured in red) and cluster 3 being mainly a combination of S phase and G2M phases. Cluster 2 displayed the least cell cycle labelling, with 62.1% labelled as non-cycling (grey). For the cluster 2 cells that were labelled, the large majority, 29.6%, were considered late G1 (coloured in turquoise). This result is perhaps consistent with the G1/G0 cell cycle phase seen in other non-

replicative infective pathogenic stages in other microorganisms, such as the G1/G0 arrest found by flow cytometry described in *T. brucei* metacyclic forms derived from tsetse flies (Shapiro et al., 1984).

In G, flow cytometry data of DNA content in the three timepoints is shown, where samples from the scRNA-seq Rep2_Pro timepoints were taken and fixed in tandem with the scRNA-seq sample preparation. Samples were stained with propidium iodide. As can be seen in these flow cytometry histogram plots, an enrichment for the G1 stage is found across the three timepoints, increasing from 41.5% and 50.1% in the first two log phase timepoints to 80.6% in the third, stationary phase timepoint, predicted to be enriched for the metacyclic promastigotes. This doubling of the G1 phase by flow cytometry may reflect the predominant G1 phase labelling seen in cluster 2 (Figure 3-17, E).

3.3.3.3 Promastigote life cycle stage markers

In order to further elucidate possible life cycle regulation and development in *L. mexicana* using these scRNA-seq data, a method of life cycle stage labelling developed by Briggs *et al.* (2021b) was recapitulated and applied with life cycle stage markers from bulk RNA-sequencing in *L. infantum* by Coutinho-Abreu *et al.* (2020). No equivalent *L. mexicana* life cycle markers have been yet described, and as such, orthologs from the Coutinho-Abreu dataset were found for *L. mexicana* using BLASTp searches through TriTrypDB (Amos et al., 2022). These stages of promastigote being procyclic, nectomonad, leptomonad and metacyclic promastigotes. The lists of orthologs can be found in Appendix IV, showing 18 orthologs found from *L. infantum* markers in *L. mexicana* for the replicating procyclic promastigote stage, 16 markers for nectomonads, 15 for leptomonad, and 90 for the infective, non-replicating metacyclic promastigotes. Proportions of cells labelled with promastigote life cycle phases are shown below in Figure 3-18, B - C.

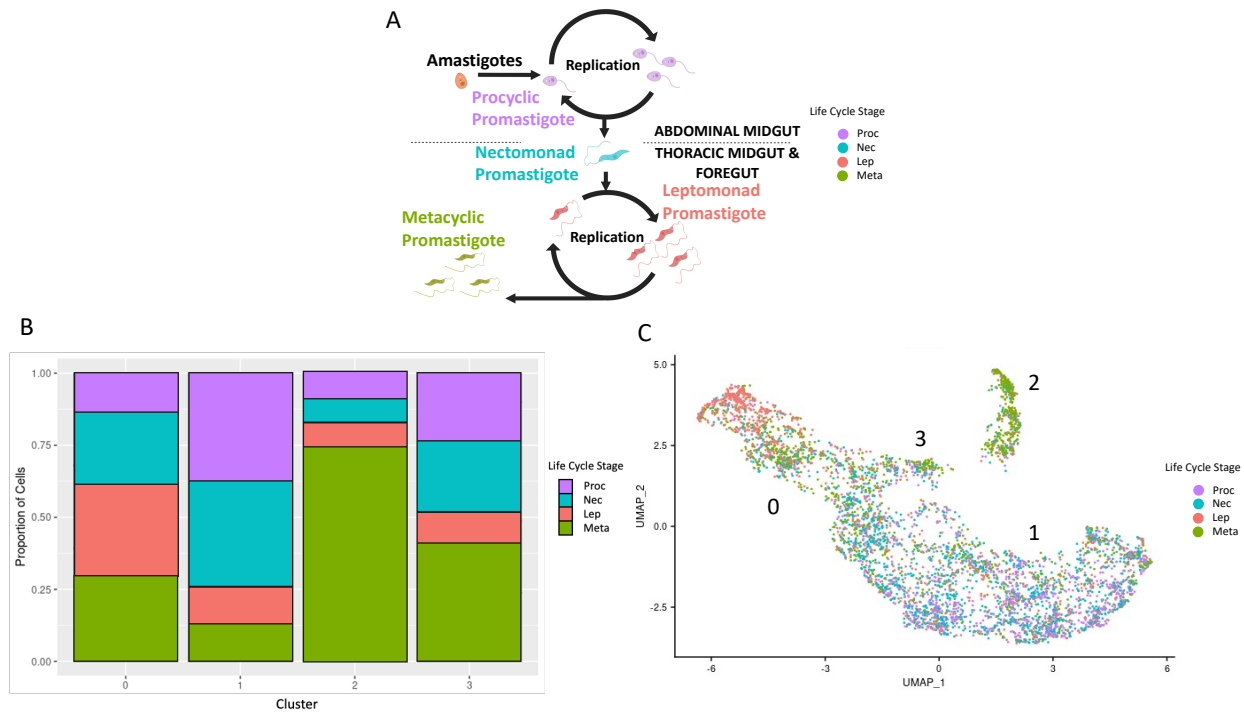


Figure 3-18 Promastigote orthologous labelling from procyclic promastigotes to metacyclic promastigote forms.

A) Schematic of promastigote stage specific progression from procyclic promastigotes (purple) which have differentiated from amastigotes replicating within the abdominal midgut of the sandfly vector. Non-replicative nectomonad promastigotes (turquoise) migrate from the abdominal midgut to the thoracic midgut and foregut where they develop into replicating leptomonad forms (red). Leptomonad forms then develop into the non-replicative infective form the metacyclic promastigote (green). **B)** Orthologous genes for stage specific forms of promastigotes found from bulk RNA-sequencing were used to allocate cells within the Rep2_Pro sample a promastigote stage and plotted as a proportion of cells across clusters and coloured by stage as in **A)**. **C)** UMAP of Rep2_Pro sample labelled by promastigote stage and coloured as in **A)** and **B)**. Schematic in **A)** adapted from Bates, (2018) and created in BioRender.

Figure 3-18 **A** is a schematic to demonstrate life cycle progression in the promastigote forms. From replicating procyclic promastigotes (coloured in purple) in the abdominal midgut which are differentiated from amastigotes, progression leads to the non-replicating nectomonad promastigotes into the replicating leptomonad promastigotes (coloured in red) across the thoracic midgut and foregut, with finally residing as the infective, non-replicating metacyclic promastigote forms (coloured in green) in the sandfly proboscis ready for egestion via the next bloodmeal (Sunter and Gull, 2017; Bates, 2018).

Labelling of these promastigote life cycle stages in **C**, using the same colours as in **A**, showed a significant overlap of both procyclic and nectomonad forms in cluster 1, with an enrichment of leptomonad cells toward the left edge of cluster 0, despite being generally mixed overall in this cluster for labelling in **B**. A large enrichment of metacyclic labelling was seen in cluster 2, consistent with non-replicating cell cycle labelling (Figure 3-17, **E** & **F**). Cluster 3, the smallest cluster, appeared to be a mix of life cycle stages with no particular enrichment

for any promastigote stage. To further investigate the overlap of life cycle and cell cycle stages in promastigote forms, a proportion of labelling of each data sets was undertaken in Figure 3-19 below.

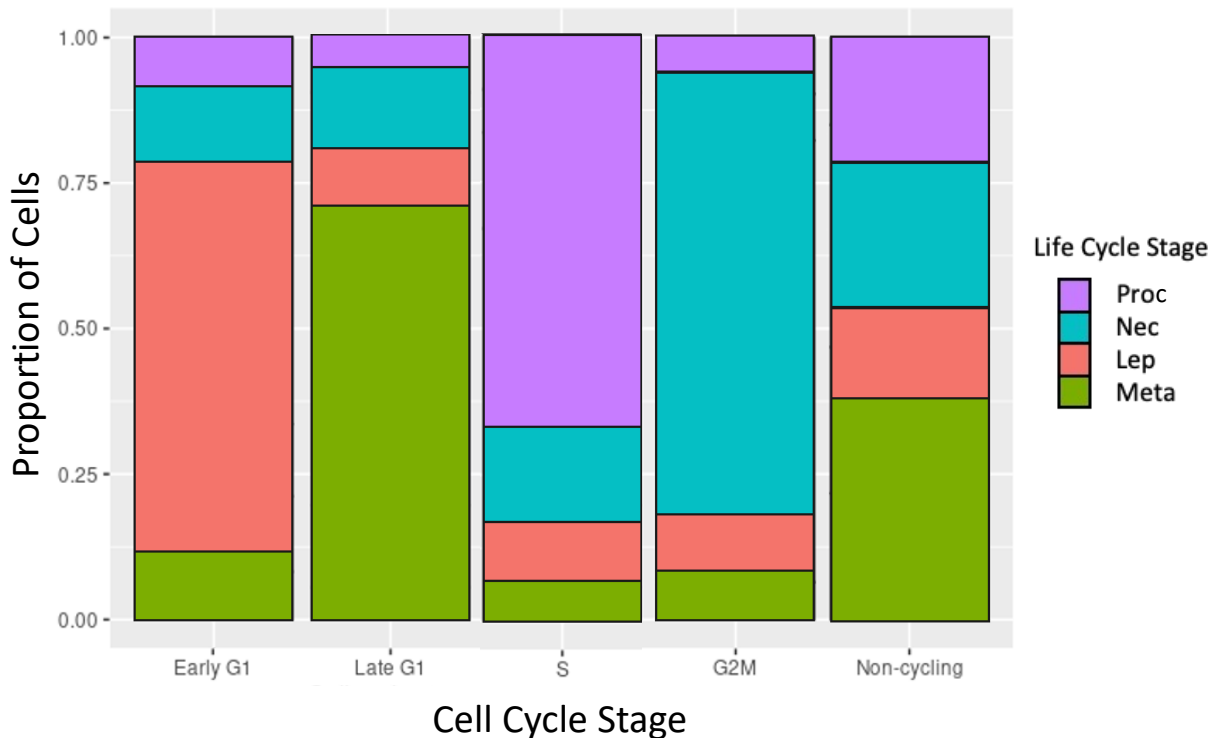


Figure 3-19 Proportions of labelled promastigote life cycle stages in respectively labelled cell cycle stages in the Rep2_Pro scRNA-seq sample.

Transcriptomes of cells categorised by cell cycle phase orthologous markers were coloured by promastigote stage specific markers to represent proportions of each promastigote life stages within their respective cell cycle stage. Cells not overexpressing any cell cycle specific markers above a 10% ratio were labelled as non-cycling, as in Figure 3-17. Promastigote stages coloured as procyclic promastigotes in purple, nectomonad promastigotes in turquoise, leptomonad promastigotes in red and metacyclic promastigotes in green.

By plotting each labelled cell cycle phase as a proportion of the promastigote life cycle stage the same cells were labelled with, we find a striking pattern, with each cell cycle phase being largely made up of one promastigote life cycle stage. Interestingly, the sequencing of cell cycle progression matched the order of promastigote life cycle development, where procyclics (coloured in purple) made up 66.4% of S phase labelled cells, nectomonads made up 75.1% of G2M, leptomonads made up 66.3% of early G1, and metacyclics made up 70.9% of late G1 cells. Nectomonads and metacyclics cells, known for being non-replicative (Gossage et al., 2003), also made up the two largest promastigote phases in the non-cycling cells. These data implicate metacyclic exist in late G1 or putative G0 cell cycle stage, before moving into differentiating forms needed for amastigote formation. This also provides further evidence that differentiation

between life cycle forms in promastigotes may also be linked to cell division and cell cycle stages (Wheeler et al., 2011; Sunter and Gull, 2017).

3.3.3.4 Marker analysis of second promastigote only replicate

To further investigate novel markers associated with different life cycle stages in the Rep2_Pro sample, marker analysis between clusters was undertaken. Following cluster identity investigations, a UMAP was replotted with life cycle stages assigned (Figure 3-20, **A**), named Lep/Meta for the leptomonad and metacyclic mixed cluster (red), Proc/Nec for the procyclic and nectomonad mixed cluster (green), Meta for the metacyclic only cluster (turquoise), and Cyto for the smallest cluster (purple) shown in **C**, so named after further marker investigation explained below. Proportion of cells in each cluster is shown in **C**. The top ten differentially expressed markers are plotted in a heatmap in **B**. Also plotted was a Dot plot with clusters on the y-axis placed in order of life cycle progression, and promastigote to amastigote life cycle markers, as described in Section 3.3.2.2, ordered left to right on the x-axis, shown in **D**. These plots are given below in Figure 3-20.

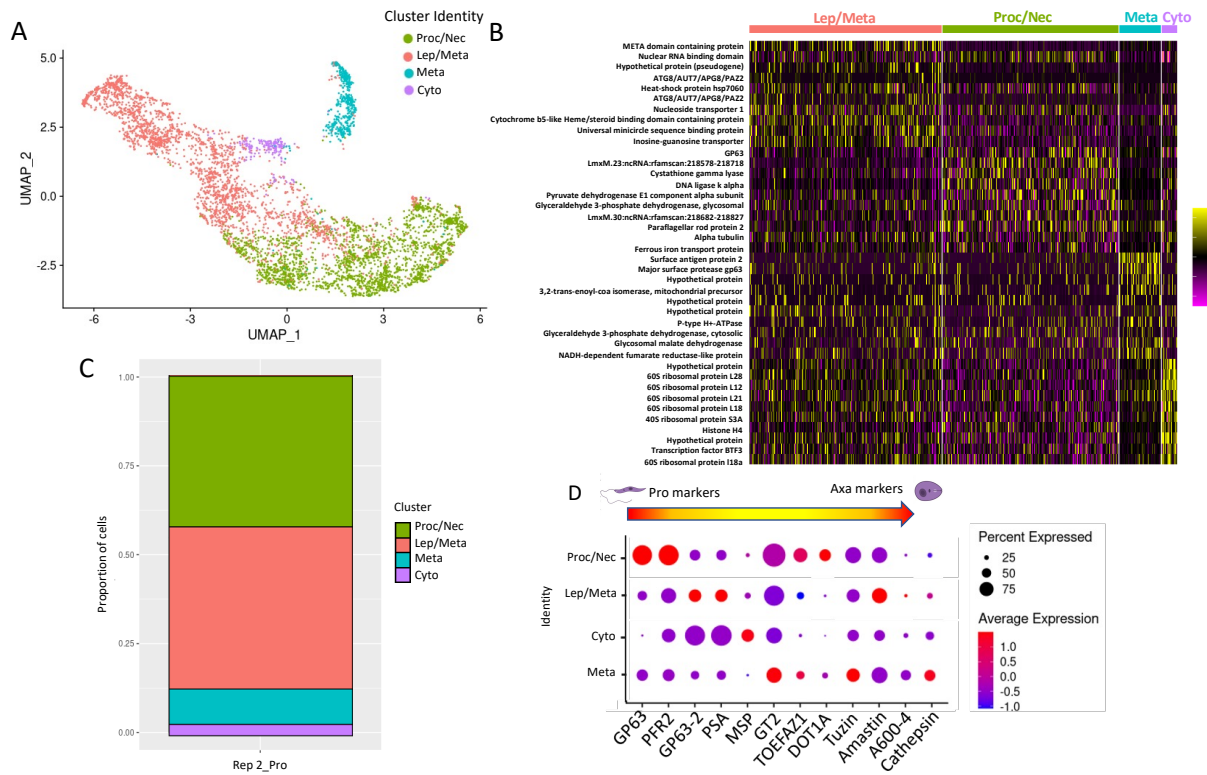


Figure 3-20 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep2_Pro sample.

A) labelled clusters by life cycle stage as determined by expression patterns of life cycle stage specific markers. **B)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Dot plot of marker genes for life cycle specific stages ordered from left to right for promastigote to amastigote, with life cycle progression for clusters ordered from top to bottom. Cell proliferation markers TOEFAZ1 and DOT1A are set in between life cycle stage markers. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot.

Gene scaled expression for the top ten expressed markers shown in Figure 3-20 B are ordered by size of the clusters, from left to right annotated on top with each column representing the expression profile of a cell, showing distinct expression profiles for each cluster, with distinct markers for the Meta cluster (coloured in turquoise) compared to the other clusters.

In total, 2890 differentially expressed markers were found in this dataset: 93 differentially expressed markers were found for the Lep/Meta cluster, 19 of them being hypothetical proteins; 2332 markers for the Proc/Nec cluster with 746 hypothetical proteins; 13 for Meta with 4 being hypothetical proteins; and 452 markers for the Cyto cluster with 66 being hypothetical proteins. Shown in B, the Lep/Meta cluster was made up of classically associated leptomonad markers, such as two forms of the ATG8 proteins and heat-shock protein hsp7060, a hypothetical protein pseudogene, and nucleoside transporter 1 and

META domain containing proteins (Coutinho-Abreu et al., 2020). Of interest are also the markers for a Nuclear RNA binding domain, Cytochrome b5-like Heme/steroid binding domain containing protein, Universal minicircle sequence binding protein and Inosine-guanosine transporter.

Three of the top ten markers for the Proc/Nec cluster were known promastigote markers (Medina-Acosta, et al, 1989; Mishra et al., 2003), two of which were promastigote markers previously used in cluster analysis here, being GP63, PFR2, and alpha tubulin (Ramírez et al., 2013). Of note is the appearance of the same ncRNA as seen in the top ten markers of Rep1 in this cluster. Also included in the Proc/Nec cluster was the DNA ligase k alpha marker previously associated with cell cycling populations and consistent with the large proportion of S phase labelled cells for this cluster. Other markers in the top ten for the Proc/Nec cluster included Cystathione gamma lyase, Pyruvate dehydrogenase E1 component alpha subunit, Glyceraldehyde 3-phosphate dehydrogenase (glycosomal) and a Ferrous iron transport protein. Three of the top Meta markers were hypothetical proteins, while with two other markers have previously been associated with metacyclics, being Surface antigen protein 2 and Major surface protease gp63. Other markers for the Meta cluster were a mitochondrial precursor, P-type H⁺-ATPase, Glyceraldehyde 3-phosphate dehydrogenase (cytosolic), Glycosomal malate dehydrogenase, and NADH-dependent fumarate reductase-like protein. For the smallest cluster, named Cyto here for the large proportion of cytokinesis and translation gene ontology markers present, six of the top ten markers were ribosomal genes, as seen in the Axa replicating cluster in Rep1, 40S ribosomal protein S3A also being repeated in the six ribosomal proteins here. Histone H4 was also seen in this cluster, although a different copy of H4 here than as seen in the Axa Rep cluster in Rep1. Two hypothetical proteins were found in the top ten of the Cyto cluster, as well as the Transcription factor BTF3, which is known to form a complex with RNA polymerase II in humans and yeast (X. M. Zheng et al., 1987) and has recently been associated with playing a role in *Leishmania donovani* and *Leishmania infantum* infections in dual bulk RNA-Sequencing analysis (Forrester et al., 2022). Lastly, the Dot plot, shown in **D**, shows a general trend from top left to bottom right of promastigote to amastigote markers, with some amastigote markers appearing in leptomonad and metacyclic clustered cells compared to

the other cells in this sample, potentially representing transcripts being produced in the latter promastigote stages in preparation of life cycle development into amastigote forms. This last finding is also consistent with amastigote associated transcripts, such as several amastin-like surface proteins, being up-regulated in metacyclic cells when examined by bulk RNA-Seq by Dillon et al., (2015) in *L. major*.

3.3.3.5 Pseudotime life cycle progression analysis of second promastigote only replicate

Currently, it is not fully understood to what extent the timing and expression patterns of genes that modulate the development of life cycle stages is controlled. To investigate this further life cycle trajectory and pseudotime analysis of the Rep2_Pro dataset was undertaken. Having collectively found evidence of life cycle populations progressing through promastigote stages in the Rep2_Pro sample via the analysis described above, the Rep2_Pro sample was then used for trajectory and pseudotime analysis to infer further if the life cycle progression and gene expression already demonstrated may also be ordered by their life cycle stages. The analysis of which can be found below in Figure 3-21.

Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) is a dimensionality reduction technique that uses a novel geometric framework to identify low-dimensional representations of high-dimensional data, such as scRNA-seq data. PHATE captures the intrinsic structure and geometry of the data and can be used for a variety of tasks, including trajectory inference.

PHATE's approach to trajectory inference is particularly relevant for scRNA-seq data because it is designed to handle the challenges associated with capturing the nonlinear and heterogeneous relationships between cells. PHATE creates a representation of the data that captures both the global and local structure of the data, allowing for accurate identification of cellular trajectories and differentiation pathways. Additionally, PHATE can handle data with missing values, which is common in scRNA-seq experiments.

PHATE has been shown to outperform other commonly used trajectory inference methods, such as Monocle (Saelens et al., 2018), in terms of accuracy and robustness to noise and missing data. Furthermore, PHATE can be used to identify previously unknown cellular states and transitions, which is useful for discovering new cell types and developmental pathways.

PHATE is a particularly relevant way to perform trajectory inference in scRNA-seq data due to its ability to capture the complex, nonlinear relationships between cells and its robustness to noise and missing data.

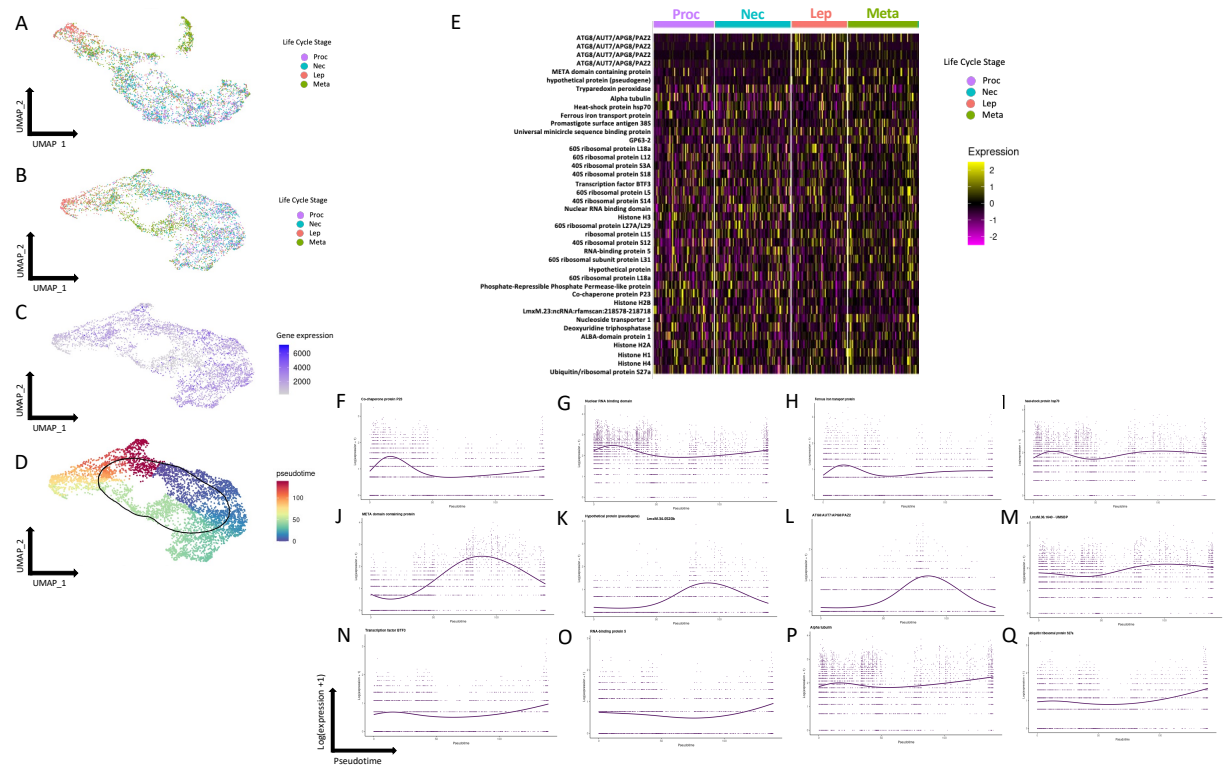


Figure 3-21 Resolving markers differentially expressed in promastigote life cycle stage clusters in Rep2_Pro.

A) UMAP of promastigote clusters in Rep2_Pro sample labelled with promastigote stage markers (procyclic promastigotes in purple, nectomonad promastigotes in turquoise, leptomonad promastigotes in red and metacyclic promastigotes in green) as in Figure 3-18. **B**) Labelled promastigote clusters in **A**) replotted in a UMAP using variable genes in labelled promastigotes with enrichments for each lifecycle stage. **C**) Raw counts of replotted Rep2_Pro UMAP in **B**) with gene expression in purple. **D**) UMAP of promastigote life cycle stages replotted as in **B**) coloured by assigned pseudotime value. The inferred circular trajectory represented by the black circular line starting in procyclics and ending in metacyclics. **E**) Heatmap of top ten distinguishing markers differentially expressed in labelled life cycle phase populations (log2 normalised z-score), coloured as in **A**). **F** – **Q**) Promastigote life cycle pseudotime (x-axis) against expression levels (y-axis; log2(expression + 1)) of selected life cycle markers present in top ten variable markers with smoothed average expression indicated by purple line. Each point represents expression level for one cell's transcriptome. Marker **F**) Co-chaperone protein P23; LmxM.34.4470, **G**) Nuclear RNA binding domain; LmxM.31.075, **H**) Ferrous iron transport protein; LmxM.30.3070, **I**) Heat-shock protein hsp70; LmxM.28.2770, **J**) META domain containing protein; LmxM.17.089, **K**) Hypothetical protein (pseudogene); LmxM.34.0520b, **L**) ATG8/AUT7/APG8/PAZ2; LmxM.19.0870, **M**) Universal minicircle sequence binding protein; LmxM.36.1640, **N**) Transcription factor BTF3; LmxM.36.3770, **O**) RNA-binding protein 5; LmxM.09.0060, **P**) Alpha tubulin; LmxM.13.0390, **Q**) Ubiquitin/ribosomal protein S27a; LmxM.14.1270.

Shown above in Figure 3-42 A) are the previously labelled promastigote life cycle stages for the Rep2_Pro replicate, with life cycle stages coloured as previously in Figure 3-18. This dataset was then replotted in a UMAP based on life cycle stage markers labels, showing a loop of cells as shown in **B**. Some enrichment for each of the life cycle phases is seen when the replotted cells are relabelled by life cycle phase in UMAP, shown in **B**, with procyclic and nectomonad cells collecting in the top and right portion of the replotted cells, following round to leptomonads cells on the left edge and ending in metacyclic promastigotes towards the top of the loop. Gene expression counts shown for the replotted

cells in **C**, where gene expression is coloured in purple. A trajectory is inferred in this replotted UMAP using the `principal_curve()` function, drawn in a black line, and pseudotime values assigned along the trajectory, as shown in the UMAP in **D**. The circular trajectory begins at the top of the replotted loop of cells (indicated in blue pseudotime colouring) and continues clockwise around finishing the loop where it started (indicated in red). These life cycle marker genes were then grouped by phase, reordered by life cycle progression, and replotted on a Heatmap in **E**, where relative expression levels (log₂ normalised z-score) were assigned to each gene and the top ten markers were plotted with each column representing the expression profile of a cell. For differentially expressed genes assigned to each life cycle phase, 717 markers for cells in procyclic promastigotes were found, with 396 for nectomonads, and 79 for leptomonads, and a further 126 markers for metacyclic promastigotes. Amongst the top ten markers for each phase via Heatmap reveals promastigote life cycle stage markers matching expression profiles for their respective stages, such as 4 forms of the ATG8 marker used for leptomonad identification and enriched in expression here, and the META domain containing protein and GP63-2 seeing upregulation in the metacyclic labelled cells. 2 hypothetical proteins and 5 histones are present, and interestingly 12 genes associated with ribosomal formation.

Other notable cell cycle markers include the same pseudogene and ncRNA as previously seen in other heatmap analysis, a Nuclear RNA binding domain gene, Co-chaperone protein P23, RNA-binding protein 5, Heat-shock protein hsp70, and the ALBA-domain protein 1 previously being shown as being required for proteome remodelling during trypanosome differentiation and host transition (Bevkal Id et al., 2021). Changes in expression profiles along the inferred trajectory was further confirmed when looking at smoothed expression plots along pseudotime for a selection of life cycle associated markers and hypothetical proteins enriched in replicating clusters which demonstrated higher expression, shown in **F - Q**, where the $\log(\text{expression} + 1)$ on the y-axis for each selected gene is plotted, and each dot represents the expression for an individual cell, with its smoothed expression (indicated by the purple line) plotted across the pseudotime trajectory on the x-axis

Following this analysis, the Rep2_Pro sample was then used for trajectory visualisation and analysis along promastigote differentiation using a Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE) map. Using PHATE, cells were again replotted onto a map to visualise the complete branching trajectory structure in high-dimensional data, plotting onto two-dimensional space, while retaining global and local structure associated with the developmental progression of gene expression found within the object. This PHATE map can then be used for pseudotime ordering of trajectories to further group matching expression profiles along the newly inferred trajectory (Moon et al., 2017). A PHATE map for the Rep2_Pro replicate was produced and shown below in Figure 3-22.

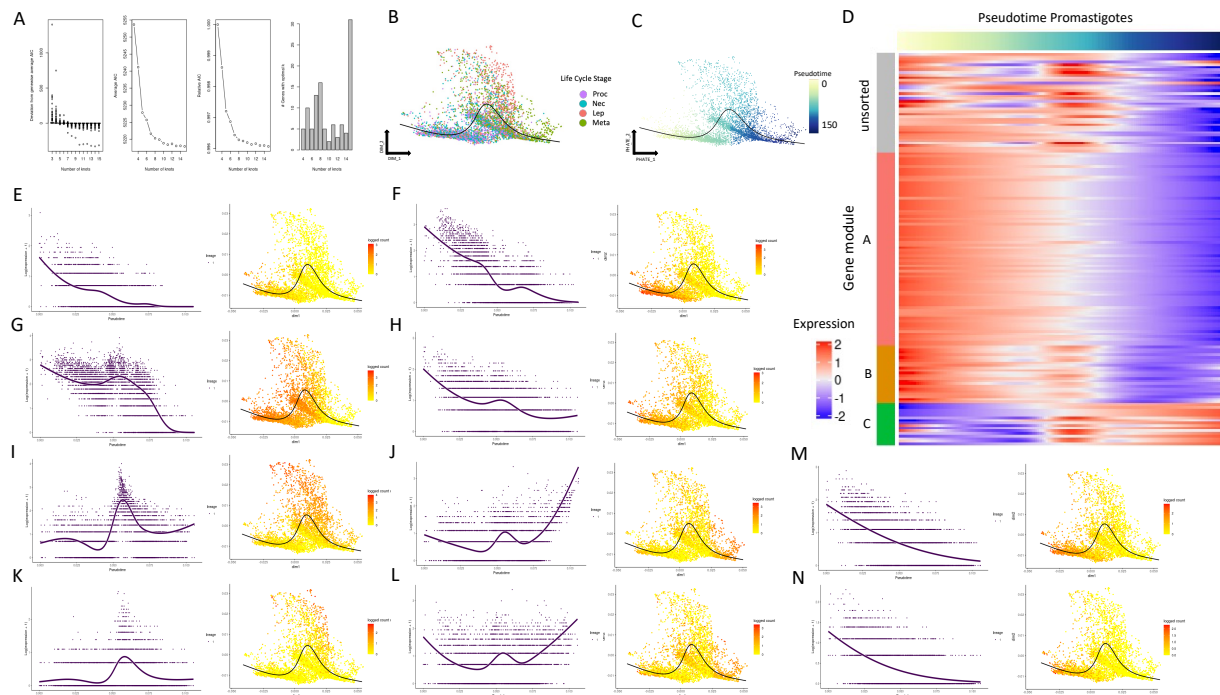


Figure 3-22 Expression of promastigote marker genes across pseudotime trajectories.

A) TradeSeq negative binomial generalized additive model. The models are fit by running the **fitGAM** function. By default, the gene-wise NB-GAM estimates one smoother for the lineage using the negative binomial distribution and then run through diagnostic plots using the Akaike Information Criterion (AIC). Number of knots compared to deviation from genewise average AIC, Average AIC, Relative AIC and number of genes with optimal k . 7 knots are chosen as the best fit to optimally select for lowest number of knots and the greatest number of genes with optimal k . **B)** PHATE plots of Rep2_Pro *L. mexicana* transcriptomes where cell are ordered across a pseudotime by comparing expression patterns. Coloured by cluster identity as in Figure 3-18. Trajectory through cells represented by the black line, starting on the left in procyclic promastigote cells (purple) and ending on the right in metacyclic promastigote labelled cells (green). **C)** PHATE map plot of cells on trajectory coloured by pseudotime from yellow to blue, starting in procyclics and ending in metacyclics. **D)** Heatmap plot of 3 gene modules in life cycle progression along pseudotime (top from yellow to blue) showing relative expression levels (log₂ normalised z-score) of genes with differential expression significantly related with the pseudotime trajectory (as determined by associationTest; p -value > 0.05, FC > 2). Unsorted genes assigned to a separate gene module, coloured in grey. **E) – N)** Trajectory expression plots across pseudotime with logged expression by PHATE plots for promastigote life cycle markers. Included for each marker are trajectory pseudotime plots (x -axis) against expression levels (y -axis; log₂(expression + 1)) of promastigote life cycle stage markers with smoothed average expression of pseudotime indicated by the purple line. Each point represents expression level for one cell's transcriptome placed in order of pseudotime. To the right of smoothed expression plots are logged transcript counts assigned to PHATE plots, coloured by over expression in red against cells not expressing in yellow. **E)** histone H2B; LmxM.19.0050, **F)** GP63; LmxM.10.0460. **G)** nucleoside transporter 1; LmxM.15.1240, **H)** ALBA-domain protein 1; LmxM.13.0450, **I)** META domain containing protein; LmxM.17.0890, **J)** promastigote surface antigen 38S; LmxM.12.0980, **K)** ATG8/AUT7/APG8/PAZ2; LmxM.19.0870, **L)** GP63-2; LmxM.10.0470. Markers **M)** & **N)** are included as the top 2 annotated genes with the highest variance at the start of the trajectory - **M)** polyadenylate-binding protein 1; LmxM.34.5040 and **N)** eukaryotic initiation factor 5a; LmxM.25.0730.

Shown in Figure 3-22 **A)** is the output of the fitGAM() function from the tradeSeq package (van den Berge et al., 2020), to determine the number of knots, or sections, across the trajectory. Number of knots (k) is calculated by comparison to deviation from genewise average Akaike Information Criterion (AIC), Average AIC, Relative AIC and number of genes with optimal k via generalized additive models (GAM) contained within the package. Where the lowest number of knots

while providing the highest number of genes with optimal k should be selected. Here, 7 knots were selected for the trajectory across the PHATE map, shown in **B**. The PHATE shows grouping of both procyclic and nectomonad cells (coloured in purple and turquoise, respectively) together in the bottom left of the PHATE map, with progression along the linear trajectory (indicated by the black line) up to leptomonad cells (coloured in red) centrally and above in the plot, followed by the trajectory, before finishing in a grouping of metacyclic cells in the bottom right of the PHATE map. The same PHATE map is shown in **C** coloured by pseudotime, progressing from yellow to blue. In total, the expression of 2140 genes were found to be associated with this trajectory, where they were differentially expressed as a function of pseudotime (p -value < 0.05 , fold change > 2). These were then assigned into three modules (A-C) based on similarity of expression profiles. Those not grouped together were assigned as being unsorted (coloured in grey). These modules were then plotted together on a heatmap, shown in **D**, with pseudotime running from left to right, indicated on top by pseudotime colouring from yellow to blue, as in **C**, and ordered manually by module. Of note, is the trend across pseudotime for the three modules, with overlap seen in the centre of the trajectory being modules B and C. Possibly indicating genes that may be involved with activation of genes associated with metacyclic commitment and formation. Genes cluster in module A see a progressive decline in expression across pseudotime, indicating transcripts that are reduced in expression as the promastigotes move from procyclic to metacyclic forms, seemingly reduced below average expression around the leptomonad form, as indicated in **B**.

Genes plotted in **E - L** were selected from top markers derived from previous trajectory inference, where the $\log(\text{expression} + 1)$ on the y -axis for each selected gene is plotted, and each dot represents the expression for an individual cell, with its smoothed expression (indicated by the purple line) plotted across the pseudotime trajectory on the x -axis. Next to each smoothed expression plot is also the matching gene count plotted onto a PHATE map with logged counts of the expressed gene indicated as an overexpression coloured in red and no expression in yellow. Genes **M -N** were selected as two of the top 4 markers differentially expressed along pseudotime with high expression at the start of the trajectory, the other two markers being hypothetical proteins (not

shown here), with matching gene count plots adjacent, as detailed below in Section 3.3.4.

3.3.4 Single cell RNA-sequencing analysis for second, axenic amastigote only, replicate

Following pseudotime analysis of the Rep2_Pro sample, further investigations were made with the Rep2_Axa replicate. Mapping of reads generated for the Rep2_Axa only sample, as described above in Section 3.3.1,. Sample data with timepoints shown for each combined sample is shown below in Figure 3-23.

3.3.4.1 Experimental design of single cell RNA-sequencing for second axenic amastigote only replicate

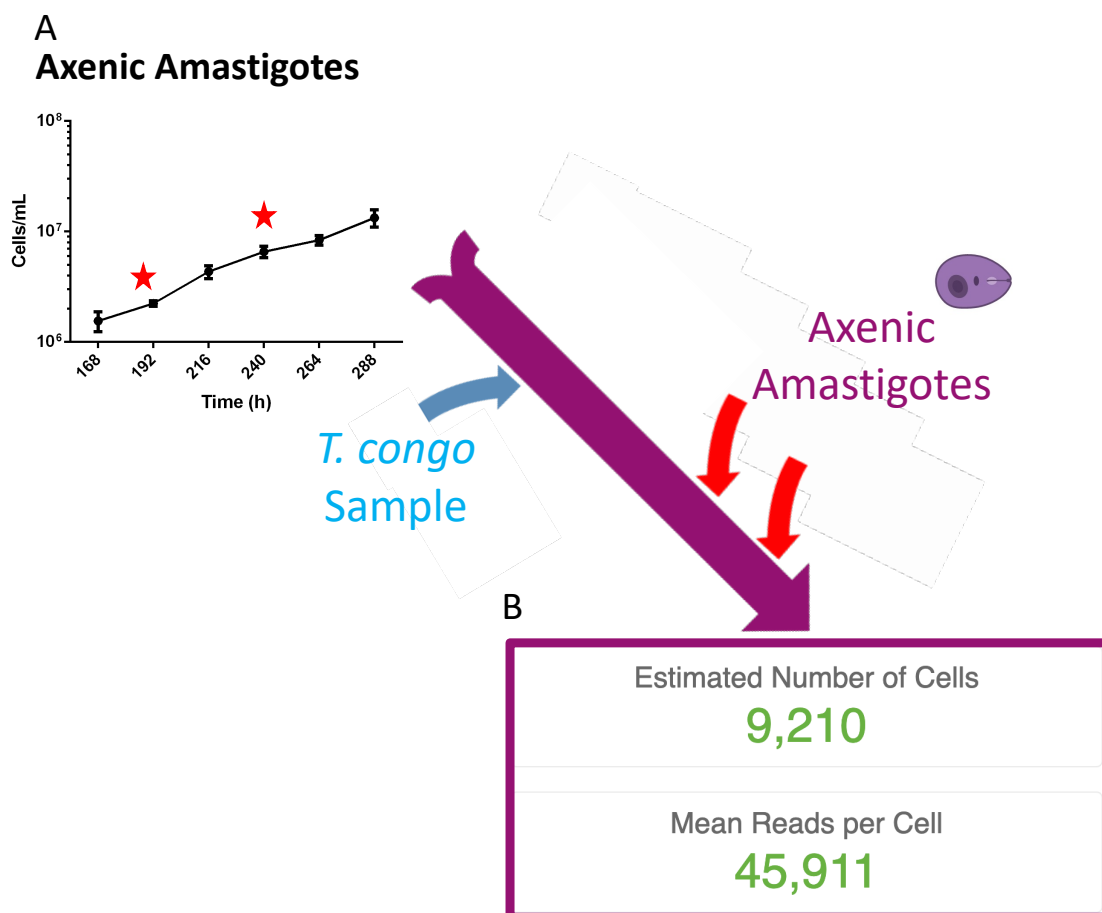


Figure 3-23 Experimental sample collection and sample information for second replicate of axenic amastigotes timepoints.

Experimental set-up for second replicate *Leishmania* sample run in single cell RNA-Sequencing over 2 timepoints (red stars) across axenic amastigote life cycle progression (**A**). Replicate Rep2_Axa contained 2 timepoints with 3000 cells taken from each timepoint with the aim to capture cells still developing from one life cycle to the next. *Leishmania mexicana* samples were mixed with *Trypanosoma congo* cells to estimate multiplets in sequencing sample. **B**) In total, 9,210 cells were sequenced with mean reads per cell of 45,911.

Here, the combined samples of *L. mexicana* and *T. congolense* contained an estimated 9,210 cells with a combined mean read per cell of 45,911. As explained in Section 3.3.1, following the generation of reads, quality control of the samples was run to limit the impact of noise generated due to biological and technological variations (Kolodziejczyk et al., 2015), as described above. Quality control was performed as for previous samples, with quality control plots for the Rep2_Axa sample can be found in Appendix I.

The top 9 PCs were used for clustering cells by likeness before determining the best resolution to group cells into biologically relevant clusters, shown below in Figure 3-24

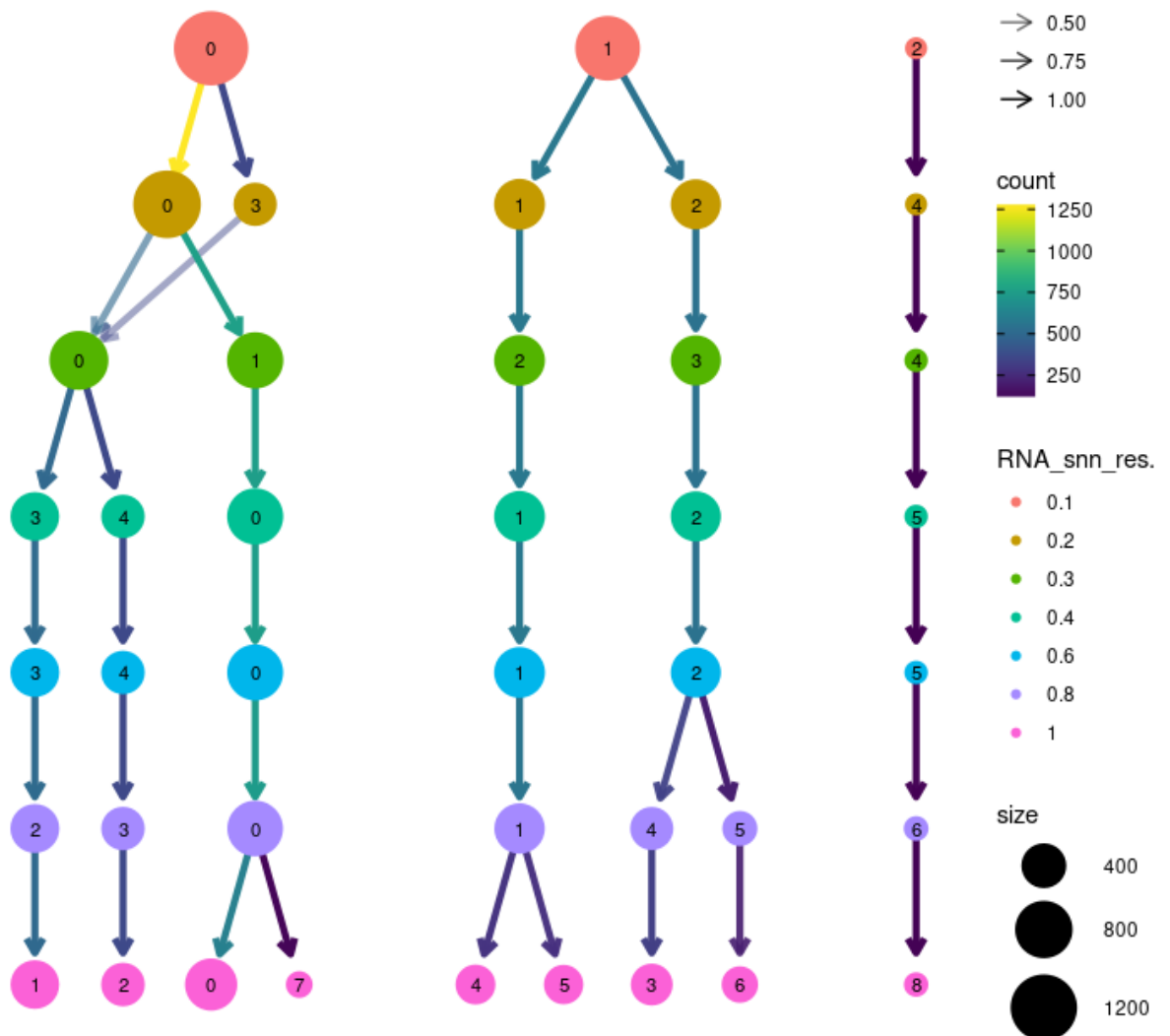


Figure 3-24 Clustree plot demonstrating Rep2_Axa sample movement with increasing cluster resolution.

Clustree plot for Rep2_Axa sample showing changes in sample movement as resolution increases (RNA_snn_res). A resolution of 0.2 is chosen resolving into 5 clusters for further clustering analysis. Proportion of cells moving from between clusters as resolution increases (from top to bottom) depicted by colour of arrows, and size of cells within clusters indicated by the relative size of depicted circles.

Shown in Figure 3-24 is Clustree analysis, a resolution of 0.2 (coloured in ochre) partitioned the dataset into 5 distinct clusters. Although 2 timepoints were contained in the Rep2_Axa sample, the top ten variable features indicated a possible late-stage promastigote population, therefore a resolution of 0.2 was selected for further clustering analysis, explored below in Section 3.3.4.2 to consider the possibility of a translational population still differentiating between late promastigote and early amastigote forms.

3.3.4.2 Individual clustering analysis of single cell RNA-sequencing for second axenic amastigote only replicate

Cluster of variable features was performed as for previous samples, described above. Plotting via UMAP reveals the clustering patterns as displayed below in Figure 3-25.

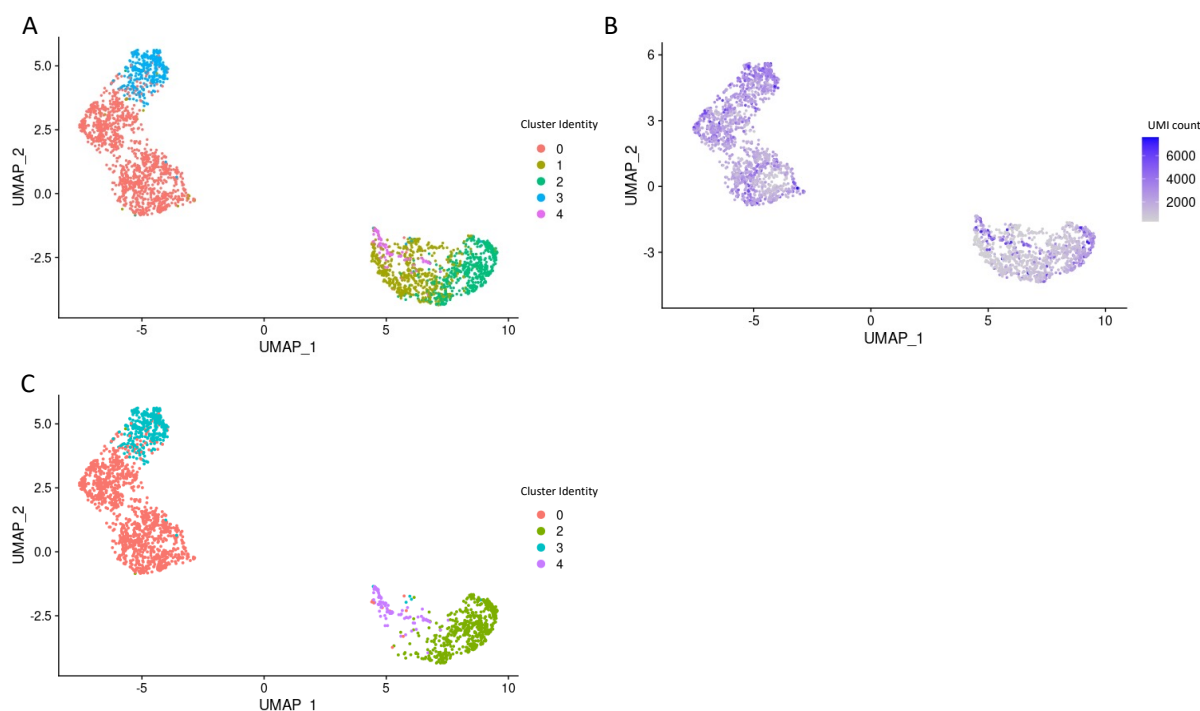


Figure 3-25 Clustering of *L. mexicana* transcriptomes across life cycle development in newly developed axenic amastigotes.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in Rep2_Axa sample to visualise relative relationships between individual transcriptomes. **A)** UMAP of Rep2_Axa sample at a resolution of 0.2 reveals 5 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** UMAP of Rep2_Axa sample as in **A)** with low transcript cluster (cluster 1 in ochre) removed for further marker analysis.

Preliminary clustering shown in Figure 3-25 **A)** revealed five distinct clusters, with clusters 0 and 3 (coloured in red and blue, respectively) separated from the other three clusters (cluster 1 in ochre, 2 in green, and cluster 4 in purple). In **B)**

the raw gene expression count is displayed for each cell, coloured in purple for cells with higher expression counts. Of note is the very low counts in cluster 1 (ochre in **A**). Following reiterations of analysis of clusters, and biologically relevant markers found within them, cluster 1 was removed from the dataset in the downstream marker analysis, giving a UMAP with 4 clusters, as shown in **C**, due to no significant markers being found during DE analysis. Where cluster 0 is coloured in red, 2 in green, 3 in turquoise and 4 in purple.

After clustering, each cluster was assigned an identity using known cell type markers. As before, 12 markers are shown to determine axenic amastigotes and replicating clusters.

The same marker analysis used in Rep1 as described in Section 3.3.2.2 is used for comparison below for Rep2_Axa in Figure 3-26.

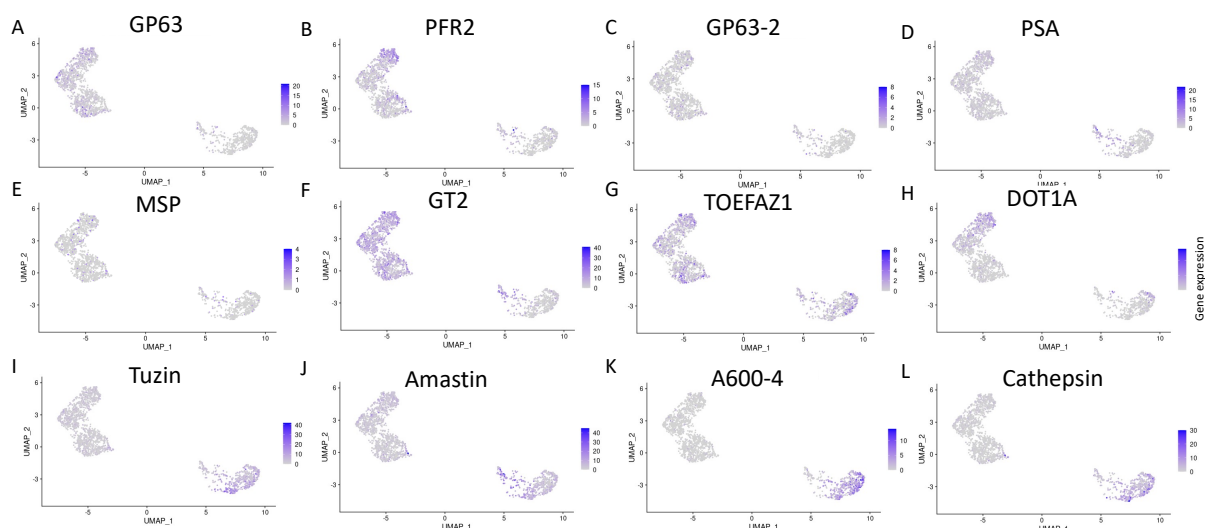


Figure 3-26 UMAP life cycle marker plots in Rep2_Axa scRNA-seq sample to determine stages defined in clustering analysis.

Uniform Manifold Approximation and Projection (UMAP) of Rep2_Axa sample coloured by selected life cycle stage specific markers. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F**) GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** – **H**) (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L**) (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

In Figure 3-26 promastigote markers from **A**, **B**, **D** and **G** showed enrichment across cluster 3 (coloured in turquoise in Figure 3-26, **C**). Proliferative markers TOEFAZ1 and DOT1A in **G** and **H** also showed enrichment in cluster 3 with little comparative expression seen in other clusters. Interestingly, some expression of

metacyclic markers in **D** and **E** was also seen in cluster 4 (Figure 3-26, **C**). This further indicates capture of a population of promastigotes in the process of differentiating into amastigotes, likely captured within the 24 h timepoint after passage into amastigote culture conditions. Amastigote markers in **I** to **L** saw expression in cluster 2 (green in Figure 3-26, **C**) with Amastin in **J** also mostly expressed in cluster 4, potentially indicating a metacyclic population preparing for differentiation into amastigotes.

To demonstrate these data in another representation Violin plots were also produced as previous analysis, shown below in Figure 3-27.

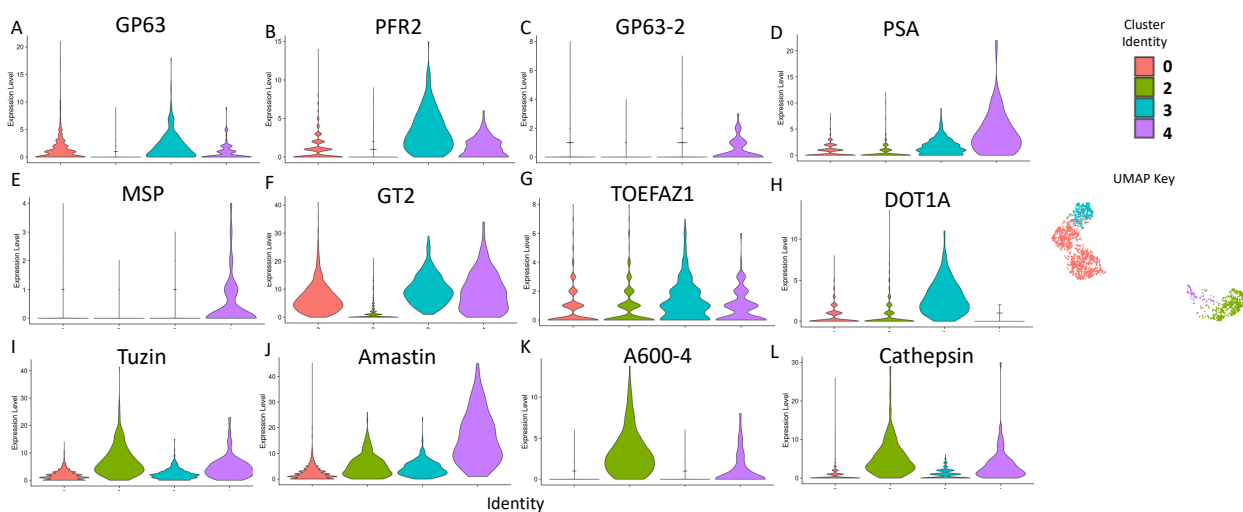


Figure 3-27 Violin plots of selected marker genes in Rep2_Axa scRNA-seq sample.

Violin plots of Rep2_Axa sample divided and coloured by clusters. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F**) GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** – **H**) (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L**) (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

As shown in Figure 3-27, expression trends follow those demonstrated in Figure 3-26, where some promastigote expression in **A** - **F** was seen in clusters 0, 3 and 4, with little to no expression present in cluster 2. Conversely, cluster 2 showed high expression of amastigote markers in **I** - **L**, with amastin also being highly expressed in cluster 4. Cell cycle markers **G**- **H** were mostly expressed in cluster 3.

3.3.4.3 Cell cycle analysis of second replicate in axenic amastigote stages

In order to further elucidate possible cell cycle based regulation of life cycle development in *L. mexicana* using this scRNA-seq data, the same method of cell cycle stage labelling developed by Briggs *et al.* (2021b) was recapitulated and applied again here for the Rep2_Axa sample, as described above in Section 3.3.2.4., shown in Figure 3-28.

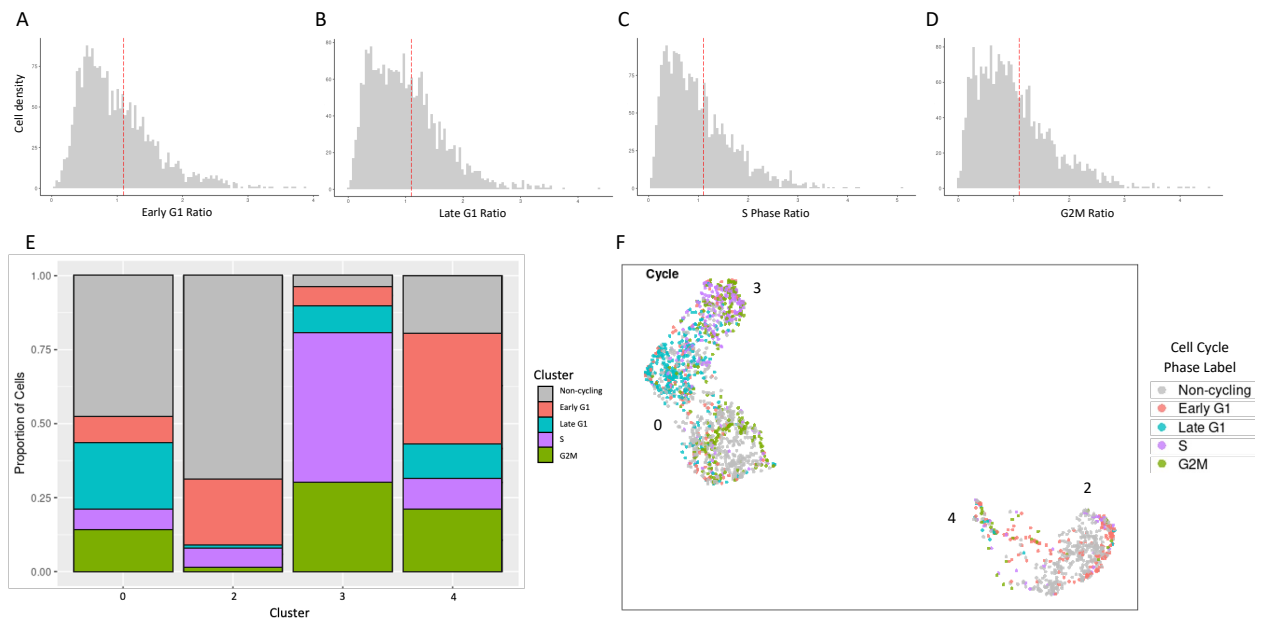


Figure 3-28 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in sample Rep2_Axa.

A) – D) ratio of orthologous cell cycle progression markers from *T. brucei* with no cell cycle stage specific marker determined by those not over-expressing any cell cycle stage specific markers below a 10% ratio cut-off, indicated by the red dashed line. **E)** Proportion of cells labelled by cell cycle phase orthologous markers across clusters, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **F)** UMAP of cells in Rep2_Axa sample assigned and coloured by cell cycle stage, as in **E)**.

Assigning cell cycle stages to cells over-expressing these markers and labelling each phase by colour, as shown in Figure 3-28 F), again revealed distinct groupings of cells enriched for different cell cycle phases. As shown in E, Cluster 3 showed the largest cell cycle labelling with 95.2% of cells being labelled, 49.8% of these in the S phase (coloured in purple), which indicated an actively proliferative cluster potentially confirming a replicating, or alternatively a S phase rich, promastigote-like cluster this sample, associated with cells still differentiating into amastigotes. Cluster 2 saw the least labelling with approximately 66.0% labelled as non-cycling (coloured in grey). Of note is that this most amastigote like cluster was also enriched in the Early G1 cell cycle phase (coloured in red, 21.8%), as indicated in Rep1 cell cycle labelling (Figure 3-11). Cluster 0 showed 51.3% labelled as non-cycling with most cells in late G1

(coloured in turquoise, 21.0), indicating again the possible presences of cell cycle arrest forms, before fully committing to differentiation into amastigotes, as seen in the Rep2_Pro late life cycle, metacyclic forms. And lastly, cluster 4 being highly labelled with all cell cycle markers (79.6%), labelled mostly as Early G1(35.2%).

3.3.4.4 Marker analysis of second axenic amastigote only replicate

Following cluster identity investigations, a UMAP was replotted with life cycle stages assigned (Figure 3-29, A)), named Pro Rep for the cluster showing both promastigote and S phase markers (turquoise), Post Pro for the promastigote like cluster following the Pro Rep cluster (red), Pre-Axa for the smallest cluster (purple) proximal to the axenic amastigote cluster (green). Markers are generated, as previously described in Section 3.3.2.3, shown below in Figure 3-29.

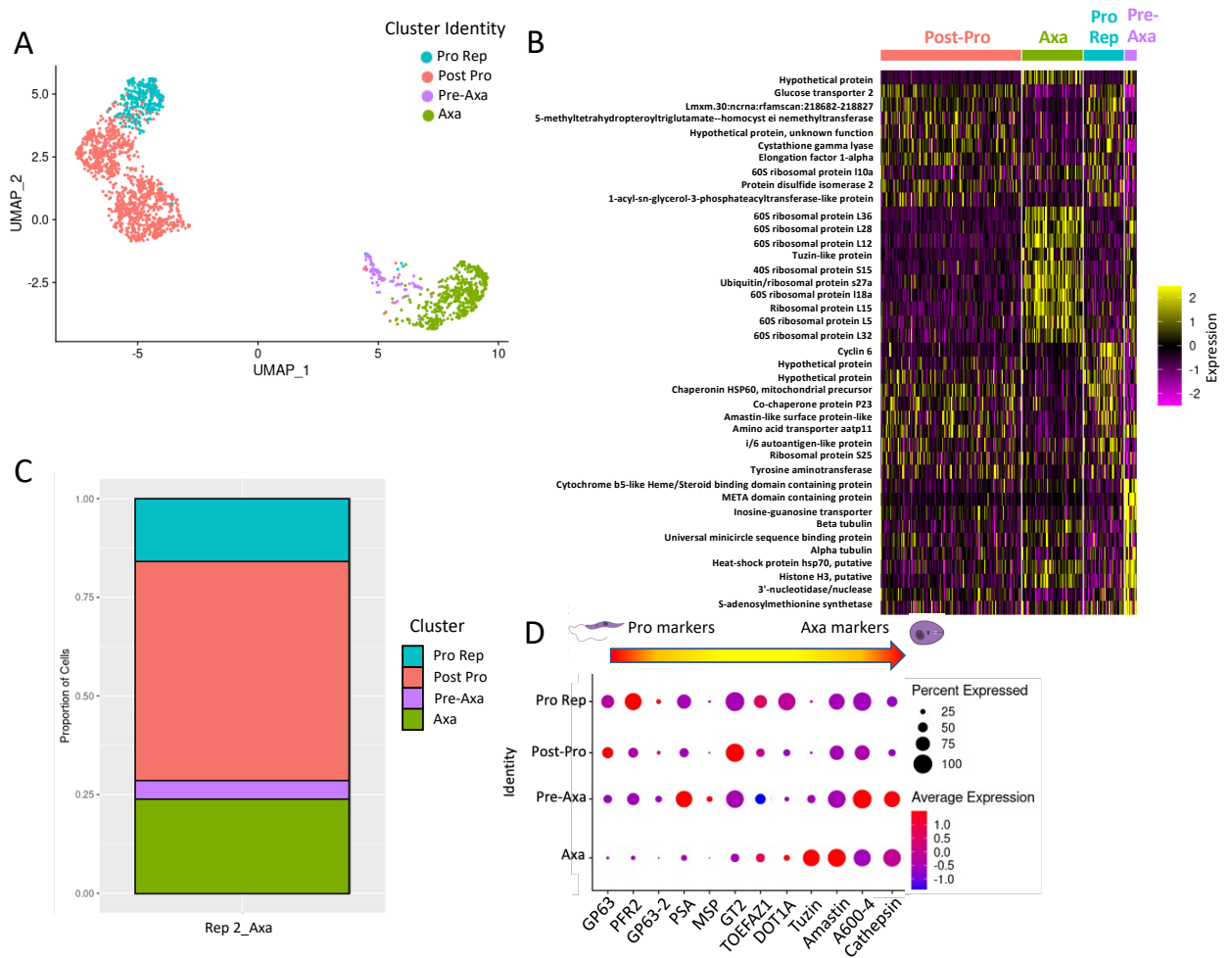


Figure 3-29 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in Rep2_Axa sample.

A) labelled clusters by life cycle stage as determined by expression patterns of life cycle stage specific markers. **B)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Dot plot of marker genes for life cycle specific stages ordered from left to right for promastigote to amastigote, with life cycle progression for clusters ordered from top to bottom. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot.

Gene expression scores for the top ten expressed markers shown in Figure 3-29 **B)** are ordered by size of the clusters, from left to right annotated on top with each column representing the expression profile of a cell, showing distinct expression profiles for each cluster, with some expected overlap of expression profiles in Post-Pro and Pro Rep clusters and some overlap in Axa and Pre-Axa clusters.

In total, 4393 differentially expressed markers were found in this dataset: 153 differentially expressed markers were found for the Post Pro cluster, 31 of them being hypothetical proteins; 331 markers for the Axa cluster with 31 hypothetical proteins; 1784 for Pro Rep with 613 being hypothetical proteins; and 2210 markers for the Pre-Axa cluster with 831 being hypothetical proteins.

Shown in **B**), two of the top ten differential markers for the Post-Pro cluster are made up of hypothetical protein genes being expressed, the top marker being a hypothetical protein which is an orthologue for amastin in other *Leishmania* species and down-regulated in both Post-Pro and Pro Rep clusters, while conversely being up-regulated in both amastigote-like clusters. The promastigote marker GT2 also make the top ten for the Post-Pro cluster as well as the same ncRNA seen in the previous Rep2_Pro sample. Two transferases, a ribosomal protein, Cystathione gamma lyase, Elongation factor 1-alpha and Protein disulfide isomerase 2 making up the other markers for the Post-Pro cluster. For the Axa cluster nine out of the top ten markers are ribosomal proteins, with two of them also being present in the top 10 markers of Axa clusters in Rep1. The other marker being the Tuzin-like protein previously used as an amastigote marker. The top marker for Pro Rep in Cyclin 6, a well-studied cell cycle marker, two hypothetical proteins and Chaperonin HSP60, mitochondrial precursor also seen as a top marker for amastigote clusters in Rep1. Also present are the markers Co-chaperone protein P23, Amastin-like surface protein-like, Amino acid transporter aatp11, i/6 autoantigen-like protein, Ribosomal protein S25, and Tyrosine aminotransferase. And for the Pre-Axa cluster, unexpectedly, six out of the top ten markers are seen in the top markers for the Rep2_Pro sample. These six markers being Cytochrome b5-like Heme/Steroid binding domain containing protein, META domain containing protein, Inosine-guanosine transporter, Universal minicircle sequence binding protein, Alpha tubulin, and Heat-shock protein hsp70, putative. Also present in the top markers for his cluster was the marker Histone H3, putative shown previously as a top marker for the Axa Rep cluster in Rep1. And markers 3'-nucleotidase/nuclease and S-adenosylmethionine synthetase are the last two markers for the Pre-Axa cluster. Lastly, the Dot plot, shown in **D**) shows a general trend from top left to bottom right of promastigote to amastigote markers as expected with clusters ordered in a proposed life cycle transition order. Some promastigote markers are shown to be in both Pro-Rep and Post-Pro markers with metacyclic markers PSA and MSP seen in the Pre-Axa cluster. With a large enrichment for amastigote markers seen in Axa and Pre-Axa clusters, as expected.

3.3.5 Single cell RNA-sequencing analysis of integrated *L. mexicana* replicates

Following analysis of the replicate scRNA-seq samples individually, the samples were then integrated into one object for further analysis as described below. As previously for the samples when analysed individually, Cell Ranger outputs were read for each replicate, as explained in Section 3.3.2.1, and quality control filters applied.

3.3.5.1 Quality control filtering of single cell RNA-sequencing integrated replicates

Outputs from Cell Ranger were then read, as explained for Rep1 in Section 3.3.2.1, with all consistent pre-processing quality controls for removing multiplets and contaminating transcripts applied.

Shown in Figure 3-30 below were the quality control filters considered for the number of unique features, UMIs, kDNA and rRNA contaminates, as performed for previous individual replicates.

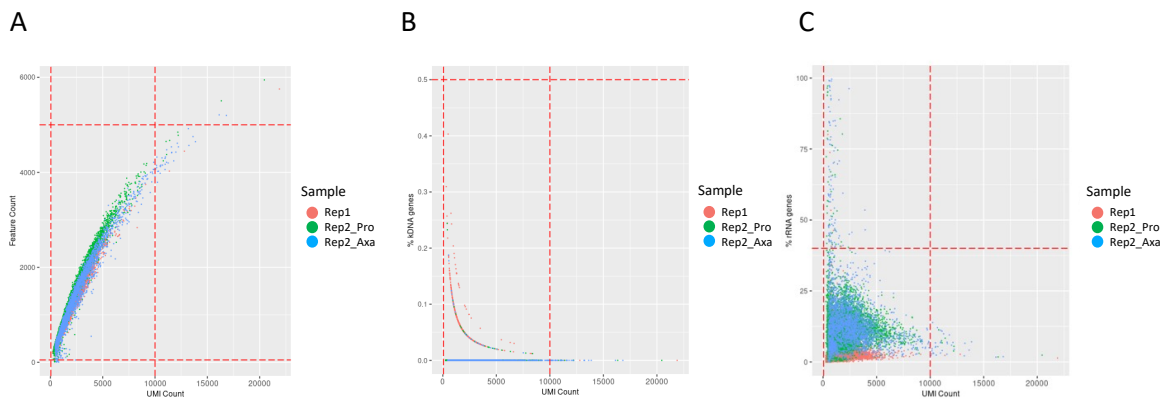


Figure 3-30 Quality control and filtering of integrated transcriptomes for unique molecular identifiers, feature counts, ribosomal RNA and kinetoplast DNA.

Scatter plots for integrated biological replicates across five timepoints. Five timepoints in replicate 1 from promastigote to axenic amastigote (red), three promastigote timepoints (green) and two axenic amastigote timepoints (blue) in biological replicate experiments. Each data point represents the captured transcriptome of one cell. Plots show the cut-offs for **A**) the Unique Molecular Identifiers (UMI) <5000, >50, against number of features (genes), **B**) percentage of features for ribosomal RNA (rRNA) <40% and **C**) percentage of features encoded on the kinetoplastid (kDNA) maxi circle genome added to mapped reads serving as an equivalent of the mitochondrial quality control for dying cells. Red dashed lines indicate cut-offs used to filter cells per experiment.

In Figure 3-30 A) quality control cut-offs were applied, as described previously for individual sample analysis.

Once quality control cut-offs had been determined and applied to the individual Seurat objects to remove unwanted cells, normalisation using the SCRAN package was applied as for previous individual samples. Replicates were then integrated with the data and regressed variables scaled to total RNA. An Elbow plot was then used to guide dimensionality of the integrated samples, as shown below in Figure 3-31.

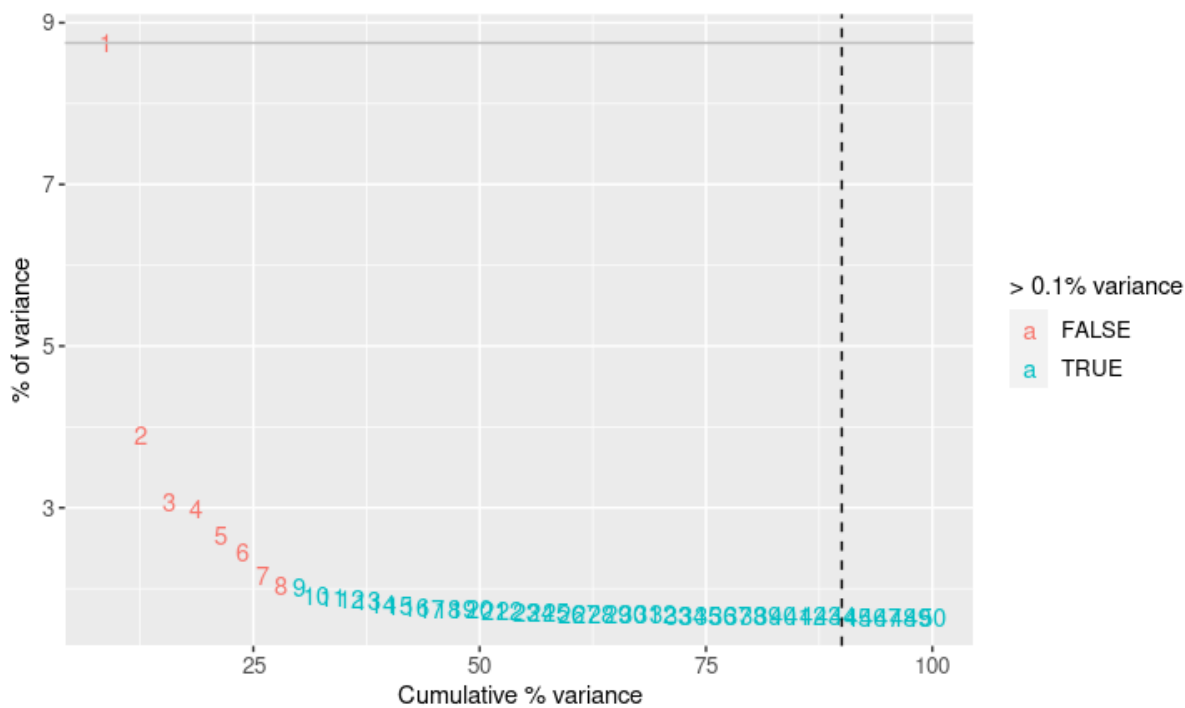


Figure 3-31 Elbow plot of variance found in principle components observed in integrated Rep1, Rep2_Pro and Rep2_Axa samples.

Elbow plot to rank principle components (PC) found in integrated samples assessing the percentage of variance explained by each PC. In these integrated samples, an “elbow” is seen around PC 7-8. Following several iterations of the clustering 7 PCs were chosen giving the most robust clustering.

In Figure 3-31 an ‘elbow’ is present around PCs 7 -8. Following extensive reiterations of the clustering 7 PCs were therefore chosen for the number of dimensions in the dataset for further downstream analysis to provide the most robust biological clustering.

The top 7 PCs were used for clustering cells by likeness using the FindNeighbors() and FindClusters() functions before determining the best resolution to group cells into biologically relevant clusters with the Clustree() function (Zappia & Oshlack, 2018), shown below in Figure 3-32.

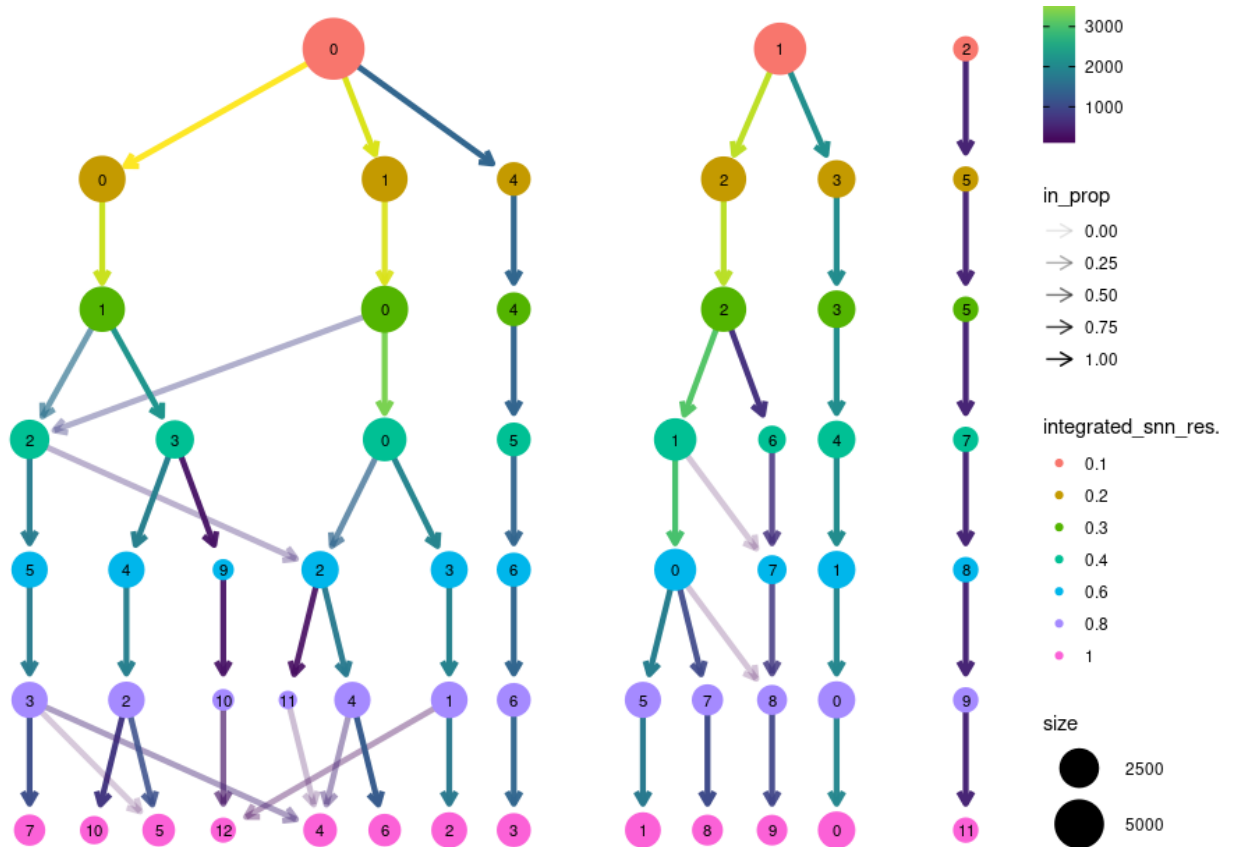


Figure 3-32 Clustree plot demonstrating cell movement in integrated samples with increasing cluster resolution.

Clustree plot for integrated Rep1, Rep2_Pro and Rep2_Axa samples showing changes in sample movement as resolution increases (RNA_snn_res). A resolution of 0.4 is chosen resolving into 8 clusters for further clustering analysis. Proportion of cells moving from between clusters as resolution increases (from top to bottom) depicted by colour of arrows, and size of cells within clusters indicated by the relative size of depicted circles.

Shown in Figure 3-32 using Clustree analysis, a resolution of 0.4 (coloured in lime green) partitions the dataset into 8 distinct clusters. A small number of cells can be seen moving from cluster 0 at a 0.3 resolution to a newly separated cluster 2 at a 0.4 resolution. Clusters resulting from different resolutions were investigated in several iterations of analysis. Here, the results from resolution 0.4 are discussed which resulted in 8 clusters with distinct marker genes. Using life stage markers, already explored in previous individual analysis, clusters are allocated life cycle stages and explored below in Section 3.3.5.2.

3.3.5.2 Clustering analysis of single cell RNA-sequencing for integrated sample replicates

As for individual replicate analysis described previously, the results of integration and clustering was visualised with UMAP in Figure 3-25.

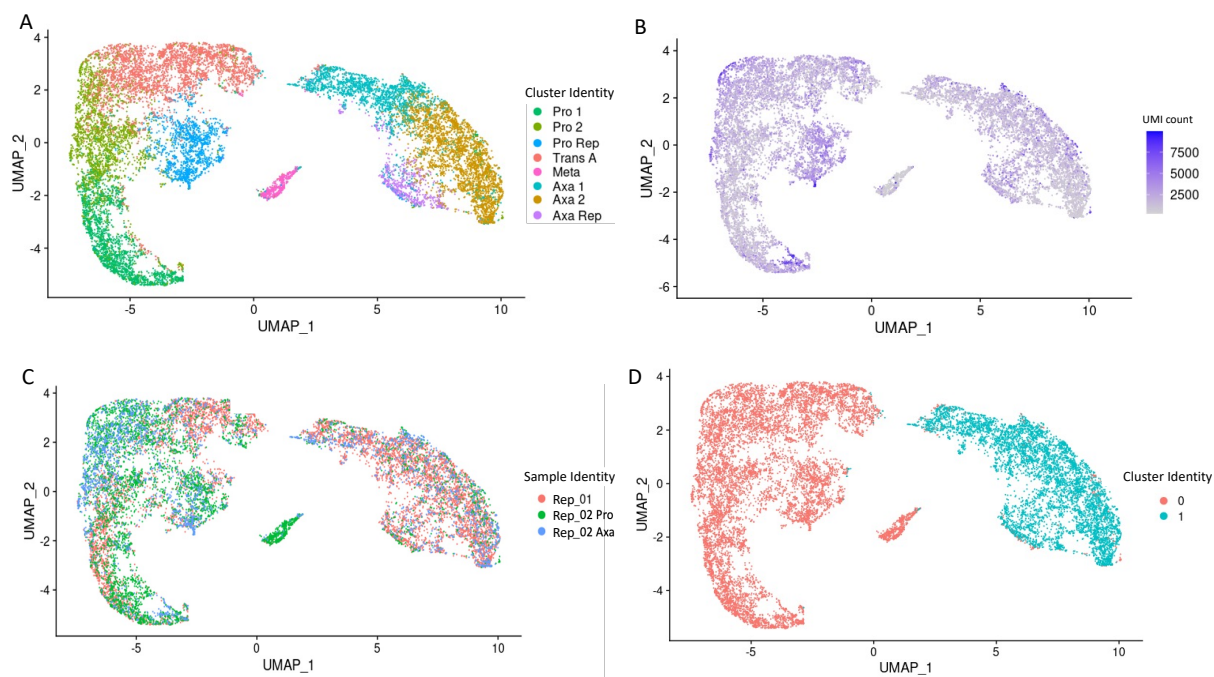


Figure 3-33 Clustering of integrated *L. mexicana* transcriptomes across life cycle stage development from promastigote to axenic amastigote in two biological replicates.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in integrated samples to visualise relative relationships between individual transcriptomes. **A)** UMAP of integrated samples at a resolution of 0.4 reveals 8 distinct clusters labelled using distinguishing markers previously selected for life cycle stages, coloured by cluster: Early promastigote 1 (Pro 1) cluster (lime green), late Pro 2 (olive green), Pro replicating (Pro Rep) cluster (blue), transitional A (Trans A) cluster (red), metacyclic (Meta) cluster (pink), early axenic amastigote (Axa 1) cluster (turquoise), late axenic amastigote (Axa 2) cluster (ochre), and replicating axenic amastigote (Axa Rep) cluster (purple). **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** UMAP of integrated samples coloured by original sample identity; Rep1 (red), Rep2_Pro (green) and Rep2_Axa (blue). **D)** Low resolution (0.025) UMAP of integrated samples to generate only 2 clusters for determining similarity of central Meta cluster to Pro and Axa clusters placed either side at higher resolutions, demonstrating Meta cluster is more similar to Pro clusters (red) than Axa clusters (turquoise).

Preliminary clustering shown in Figure 3-33 A revealed 8 distinct clusters.

Applying previous marker analysis used in individual replicates, clusters were defined by life cycle stages. 4 promastigote clusters were revealed named Pro1 (coloured in lime green), Pro 2 (coloured in olive green), Pro Rep for a possible proliferative cluster (coloured in blue) and Meta (coloured in pink). Three axenic amastigote clusters were allocated and named Axa 1 (coloured in turquoise) Axa 2 (coloured in ochre) and Axa Rep (coloured in purple) for a potentially proliferative amastigote cluster. The same transitional cluster, named Trans A here, is coloured in red and again sat between promastigote and amastigote clusters. Of note is the location of the Trans A cluster in the integrated object, where it is placed proximal to the Pro 2 and Pro Rep clusters. Also of note is the size of the cluster, the biggest in this integrated object. In B the raw gene expression count is displayed for each cell, coloured in purple for cells with higher expression counts. Consistent with previous analysis of the Rep2_Pro

metacyclic-like cluster, very low transcript levels were seen in the Meta cluster (coloured in pink in **A**, consistent with a non-cycling life cycle stage. The Meta cluster also sat noticeably separate from the rest of the promastigote cells, placed between promastigote and amastigote clusters. Shown in **C** are the cells coloured by the replicate they originated from following integration, where Rep1 is coloured in red, Rep2_Pro coloured in green and Rep2_Axa coloured in blue. Here, the Meta cluster is predominately made up of cells from the Rep2_Pro replicate, while the Trans A cluster is made up mostly of the Rep1 and Rep2_Pro replicates, with some cells from the Rep2_Axa replicate present as well, consistent with being an intermediate cell population. Shown in **D** is an UMAP with dimensionality reduced to 0.025 so that cells are grouped into only two clusters. Here we can see that the Trans A cluster, from plot **A**, is combined with the promastigote clusters (red in plot **A**), as seen in the Rep1 sample at lower resolutions (Figure 3-7, **D**). This again potentially suggests, that in terms of profile expression, cells grouped in Trans A are more like promastigote-like than amastigote-like, further indicating an intermediate cell population. Reassuringly, the Meta cluster also clustered with promastigote like clusters. Marker analysis, as for individually analysed samples, was then undertaken, shown below in Figure 3-34.

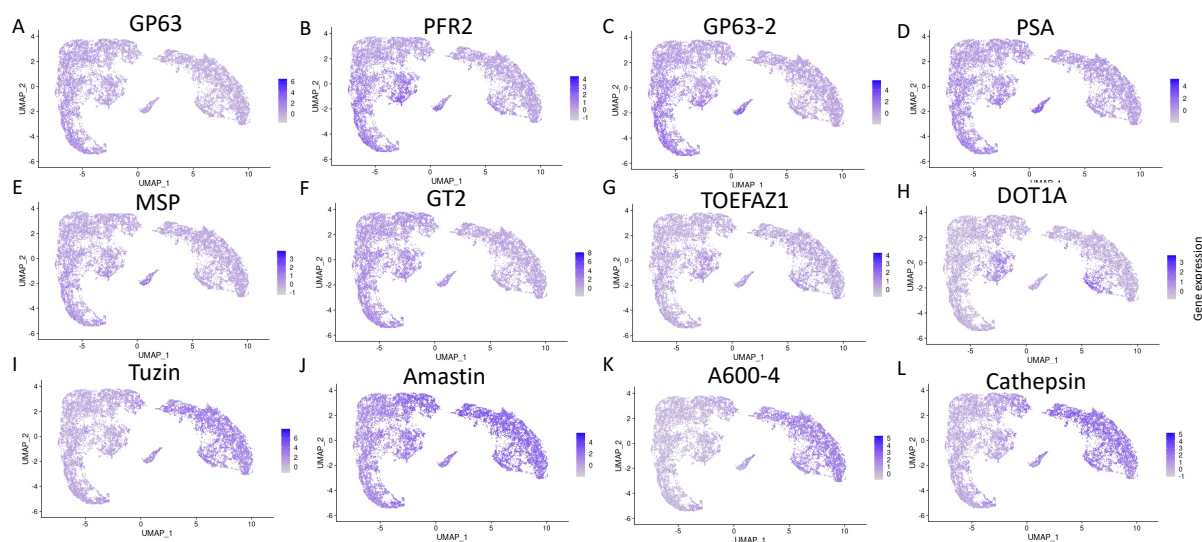


Figure 3-34 UMAP life cycle marker plots in integrated scRNA-seq samples to determine stages defined in clustering analysis.

Uniform Manifold Approximation and Projection (UMAP) of integrated samples coloured by selected life cycle stage specific markers. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F** GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** - **H** (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L** (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

In Figure 3-34 promastigote markers shown in **A** to **F** showed enrichment in the leftmost clusters. proliferative markers TOEFAZ1 and DOT1A in **G** - **H** were enriched in replicating-like clusters placed on the underside of their respective life cycle clusters (as shown in Figure 3-33, **A**). Metacyclic markers **C** - **E** showed enrichment in the centrally placed Meta cluster, as expected. Finally, amastigote markers **I** - **L** were found to be distinctly enriched in the rightmost clusters, suggesting again that cell populations have been clustered by life cycle stages, with the left clusters being promastigotes, the right clusters being axenic amastigotes with a separate metacyclic cluster placed in between. Interestingly, the transitional cluster (as shown in Figure 3-33, **A** coloured in red), positioned between the promastigote and amastigote clusters, again showed overlap of both promastigote and amastigote markers GT2 (in **F**) and the marker Amastin (in **J**). This same cluster was grouped with the promastigote cluster at very low resolutions (Figure 3-33 **D**) and may represent a transitional cell type between the two life cycle stages, distinct from the non-replicative metacyclic promastigotes canonically associated with infecting phagocytotic cells before differentiating into amastigotes.

Violin plots were also produced to give expression probability distributions across clusters, as shown below in Figure 3-35.

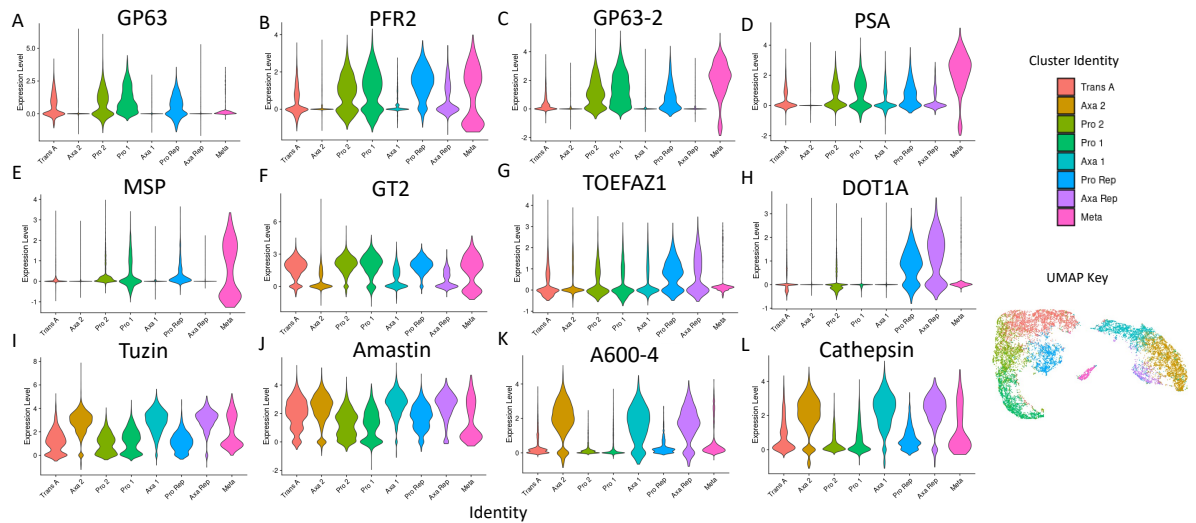


Figure 3-35 Violin plots of selected marker genes in integrated scRNA-seq samples.

Violin plots of integrated samples divided and coloured by clusters. Markers **A** - **E** (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570) are used as markers for promastigote stages, with markers **C** - **E** indicative of the metacyclic promastigote stages. **F** GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **G** - **H** (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **I** - **L** (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

As shown in Figure 3-35, expression trends followed those demonstrated in Figure 3-34, where promastigote markers **A** - **F** were over expressed only in promastigote clusters Pro 1 (coloured in lime green), Pro 2 (coloured in olive green), Pro Rep (coloured in blue) and Meta (coloured in pink) clusters. Conversely, amastigote markers **I** - **L** were expressed in clusters Axa 1 (coloured in turquoise), Axa 2 (coloured in ochre), and Axa Rep (coloured in purple). Again, the Trans A (coloured in red) placed in between the promastigote and amastigote clusters showed overlap in expression for promastigote marker GT2 (**F**) and Amastin (**J**). Proliferative markers TOEFAZ1 and DOT1A in **G** and **H** were most prominently expressed in Pro Rep (coloured in blue) and Axa Rep (coloured in purple). These clusters also express genes from their respective life cycle stages with little relative expression of their opposing life cycle stage. Also of note, were the Meta markers shown in **C**, **D** and **E** being over expressed in the Meta cluster, with some overlap of promastigote marker PFR2 (**B**) and some expression of amastigote markers (**J** and **L**), possibly in preparation for

differentiating in amastigote forms (Alcolea et al., 2009). Further cluster markers were then investigated, as described below in Section 3.3.5.3.

3.3.5.3 Marker analysis of integrated sample replicates

Following cluster identity investigations using known marker genes, differential expression analysis was used to identify all markers for each cluster. The top five differentially expressed markers for each cluster are plotted in a heatmap shown in **A**. The proportion of cells in each cluster, divided by the experiment sample they originated from, are plotted in **B** and coloured by cluster. Also plotted was the expression of promastigote to amastigote life cycle markers, as described above, for each cluster, placed in order of life cycle progression in **C**. These plots are given below in Figure 3-36.

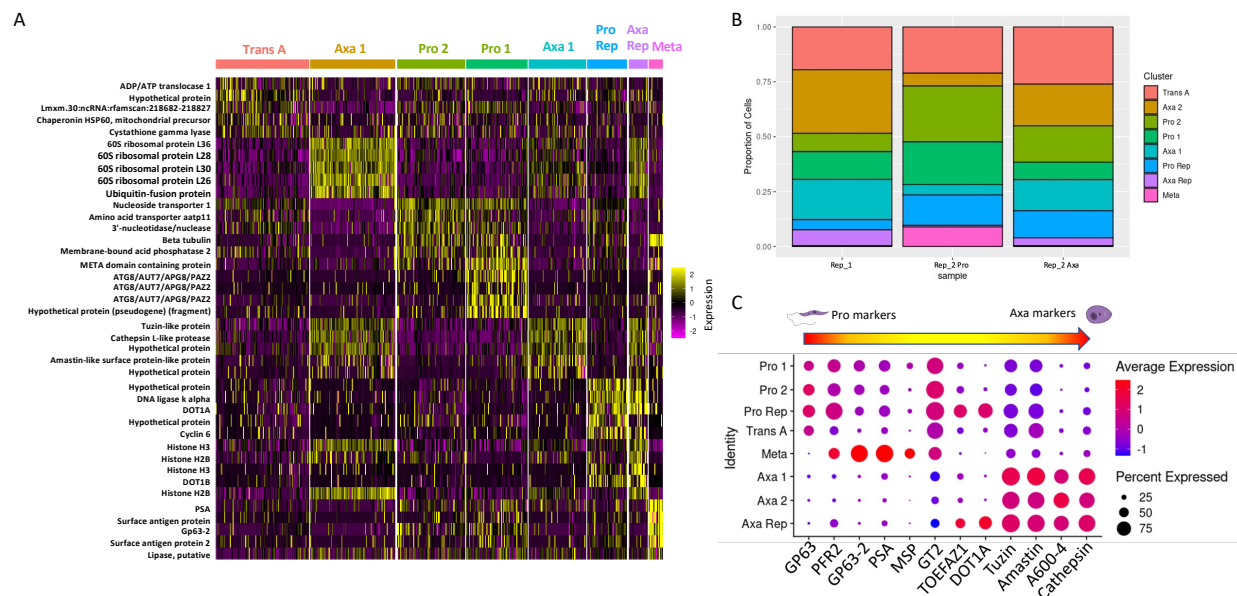


Figure 3-36 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in integrated samples.

A) Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster. **B)** Proportional groupings of cells in each cluster and coloured by cluster, as in Figure 3-33. **C)** Dot plot of marker genes for life cycle specific stages ordered from left to right for promastigote to amastigote, with life cycle progression for clusters ordered from top to bottom. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot.

Gene expression scaled expression for the top five expressed markers shown in Figure 3-36 **A** showed distinct expression profiles for each cluster, with some expected overlap for Pro clusters and Axa clusters. The Meta cluster showed mostly distinct marker expression while still expressing, albeit to a lesser extent, Pro 1, Pro 2 and Axa 1 markers. Interestingly, expression profiles for

Trans A also showed most overlap for the Pro clusters, with some of its top five markers also being expressed to lower levels in the Axa clusters. In total, 1586 differentially expressed markers were found in this dataset; 45 differentially expressed markers were found for the Trans A cluster, 5 of them being hypothetical proteins; 194 markers for the Axa 1 cluster, with 27 hypothetical proteins; 158 markers for Pro 2, with 33 being hypothetical proteins; 232 markers for the Pro 1 cluster, with 77 being hypothetical proteins; 158 markers for Pro 2, with 33 being hypothetical proteins; 194 markers for the Axa 1 cluster, with 27 being hypothetical proteins; 131 markers for the Pro Rep cluster, with 41 being hypothetical proteins; 292 markers for the Axa Rep cluster, with 43 being hypothetical proteins; and lastly 218 markers for the Meta cluster, with 45 being hypothetical proteins.

Shown in **A**, two of the top five differential markers for the Trans A cluster in the Rep1 sample are repeated in this dataset, being the same hypothetical protein (LmxM.08.0810), which has orthologues of amastin-like proteins in other *Leishmania* species when searched in TriTrypDB (Amos et al., 2021), and the Chaperonin HSP60 mitochondrial precursor also seen in top markers for the Pro Rep and Post-Pro like clusters in Rep2_Axa (Figure 3-29). Also of note, was the same ncRNA seen in all sets of top ten markers in previous samples, expressed highly in Trans A and promastigote-like clusters. Other markers in the top five for Trans A were the ADP/ATP translocase 1 and Cystathione gamma lyase. Four of the top five markers for the Axa 1 cluster are ribosomal genes, repeating trends seen in previous amastigote clusters. The other marker being a Ubiquitin-fusion protein, also seen in previous amastigote datasets. For the Pro 2 cluster three markers are repeated from previous promastigote-like clusters, being Nucleoside transporter 1, 3'-nucleotidase/nuclease, and Beta tubulin. The other two markers being Amino acid transporter aatp11 and Membrane-bound acid phosphatase 2. For the Pro 1 cluster all five of the top markers for this cluster have previously been present in the top markers of other promastigote-like clusters, three of the markers being forms of ATG8, a marker distinct for the leptomonad promastigote. The other two markers also being in late promastigote stages; the META domain containing protein and the same pseudogene present in the Lep/Meta markers in Rep2_Pro. Axa 1 markers contain three typical amastigote markers, also used in this study to identify

amastigote life cycle stages, being the Tuzin-like protein, Cathepsin L-like protease, and a Amastin-like surface protein-like protein. The other two markers being hypothetical proteins, one of which has orthologues of amastin in other *Leishmania* species while the other is listed as an unspecified product.

Clusters identified in the integrated dataset are consistent with cluster identity discussed in previous sections. As expected, individual samples were enriched for their respective life cycle stages when integrated. Interestingly, Trans A cluster cells have originated from both amastigote only and promastigote only samples, further indicating a potential overlap in forms placed between differentiation of promastigote to amastigote forms.

Seven of the top ten markers for the Pro cluster are canonical promastigote markers as represented in the literature, with several being used previously for life cycle cluster identification previously in Figure 3-8 and Figure 3-9. Interestingly, two hypothetical proteins and a ncRNA are seen in the top ten markers for promastigotes. Two of the top Trans A markers are forms of the universal minicircle sequence binding protein. Also seen are a translation elongation factor 1-beta, ADP/Translocase 1 and an ABC transporter, potentially all associated with developmental progression from one life cycle stage to another. Three of the top ten markers for Trans A are markers associated with the amastigote stage; two amastin-like proteins and the cathepsin L-like protease, further leading to the observation of a possible transitional cluster between the two life cycle stages not previously identified. For the Axa Rep cluster four genes for histones and four ribosomal protein genes make the top ten markers, the other two markers being profilin, which is associated with cell development and cytokinesis (Krishnan et al., 2009) and tuzin-like protein previously selected as an amastigote marker, again suggesting an amastigote cluster containing cells in the later stages of the cell cycle. And lastly, the Dot plot, shown in **D** shows a general trend from top left to bottom right of promastigote to amastigote markers as expected. Of the top five markers for the Pro Rep cluster three have roles in the cell cycle, being DNA ligase k alpha, Cyclin 6 and DOT1A (U. Chandra et al., 2017; Downey et al., 2005; Gassen et al., 2012). The other two markers are both hypothetical proteins, one of which excitingly having orthologs in *T. brucei*, *T. congolense*, and *T. cruzi* for RNA-

binding proteins and RNA recognition motif domains. Four of the top five markers for the Axa cluster were found to be Histones with the other marker being the DOT1B gene, also associated with cell cycle progression in *T. brucei* and *T. cruzi* (Gassen, et al., 2012; Santana Nunes, et al., 2020). And lastly, four of the top five markers for the Meta cluster being genes associated with metacyclic promastigotes, two of which used as metacyclic markers in this study. These four markers being PSA, Surface antigen protein, GP63-2, and Surface antigen protein 2 (Devault & Bañuls, 2008; Yao et al., 2003; Hallé et al., 2009). The last Meta marker being a putative Lipase.

To further investigate gene ontology for the markers associated with the Trans A and Meta clusters, markers were entered into TriTrypDB (Amos et al., 2022) where gene ontology (GO) enrichment was investigated for biological processes via searching the GO database and the resulting GO terms plotted via the Reduce and Visualise Gene Ontology (REVIGO) application via scatterplot (Supek et al., 2011). GO terms of interest are labelled in the scatterplots produced via REVIGO below in Figure 3-37.

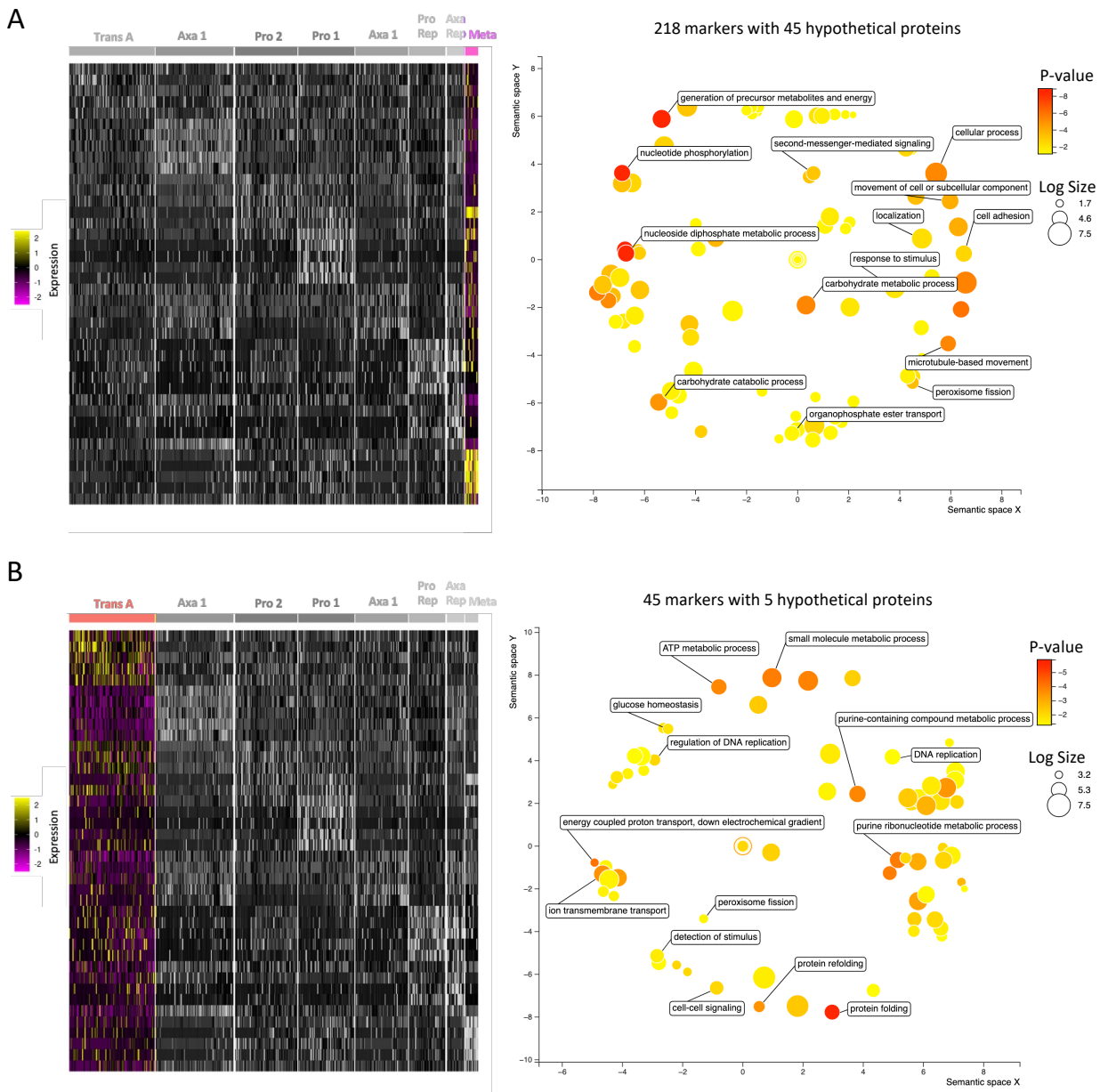


Figure 3-37 GO Term analysis by REVIGO of metacyclic and transitional A cluster. Non-linear projection of multidimensional scaling by REVIGO for gene ontology (GO) enrichment terms to reduce the dimensionality of the matrix of GO terms pairwise semantic similarities. Semantically similar GO terms clustered close together in semantic space. **A**) 218 markers for the metacyclic cluster are analysed for GO terms with TryTrypDB and ontologies for biological processes plotted by REVIGO in a scatter plot coloured by p-value with log size relative to the size of dots. GO terms of interest highlighted. **B**) 45 markers for the transitional A cluster analysed and depicted as in **A**). REVIGO designed and summarised by Supek et al., (2011).

In these scatterplots p-value of the GO terms is indicated by colour, where red indicates a lower p-value (legend in upper right-hand corner) and size of the dots indicates the frequency of the GO term in the underlying Gene Ontology Annotation database, where dots of more general terms are larger in size. The top three GO terms for the Meta cluster were “generation of precursor metabolites and energy” (GO:0006091), ”nucleoside diphosphate metabolic process” (GO:0009132), and “nucleoside diphosphate phosphorylation”

(GO:0006165), with 143 significant GO terms (p-value <0.05) found in this data. The top three GO terms for the Trans A cluster were “protein folding” (GO:0006457), “ATP synthesis coupled proton transport” (GO:0015986), and “energy coupled proton transport, down electrochemical gradient” (GO:0015985), with 127 significant GO terms found in this data. To further compare these data, Venn diagrams of each of the marker sets were compared between Meta and Trans A clusters, and then again comparing these clusters to individual promastigote and amastigote like clusters, shown below in Figure 3-38. This was performed to explore whether the Trans A cluster allows transitions between promastigote and amastigote stages while circumventing the metacyclic stage.

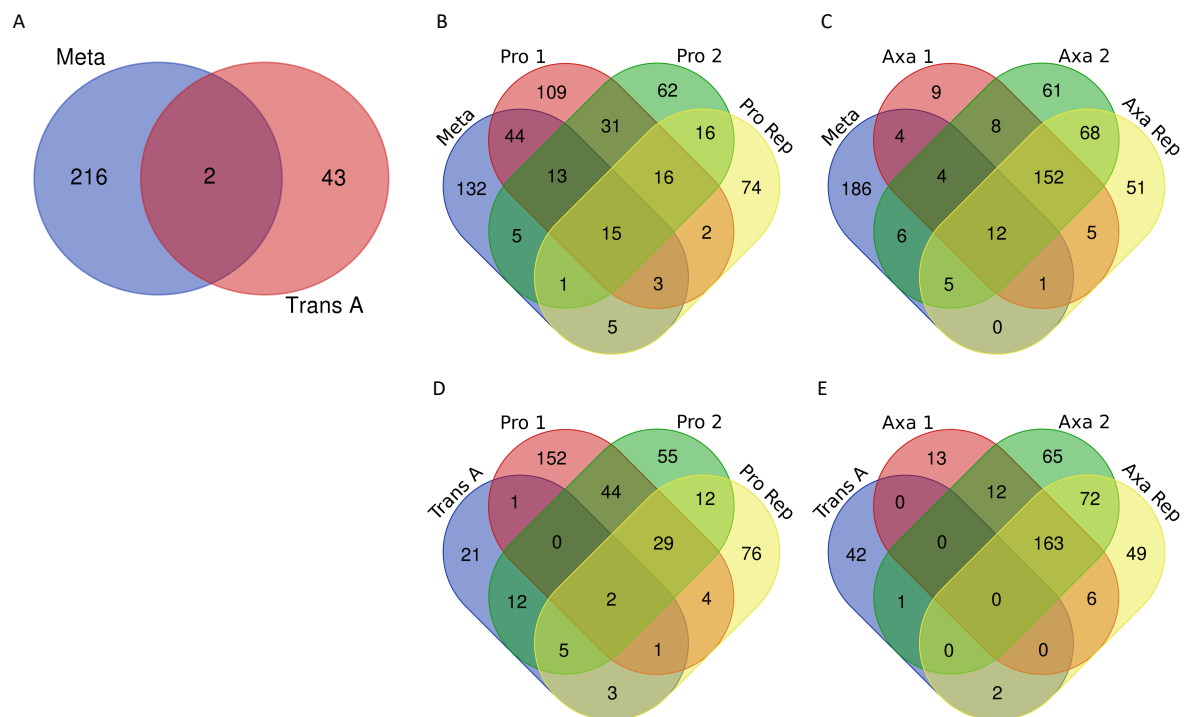


Figure 3-38 Venn diagrams of overlapping markers for life cycle stage clusters between transitional A, promastigote, and axenic amastigote clusters.

A) Venn diagram for metacyclic promastigote and transitional A cluster demonstrating little overlap between these distinct clusters. **B)** Venn diagram for metacyclic promastigote and other promastigote clusters. **C)** Venn diagram for metacyclic promastigote and axenic amastigote clusters. **D)** Venn diagram for transitional A cluster and promastigote clusters. **E)** Venn diagram for transitional A cluster and axenic amastigote clusters.

In the above Venn diagram (Figure 3-38, A) the comparison of Meta and Trans A cluster markers reveals little overlap, with only two markers shared between the clusters, being the cytochrome oxidase subunit VII and phosphate-Repressible Phosphate Permease-like protein. Comparing Meta markers with promastigote cluster markers in B reveals most overlap with the Pro 1 cluster, sharing 44 markers. Overall, the Meta cluster shares 86 markers with promastigote

clusters, indicating a partially shared expression profile, as expected. In comparison, markers for the Meta cluster shared with amastigote clusters in **C** total 32 markers, distinctly less, but with some overlap of markers present in preparation of differentiation in amastigote forms. When comparing Trans A markers to promastigote clusters in **D** we find more than half of the markers are shared between Trans A and promastigote markers, although interestingly no markers are shared between all of three of the Trans A, Pro 1 and Pro 2 markers. In **E**, comparing Trans A markers with amastigote cluster markers, interestingly, only 3 markers are shared with Axa 2 and Axa markers. The single shared marker with Axa 2 cluster being chaperonin HSP60, mitochondrial precursor and the two shared with the Axa cluster being H1 histone-like protein and ribonucleoside-diphosphate reductase large chain (putative). Taken together, this analysis both showed the similarity between the meta cluster and promastigote clusters, while being more distinct from amastigote markers. Additionally, the Trans A cluster displayed little overlap with any other cluster, sharing only 2 markers with the Meta cluster, 24 markers with the promastigote clusters, and only 3 with amastigote clusters: further distinguishing the Trans A cluster as being distinct.

To further compare these datasets Venn diagrams were drawn for promastigote and amastigote markers both with this scRNA-seq data and also to compare to bulk RNA-Seq data generated in promastigote and axenic amastigotes in *L. mexicana* by Fiebig et al., (2015), the results of which are shown below in Figure 3-39.

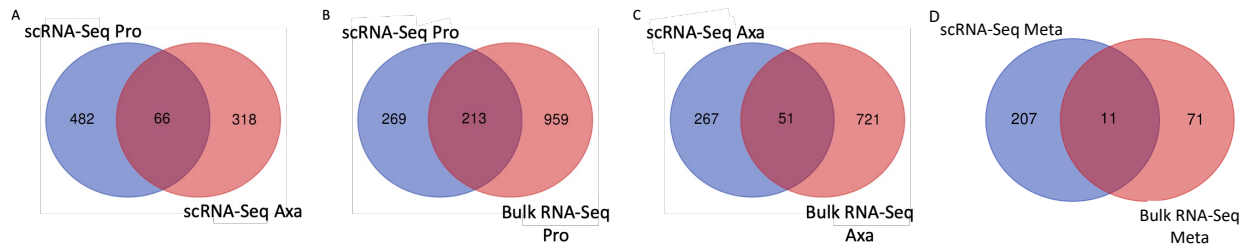


Figure 3-39 Venn diagrams of overlapping markers for life cycle stage clusters between combined scRNA-seq promastigote and axenic amastigote clusters, and bulk RNA-Seq promastigote and axenic amastigote clusters.

A) Venn diagram for scRNA-seq combined promastigote clusters (Pro 1, Pro 2, Pro Rep, Meta and Trans A) against combined axenic amastigote clusters (Axa 1, Axa 2 and Axa Rep) demonstrating moderate overlap between these distinct groups. **B)** Venn diagram for combined scRNA-seq promastigote cluster using the 482 distinct Pro markers from **A)** against Pro markers generated from bulk RNA-Seq data by Fiebig *et al.*, (2015) showing roughly half of scRNA-seq markers are present in and Bulk RNA-Seq markers, with 269 new markers found by scRNA-seq. **C)** Venn diagram for combined scRNA-seq promastigote cluster using the 318 distinct Axa markers from **A)** against Axa markers generated from bulk RNA-Seq data by Fiebig *et al.*, (2015) showing roughly one sixth of scRNA-seq markers are present in and Bulk RNA-Seq markers, with 267 new markers found by scRNA-seq. **D)** Comparison of scRNA-seq markers for Meta cluster compared to orthologues derived from bulk RNA-Sequencing of metacyclic in *Leishmania infantum*, by Coutinho-Abreu *et al.* (2020), showing only 11 marker overlapping with 207 new metacyclic markers identified.

Above in Figure 3-39 **A** is a Venn diagram comparing all promastigote markers from Pro1, Pro 2, and Pro clusters comparing to all markers in Axa 1, Axa 2, and Axa markers. When comparing these markers, 66 markers are shared between the two groups of life cycle stages. Of the 66 shared markers, the large majority are metabolic and cell cycling markers, the latter being shared between Pro Rep and Axa Rep clusters. 482 distinct promastigote markers and 318 distinct amastigote markers were found. These 482 distinct promastigote markers were then compared with the bulk RNA-Seq data generated by Fiebig *et al.*, (2015) in **B**, revealing 269 unique promastigote markers not previously seen in bulk analysis of the promastigote stage. In **C** the same comparison is made between the 318 scRNA-seq axenic amastigote markers with bulk RNA-Seq markers revealing a further 267 markers uniquely discovered by these scRNA-seq data. For consideration of these unique markers, however, are the slight variations in culture conditions for the parasites between the scRNA-seq and bulk RNA-Seq datasets. Additionally, when comparing markers for the metacyclic cluster by scRNA-seq to markers identified by bulk RNA-sequencing in *L. infantum* by Coutinho-Abreu *et al.* (2020) in **D**, 11 markers were found to overlap, with 207 novel markers for the metacyclic cluster identified by scRNA-seq.

3.3.5.4 Life cycle labelling of transitional A cluster using promastigote stage markers

As cluster Trans A showed similar transcriptomic signatures to promastigote clusters, next the promastigote form labelling approach was repeated as performed in section 3.3.3.3. Where life cycle phase specific markers from bulk RNA-sequencing by Coutinho-Abreu et al. (2020), as explained in detail in Section 3.3.3.3. Proportions of cells labelled with promastigote life cycle phases are shown below in Figure 3-40, **A**.

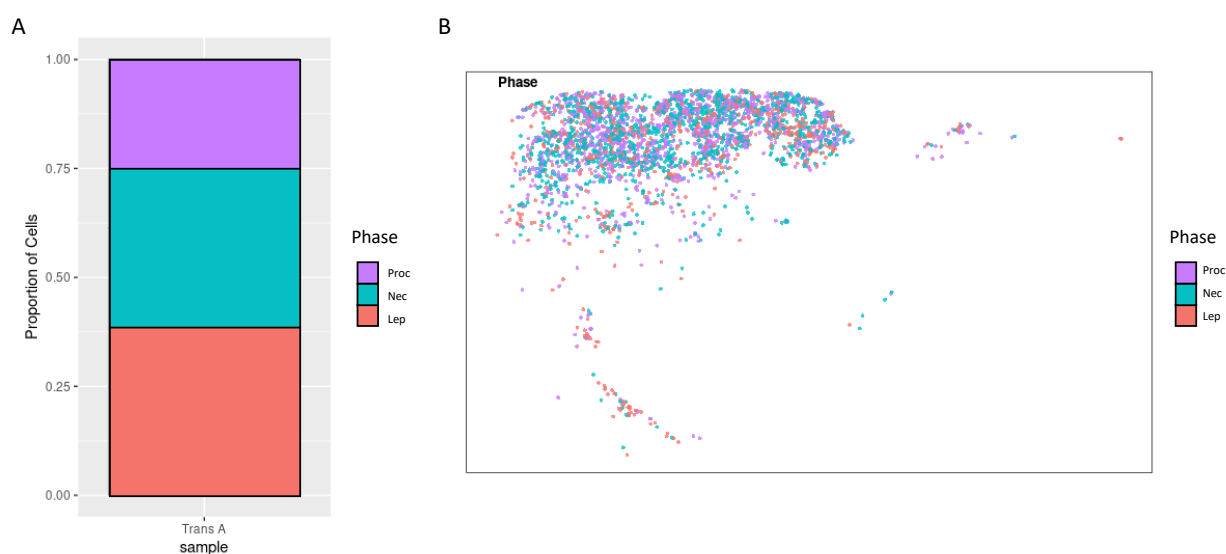


Figure 3-40 Promastigote labels for transitional A cluster reveals a mixture of promastigote stages with no distinct promastigote life cycle stage across the cluster.

A) Proportion of promastigote forms found in transitional A cluster following labelling of cells with orthologous promastigote markers. Procyclic promastigotes in purple, nectomonad promastigotes in turquoise and leptomonad promastigotes in red, as in Figure 3-18. **B)** UMAP of labelled transitional A cluster coloured as in **A**).

Shown above in Figure 3-40 **A** is a proportional plot of promastigote life cycle stages labelled as in Figure 3-18. Labelling of these promastigote life cycle stages in **A**, using the same colours as in Figure 3-18, shows an approximately equal mix of the first three promastigote life cycle stages, with procyclic promastigotes being least represented of the three in the Trans A cluster. Of note is the absence of any cells labelled as metacyclic promastigotes. The labels for each individual cell being allocated and imaged in a UMAP in **B**. This further suggests this Trans A cluster is very distinct from the Meta cluster, despite being placed between promastigote and amastigote clusters, lending further evidence to a potentially new life cycle stage for *Leishmania* differentiating between promastigotes and amastigotes, *in vitro*. To further investigate the overlap of life cycle and cell cycle stages across the integrated dataset, the same labelling

method above is again applied using orthologous *L. mexicana* cell cycle markers (as in Figure 3-11) for the integrated UMAP was undertaken, as shown in Figure 3-41 below. Additionally, inferred trajectory and pseudotime analysis was run for both combinations of Pro Rep and Axa Rep cluster together and individually, with promastigote life cycle stage labelling applied to the Pro Rep cluster.

3.3.5.5 Cell Cycle Labelling of Integrated Samples with Trajectory Analysis

In order to elucidate possible cell cycle based regulation of life cycle development in *L. mexicana* further using this scRNA-seq data, a method of cell cycle stage labelling developed by Briggs *et al.* (2021b) was recapitulated and applied, as described in Section 3.3.2.4.

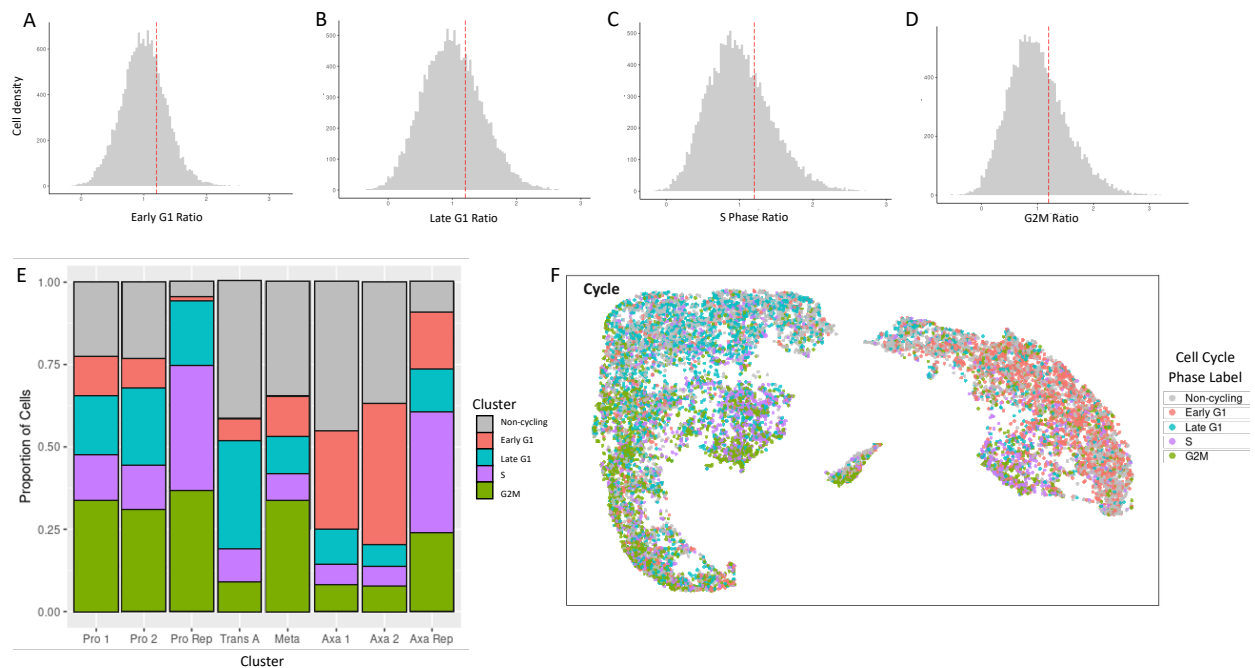


Figure 3-41 Labelling of single cell transcriptomes by orthologous cell cycle stage specific markers in integrated samples.

A) – D) ratio of orthologous cell cycle progression markers from *T. brucei* with no cell cycle stage specific marker determined by those not over-expressing any cell cycle stage specific markers below a 10% ratio cut-off, indicated by the red dashed line. **E)** Proportion of cells labelled by cell cycle phase orthologous markers across clusters, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. Clusters ordered from left to right my life cycle progression from promastigote to amastigote. **F)** UMAP of cells in integrated samples assigned and coloured by cell cycle stage, as in **E)**.

Assigning cell cycle stages to cells over-expressing these markers and labelling each phase by colour, as shown in Figure 3-41 F above reveals distinct groups of cells enriched for different cell cycle phases. As shown in E two clusters stand out as being heavily labelled with cell cycle phase markers, these being the Pro Rep and Axa Rep clusters, where only 4.2% and 9.1% of cells were labelled as

non-cycling (coloured in grey), respectively. The largest group of cell cycle markers for each of these two clusters being labelled in the S phase (coloured in purple). The same trend seen in previous individual replicate analysis for amastigote clusters is also seen, where an enrichment for the Early G1 labelling (coloured in red) is found, with 42.4% of the Axa 2 cluster being labelled as Early G1. Promastigote clusters are interestingly enriched approximately 35% in G2M labels. While the Trans A cluster once again is distinctly labelled as being in Late G1 (coloured in turquoise, 32.8%) having the largest enrichment for this phase across the integrated dataset.

Having collectively found evidence of proliferative, S phase rich, populations identified in the Pro Rep (37.6% S phase) and Axa Rep (36.3% S phase) clusters via this analysis and previous GO term enrichment, the Pro Rep and Axa Rep clusters were then subset from the integrated data set for trajectory and pseudotime analysis to infer if these two replicative life cycle stages may be ordered by their cell cycle stages. The analysis of which can be found below in Figure 3-42.

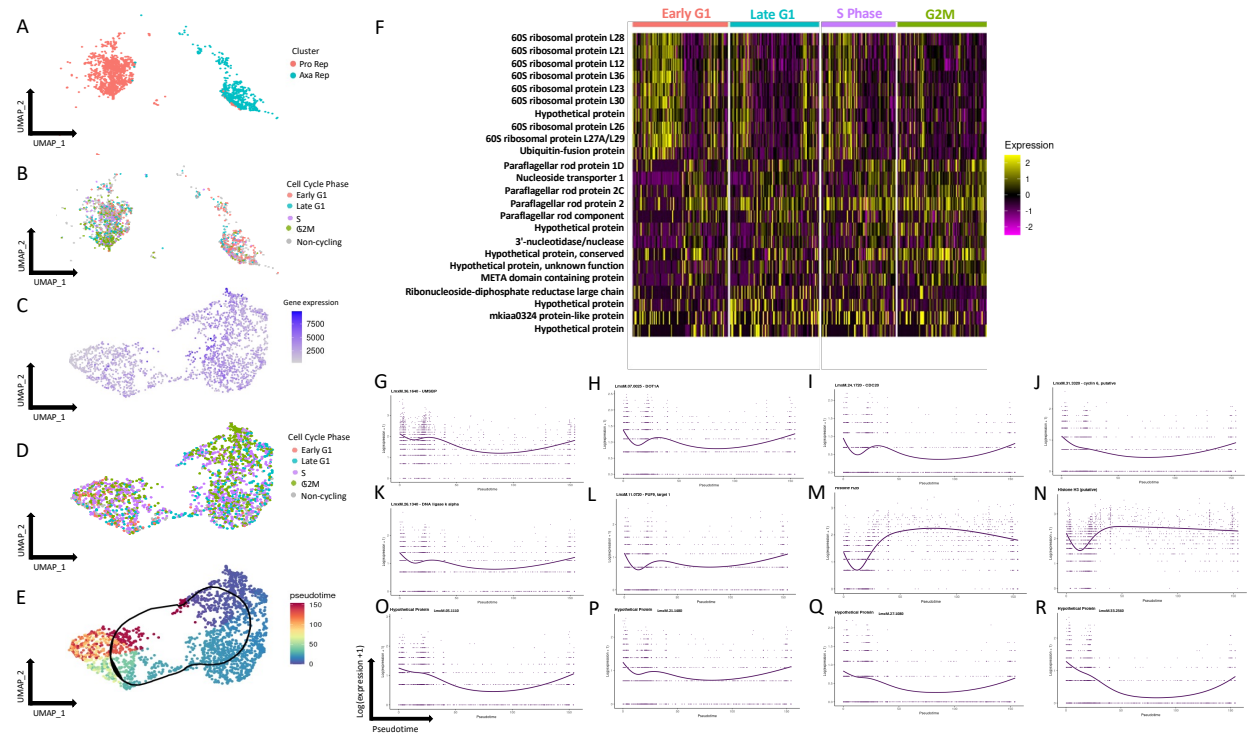


Figure 3-42 Detection of markers differentially expressed in replicating promastigote and axenic amastigote clusters.

A) UMAP of subset of replicating promastigote and axenic amastigote clusters. Promastigote replicating cluster in red and axenic amastigote cluster in turquoise. **B**) UMAP of replicating clusters labelled as in Figure 3-41, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **C**) Replicating cluster in **A**) replotted in a UMAP using only variable genes in replicating subset. Coloured by raw counts of gene expression. **D**) Relabelling of replotted replicating clusters in UMAP using the same cell cycle stage markers and colours as in **B**). **E**) UMAP of replicating promastigote and axenic cluster subset coloured by assigned pseudotime value. The inferred circular trajectory represented by the black circular line. **F**) Heatmap of top ten distinguishing markers, if present, differentially expressed in labelled cell cycle phase populations (log₂ normalised z-score). **G** – **R**) Cell cycle pseudotime (x-axis) against expression levels (y-axis; log₂(expression + 1)) of selected cell cycle markers with smoothed average expression indicated by purple line. Each point represents expression level for one cell's transcriptome. Marker **G**) Universal minicircle sequence binding protein; LmxM.36.1640, **H**) DOT1A; LmxM.07.0025, **I**) CDC20; LmxM.24.1720, **J**) Cyclin 6; LmxM.31.3320, **K**) DNA ligase k alpha; LmxM.26.1340, **L**) PUF9, target 1; LmxM.11.0720, **M**) Histone H2B; LmxM.19.0050, **N**) Histone H3 (putative); LmxM.16.0600, **O**) Hypothetical protein; LmxM.05.1110, **P**) Hypothetical protein; LmxM.21.1480, **Q**) Hypothetical protein; LmxM.27.1080, **R**) Hypothetical protein; LmxM.33.2560.

Shown above in Figure 3-42 **A** are the subset Pro Rep (coloured in red) and Axa Rep (coloured in turquoise) clusters. These clusters were then relabelled with cell cycle stages labelled for the highest scoring phase for each cell in the UMAP in **B**, as previously in Figure 3-11. Cells not expressing any cell cycle phase as a ratio below 1.2 were labelled as being non-cycling, as described in Section 3.3.2.4. The Pro Rep and Axa Rep were then replotted in a UMAP, showing a loop of cells with gene expression counts shown for the replotted cells in **C**, where gene expression is coloured in purple. Some enrichment for each of the cell cycle phases is seen when the replotted cells are relabelled by cell cycle phase in UMAP, shown in **D**. A trajectory is inferred in this replotted UMAP using the

principal curve package (Weingessel, 2015), drawn in a black line, and pseudotime values assigned along the trajectory, as shown in the UMAP in **E**. The circular trajectory begins at the top of the replotted loop of cells (indicated in blue pseudotime colouring) and continues clockwise around finishing the loop where it started (indicated in red). These cell cycle marker genes were then grouped by phase and replotted on a Heatmap in **F**, where relative expression levels (\log_2 normalised z-score) were assigned to each gene and the top ten markers (if present) were plotted with each column representing the expression profile of a cell. For differentially expressed genes assigned to each cell cycle phase, 256 markers for cells in early G1 were found, with 4 for late G1, and 38 for G2M. Strangely, no unique markers were found for the cells labelled with the S Phase, and upon closer investigation of the top 10 distinguishing markers, it became apparent these cells were being ordered by life cycle markers previously seen in life cycle cluster analysis. This was further confirmed when looking at smoothed expression plots along pseudotime for a selection of cell cycle associated markers and hypothetical proteins enriched in replicating clusters, shown in **G - R**, where the $\log(\text{expression} + 1)$ on the y-axis for each selected gene is plotted, and each dot represents the expression for an individual cell, with its smoothed expression (indicated by the purple line) plotted across the pseudotime trajectory on the x-axis, as previously shown in Figure 3-21. Following this analysis, the replicating clusters for promastigotes and amastigotes was investigated separately. The replicating promastigote cluster being investigated further first below in Figure 3-43. Before cell cycle phase labels were applied, life cycle stage markers were first assessed for the Pro Rep subset, using the same method described in Figure 3-18.

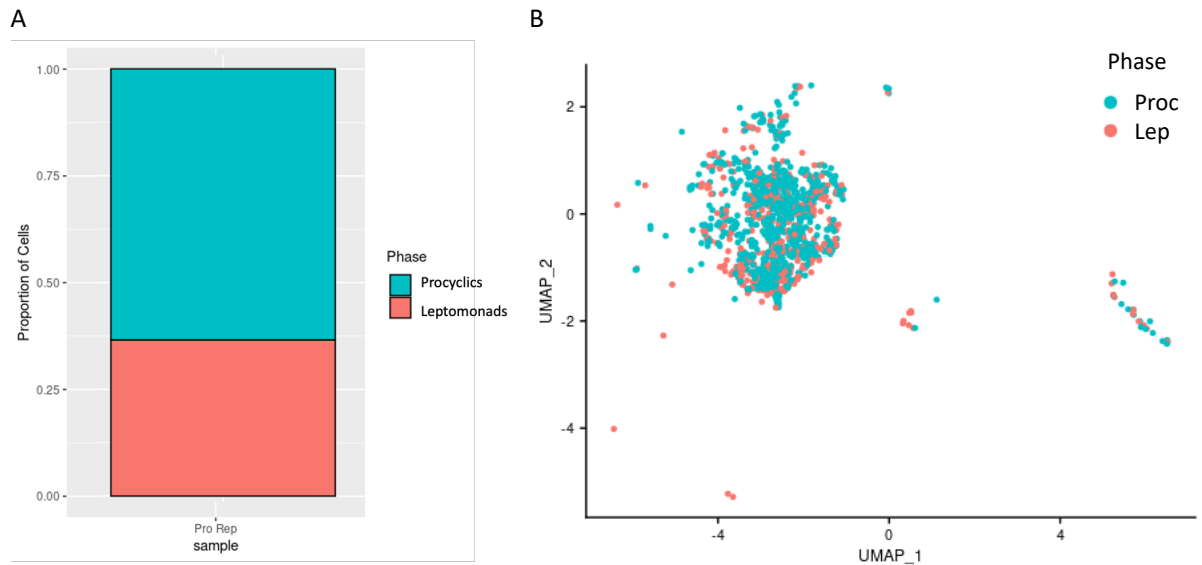


Figure 3-43 Promastigote labels for promastigote replicating cluster reveals only procyclic and leptomonad promastigote stages.

A) Proportion of promastigote stages found in promastigote replicating cluster following labelling of cells with orthologous promastigote markers. Procyclic promastigotes in turquoise, and leptomonad promastigotes in red, as in Figure 3-18. **B)** UMAP of labelled promastigote replication cluster coloured as in **A**).

Following labelling with promastigote life cycle stages, a population of only procyclic and leptomonad promastigotes is found. These two stages being determined as the two replicating life cycle stages in the promastigote.

Following the confirmation of replicating life cycle stages in this cluster subset, cell cycle labelling of the Pro Rep cluster was performed, as described in Figure 3-42.

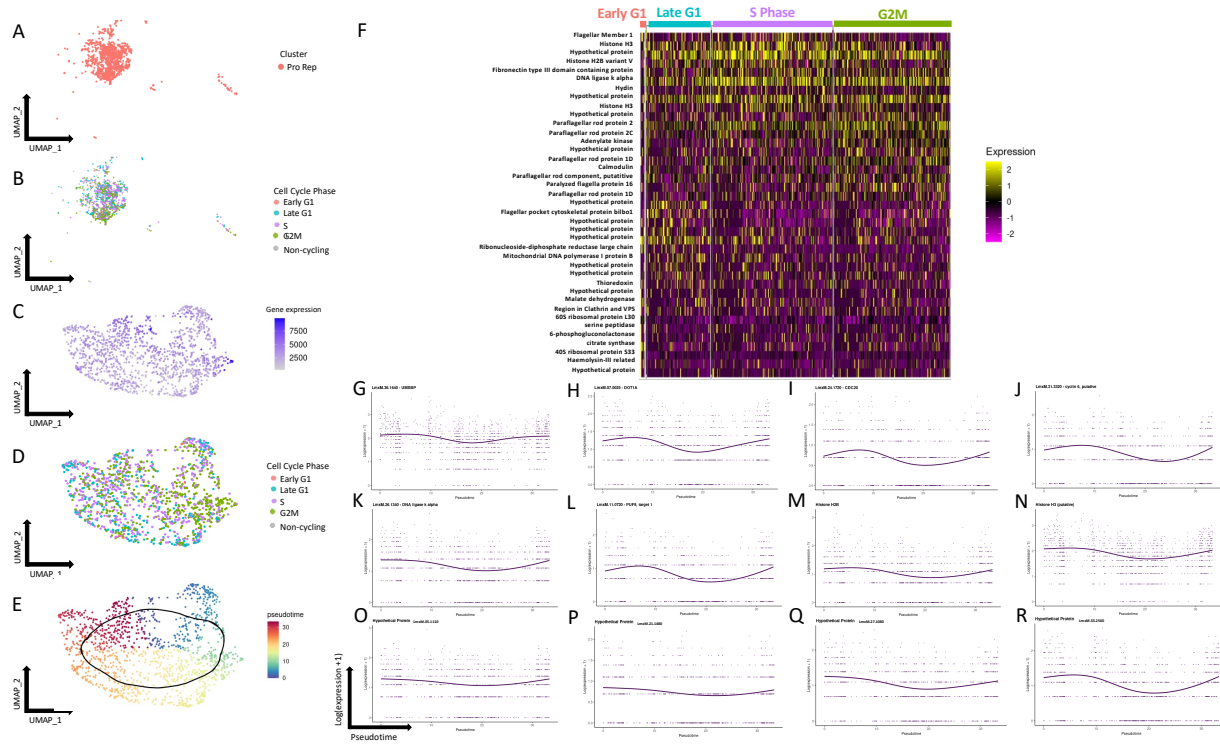


Figure 3-44 Detection of markers differentially expressed in replicating promastigote cluster.

A) UMAP of subset of replicating promastigote cluster. Promastigote replicating cluster in red. **B)** UMAP of replicating promastigote cluster labelled as in Figure 3-41, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **C)** Replicating promastigote cluster in **A)** replotted in a UMAP using only variable genes in replicating promastigote subset. Coloured by raw counts of gene expression. **D)** Relabelling of replotted replicating promastigote cluster in UMAP using the same cell cycle stage markers and colours as in **B)**. **E)** UMAP of replicating promastigote cluster subset coloured by assigned pseudotime value. The inferred circular trajectory represented by the black circular line. **F)** Heatmap of top ten distinguishing markers differentially expressed in labelled cell cycle phase populations (\log_2 normalised z-score). **G**) – **R)** Cell cycle pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of selected cell cycle markers with smoothed average expression indicated by purple line. Each point represents expression level for one cell's transcriptome. Marker **G)** Universal minicircle sequence binding protein; LmxM.36.1640, **H)** DOT1A; LmxM.07.0025, **I)** CDC20; LmxM.24.1720, **J)** Cyclin 6; LmxM.31.3320, **K)** DNA ligase k alpha; LmxM.26.1340, **L)** PUF9, target 1; LmxM.11.0720, **M)** Histone H2B; LmxM.19.0050, **N)** Histone H3 (putative); LmxM.16.0600, **O)** Hypothetical protein; LmxM.05.1110, **P)** Hypothetical protein; LmxM.21.1480, **Q)** Hypothetical protein; LmxM.27.1080, **R)** Hypothetical protein; LmxM.33.2560.

Shown above in Figure 3-44 **A** is the subset Pro Rep (coloured in red) cluster.

This cluster was then relabelled with cell cycle stages labelled for the highest scoring phase for each cell in the UMAP in **B**, as previously in Figure 3-42. Little enrichment for each of the cell cycle phases is seen when the replotted cells are relabelled by cell cycle phase in UMAP, shown in **D** with G2M cells towards the right side of the loop, and proceeding clockwise to late G1 towards the bottom, and progressing to S phase on the left. Of note was the very few cells labelled as being in early G1, as represented in the proportion plot in Figure 3-41, **E**. For differentially expressed genes assigned to each cell cycle phase, 135 markers for cells in early G1 were found, with 18 for late G1, 9 for S Phase, and 28 for G2M.

Examining the markers for each phase via Heatmap revealed 12 hypothetical proteins and 3 histones present, and 9 genes associated with paraflagellar rod formation. Other notable cell cycle markers included DNA ligase k alpha which has previously appeared as a marker in individual replicate analysis, and Mitochondrial DNA polymerase I protein B. Analysis of the same cell cycle markers and hypothetical proteins enriched in replicating clusters were used in smoothed expression plots along pseudotime, as in Figure 3-42, again, shown here in **G - R**, demonstrates greater changes in expression across pseudotime, which inferred a more interpretable trajectory being drawn for the replicating promastigote cluster when analysed individually. This same analysis was then applied to the replicating amastigote cluster, shown below in Figure 3-45.

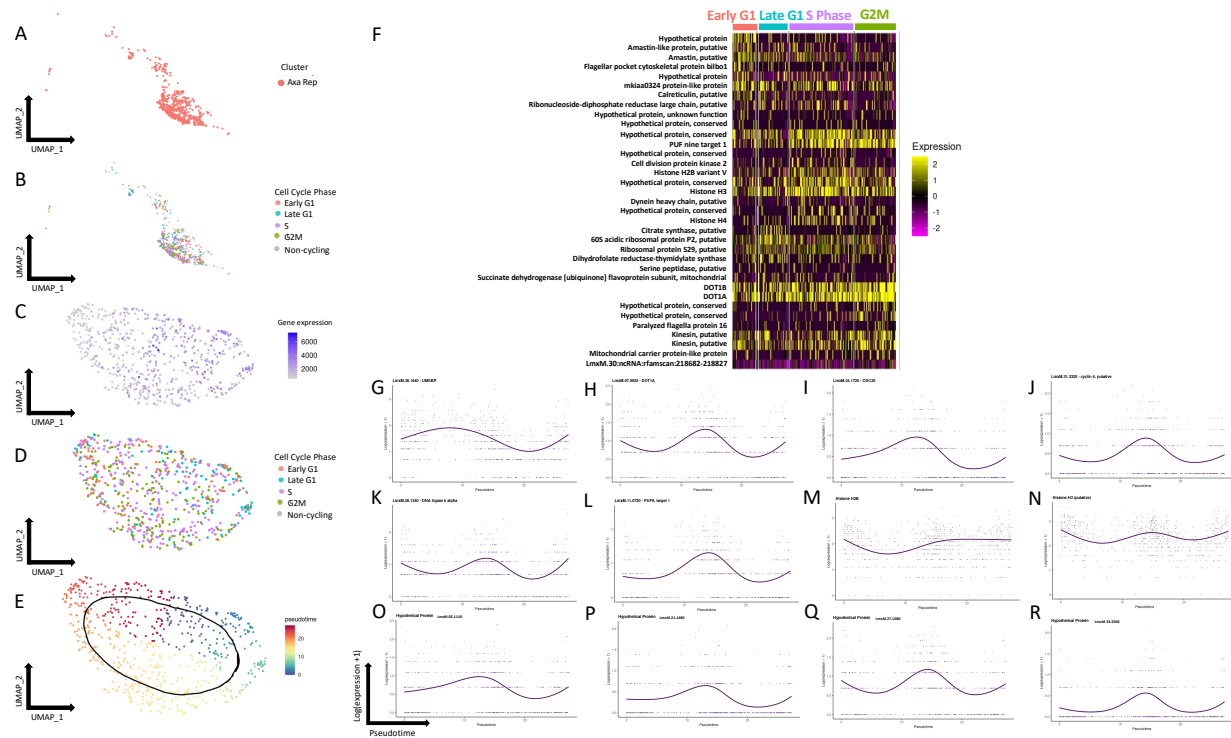


Figure 3-45 Detection of markers differentially expressed in replicating axenic amastigote cluster.

A) UMAP of subset of replicating axenic amastigote cluster. Axenic amastigote replicating cluster in red. **B)** UMAP of replicating axenic amastigote cluster labelled as in Figure 3-41, with unlabelled cells coloured grey, Early G1 in red, Late G1 in turquoise, S phase in purple and G2M in green. **C)** Replicating axenic amastigote cluster in **A)** replotted in a UMAP using only variable genes in replicating axenic amastigote subset. Coloured by raw counts of gene expression. **D)** Relabelling of replotted replicating axenic amastigote cluster in UMAP using the same cell cycle stage markers and colours as in **B)**. **E)** UMAP of replicating axenic amastigote cluster subset coloured by assigned pseudotime value. The inferred circular trajectory represented by the black circular line. **F)** Heatmap of top ten distinguishing markers differentially expressed in labelled cell cycle phase populations (log2 normalised z-score). **G) – R)** Cell cycle pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of selected cell cycle markers with smoothed average expression indicated by purple line. Each point represents expression level for one cell's transcriptome. Marker **G)** Universal minicircle sequence binding protein; LmxM.36.1640, **H)** DOT1A; LmxM.07.0025, **I)** CDC20; LmxM.24.1720, **J)** Cyclin 6; LmxM.31.3320, **K)** DNA ligase k alpha; LmxM.26.1340, **L)** PUF9, target 1; LmxM.11.0720, **M)** Histone H2B; LmxM.19.0050, **N)** Histone H3 (putative); LmxM.16.0600, **O)** Hypothetical protein; LmxM.05.1110, **P)** Hypothetical protein; LmxM.21.1480, **Q)** Hypothetical protein; LmxM.27.1080, **R)** Hypothetical protein; LmxM.33.2560.

Shown above in Figure 3-45 **A)** is the subset Axa Rep (coloured in red) cluster.

This cluster was then relabelled with cell cycle stages labelled for the highest scoring phase for each cell in the UMAP in **B)**, as previously in Figure 3-42. Some enrichment for each of the cell cycle phases was seen when the replotted cells are relabelled by cell cycle phase in UMAP, but perhaps less than previously demonstrated, as shown in **D)** with early and late G1 labelled cell collecting at the top of the loop, followed by S phase cells towards the bottom of the loop, ending in G2M towards the leftmost edge. For differentially expressed genes assigned to each cell cycle phase, 6 markers for cells in early G1 were found, with 27 for late G1, 13 for S Phase, and 18 for G2M. Examining the markers for

each phase via Heatmap revealed 9 hypothetical proteins, and 3 histones also present. Other notable cell cycle markers include PUF nine target 1, Cell division protein kinase 2, DOT1A, and DOT1B. Analysis of the same cell cycle markers and hypothetical proteins enriched in proliferating clusters were used in smoothed expression plots along pseudotime in Figure 3-42, again, shown here in **G - R**, which demonstrated greater changes in expression across pseudotime than previous analysis, which inferred a more interpretable trajectory being drawn for the proliferative amastigote cluster when analysed individually.

3.3.5.6 Pseudotime Analysis of Integrated Samples Reveals Distinct Gene Expression Patterns Through Life Cycle Progression

Following the above pseudotime analysis of replicating cluster subset, the integrated samples were then used for trajectory visualisation and analysis along promastigote to axenic amastigote differentiation using a PHATE map, as explained previously in Section 3.3.3.5. Whereby cells were again replotted onto a PHATE map to visualise the complete branching trajectory structure in high-dimensional data, plotting onto two-dimensional space, while retaining global and local structure associated with the developmental progression of gene expression found within the object. This PHATE map can then be used for pseudotime ordering of trajectories to further group matching expression profiles along the newly inferred trajectory (Moon et al., 2017). A PHATE map for the integrated replicates was produced and shown below in Figure 3-46.

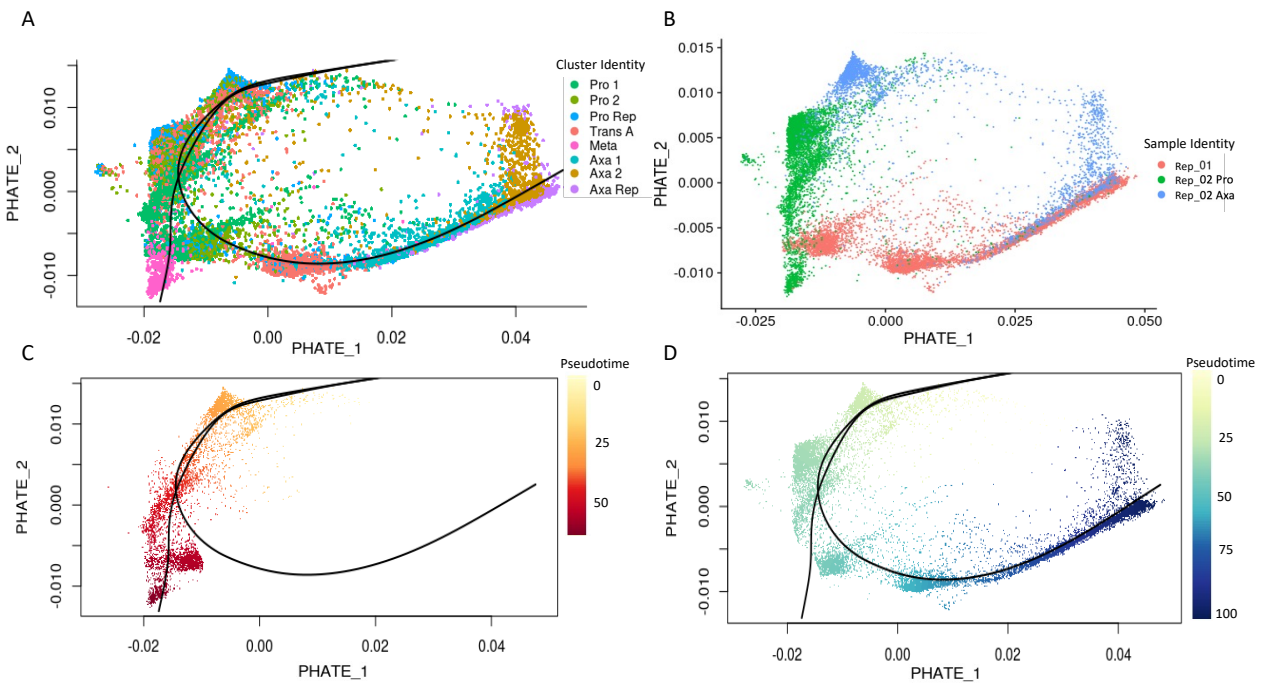


Figure 3-46 Pseudotime analysis demonstrates two distinct trajectories from Pro 1 cluster ending in Meta and Axa Rep clusters.

PHATE plots of single cell *L. mexicana* transcriptomes. **A)** Coloured by cluster identity as in Figure 3-33. Trajectories represented by black lines, both branched lineages starting in Pro 1 cluster (lime green). trajectory 1 ending in Meta cluster (pink) and trajectory 2 ending in Axa Rep cluster (purple). **B)** PHATE plot as in **A)** coloured by biological replicate samples; Rep_1 in red, Rep2_Pro in green and Rep2_Axa in blue. **C)** PHATE plot of cells on trajectory 1 coloured by pseudotime from yellow to red, starting in Pro 1 cluster and ending in Meta cluster. **D)** PHATE plot of cells on trajectory 2 coloured by pseudotime from yellow to blue, starting in Pro 1 cluster and ending in Axa Rep cluster.

As previously described for Figure 3-22, shown in **A** is a PHATE map replotted using the clusters defined and coloured in previous analysis and a trajectory inferred, as indicated by the black lines. Interestingly, here two trajectories are found, both directed to start in the Pro 1 cluster (coloured in lime green). One, shorter, trajectory ending in the Meta cluster (coloured in pink, in the bottom left), and one trajectory crossing through the Trans A cluster (coloured in red, at the bottom of the curve) finally ending in the Axa Rep cluster (coloured in purple, placed rightmost). Shown in **B** is the sample identities for each of the replicates, Rep1 coloured in red, Rep2_Pro coloured in green, and Rep2_Axa coloured in blue. Interestingly, a portion of the Rep2_Axa replicate has been grouped with the Pro Rep cluster, likely originating from the so-called Pro Rep cluster shown in Figure 3-29 **A** and shown to be overlapping in the Pro Rep cluster in Figure 3-33, **C**. Plotted on PHATE maps in Figure 3-46 **C** and **D** are the cells assigned to the Pro - Meta and Pro - Axa trajectories, respectively. Where the Pro - Meta trajectory is coloured in yellow to red (ending in the Meta cluster) and the Pro - Axa trajectory coloured in yellow to blue (ending in the Axa Rep

cluster). Additionally, a branching point between the two trajectories was found, separating along promastigote labelled clusters.

In total, the expression of 1957 genes were found to be associated with both these two trajectories, with 425 associated for the Pro-Meta trajectory, where all genes were differentially expressed as a function of pseudotime (p -value < 0.05, fold change > 2). These genes were then assigned into five modules (A-E) based on similarity of expression profiles and ordered manually, as previously explained in Section 3.3.3.5, and plotted below in Figure 3-47 A. Those not grouped together were assigned as being unsorted (coloured in grey). These modules were then plotted together on a heatmap, shown below in Figure 3-47 B, with pseudotimes running from left to right, indicated on top. Where the pseudotime across the Pro - Meta trajectory is coloured from yellow to red, and the Pro - Axa trajectory coloured from yellow to blue, as above in Figure 3-46.

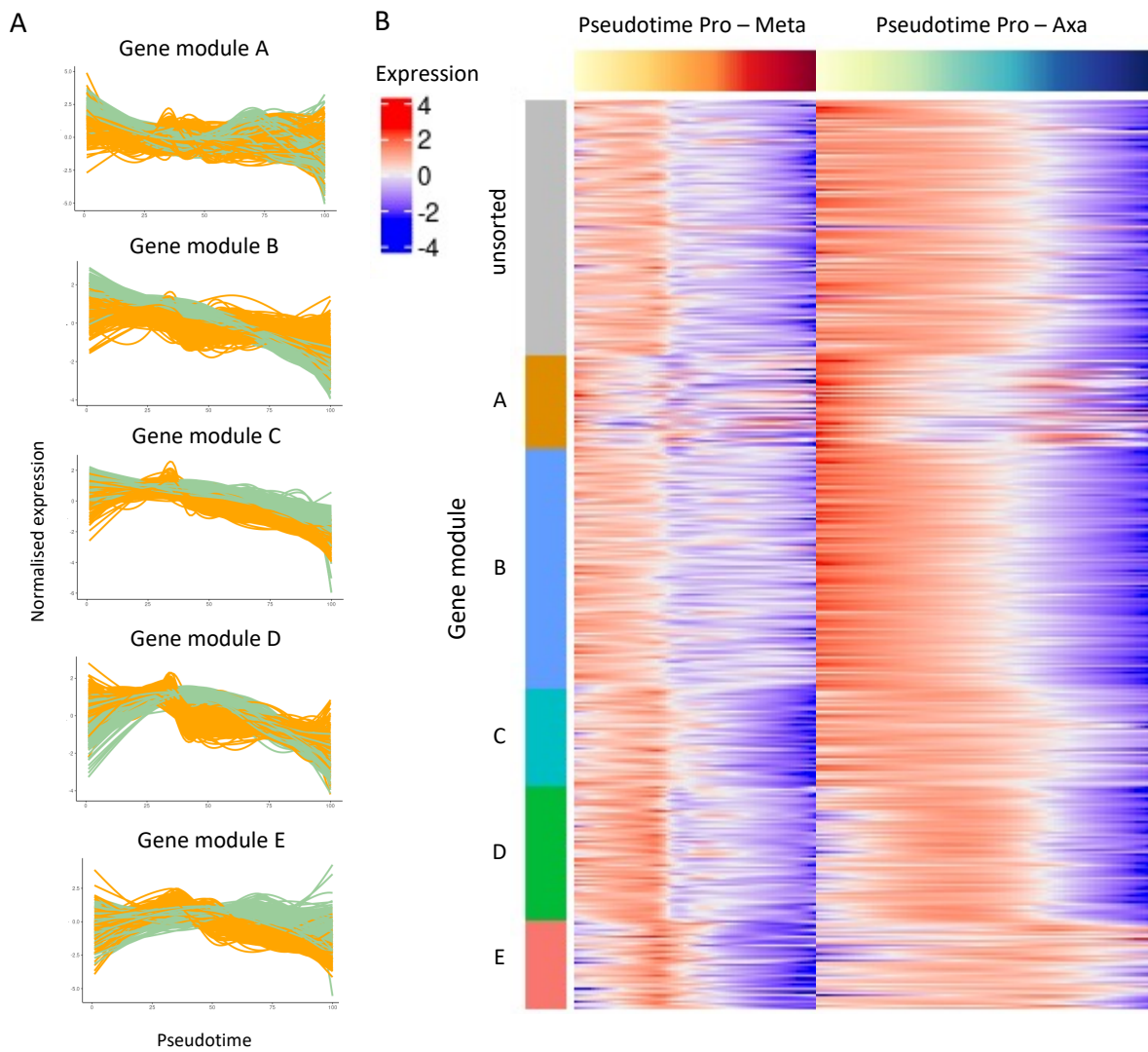


Figure 3-47 Ordering and grouping of genes differentially expressed through *L. mexicana* life cycle progression.

A) 5 assigned gene module groups for pseudotime against normalised gene expression (\log_2 normalised z-score) and grouped by co-expression, where pseudotime lineage from Pro to Meta modules are coloured in orange and Pro to Axa lineage is coloured in sage green. **B)** Heatmap plot of 5 gene modules shown in **A)** showing relative expression levels (\log_2 normalised z-score) of 1947 genes with differential expression significantly related with the pseudotime trajectory (as determined by associationTest; p -value > 0.05 , FC > 2). Left lineage from yellow to red on top of heatmap for the Pro to Meta trajectory and right lineage from yellow to blue on top of heatmap for the Pro to Axa trajectory. Unsorted genes assigned to a separate gene module, coloured in grey.

Of note, was the trend across pseudotime for the five modules, with transient expression of the Pro - Meta and Pro- Axa genes seen in modules A and D with opposite expression patterns. In module A, under expression of genes in the Pro - Axa trajectory occur when trajectories branch, with genes being overexpressed following commitment to amastigote forms after passing through the Trans A cluster. The expression of genes seen in the Pro - Axa trajectory in module D potentially being timed with expression of genes associated with the Trans A cluster. Module E also saw a transient upregulation of both Pro - Meta and Pro - Axa trajectories, although the latter occurred later in the trajectory. This

transient expression seen in E in the Pro - Meta trajectory could again indicate genes that may be involved with activation of genes associated with metacyclic commitment and formation, as shown previously in Figure 3-22. The timing for over expression of genes in the Pro - Axa matching the trajectory moving through the Trans A cluster.

Following trajectory inference with pseudotime, markers previously selected for cluster identification were selected for smoothed expression plots and logged gene count plots, as in Figure 3-22, shown below in Figure 3-48.

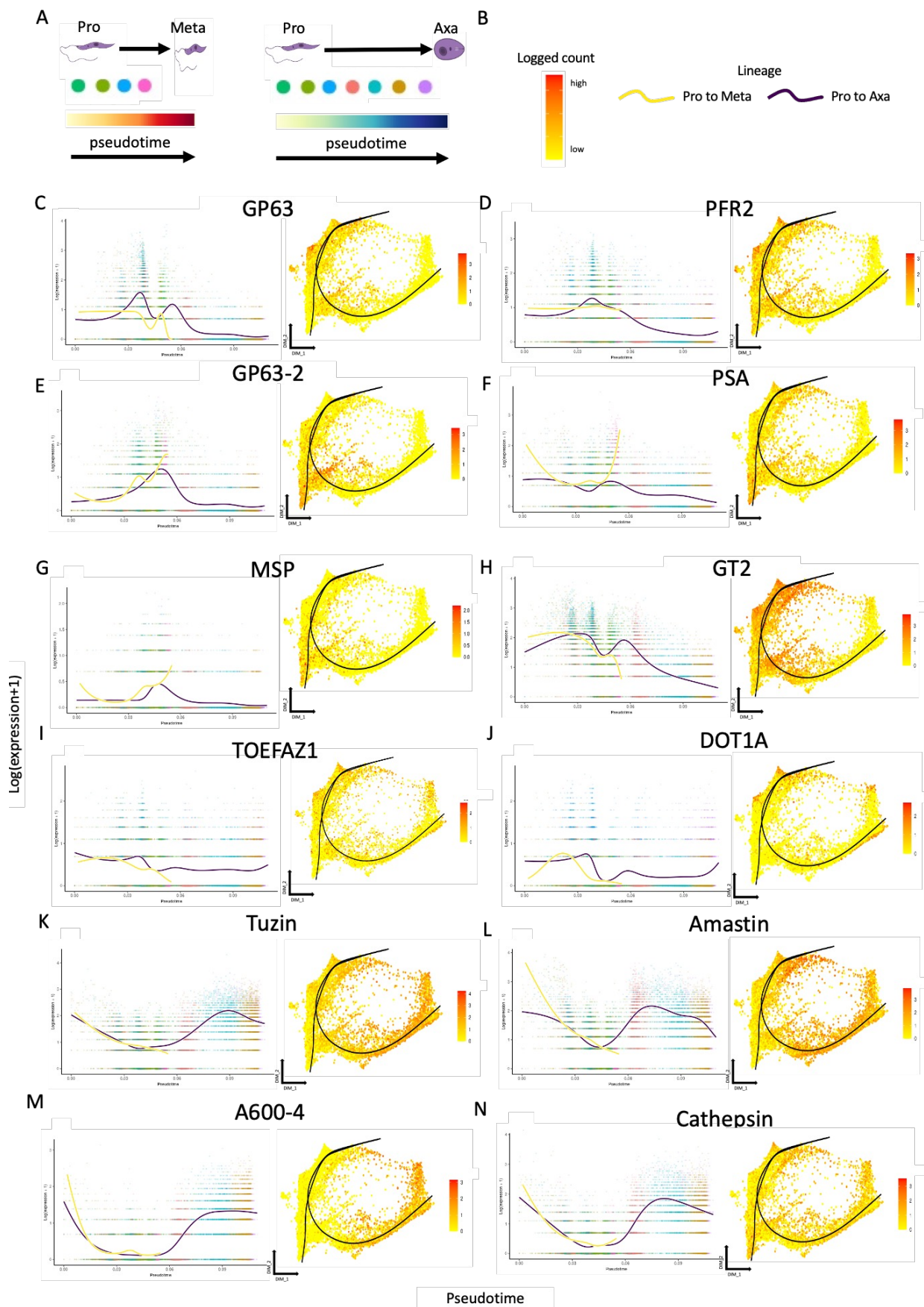


Figure 3-48 Expression of life cycle marker genes across pseudotime trajectories.

A) Schematic of pseudotime trajectories demonstrating progression through life cycle stage clusters coloured as in Figure 3-33. **B)** Expression colours for PHATE map markers, and trajectory colours for expression plot as shown in **C** – **N**. **C** – **N**) Trajectory lineage pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of selected life cycle markers with

smoothed average expression of pseudotime 1 lineage indicated by yellow line and smoothed average expression of pseudotime 2 lineage indicated by purple line. Each point represents expression level for one cell's transcriptome coloured by cluster identity as in Figure 3-33. Included with each marker are logged marker transcript counts assigned to PHATE plots, coloured by over expression in red against cells not expressing in yellow, as in **B**). Markers **C** - **G** are used as markers for promastigote stages (GP63; LmxM.10.0460, PFR2; LmxM.16.1430, GP63-2; LmxM.10.0470, PSA; LmxM.12.0980, MSP; LmxM.28.0570), with markers **E** - **G** indicative of the metacyclic promastigote stages. **H**) GT2; LmxM.36.6290, a glucose transporter known to be upregulated in promastigote stages. **I** - **J**) (TOEFAZ1; LmxM.31.2610, DOT1A; LmxM.07.0025) included as markers for populations undergoing cytokinesis and cell cycle progression, respectively. Markers **K** - **N**) (Tuzin; LmxM.33.1970, Amastin; LmxM.30.0450, A600-4; LmxM.33.3645 and Cathepsin; LmxM.08.1070) are used as markers for amastigote life cycle stages.

Shown in **A** is a schematic of the two trajectories with their progress through clusters indicated by dots coloured by cluster. In **B** is a key for lineage smoothed expression plots and logged gene counts for the life cycle cluster marker plots shown below.

Genes plotted in **C** - **N** show the smoothed expression plots for the life cycle markers used in previously cluster analysis, where the $\log(\text{expression} + 1)$ on the y-axis for each selected gene is plotted, and each dot represents the expression for an individual cell and is coloured by cluster. The smoothed expression plotted across the pseudotime trajectory on the x-axis, indicated by the yellow line for the Pro - Meta trajectory and purple for the Pro - Axa trajectory. Next to each smoothed expression plot is also the matching gene count plotted onto a PHATE map with logged counts of the expressed gene indicated as an overexpression coloured in red and no expression in yellow. As in previous marker analysis, promastigote markers shown in **C** - **H** show and over expression in promastigote clusters, with metacyclic markers in **E** - **G** being enriched in the metacyclic cluster as expected. Conversely, amastigote markers **K** - **N** are seen to be overexpressed towards the end of the Pro - Axa trajectory, each beginning its overexpression as it passes by the branching of the two trajectories in the Trans A cluster, as indicated by the red coloured cells before peaking in the Axa 1 cluster coloured in turquoise. Conversely, the Promastigote marker GT2 (**H**) is seen to begin under expression while passing through the Trans A cluster.

Lastly, genes selected for being most differentially expressed between branching trajectories were investigated and plotted below in Figure 3-49.

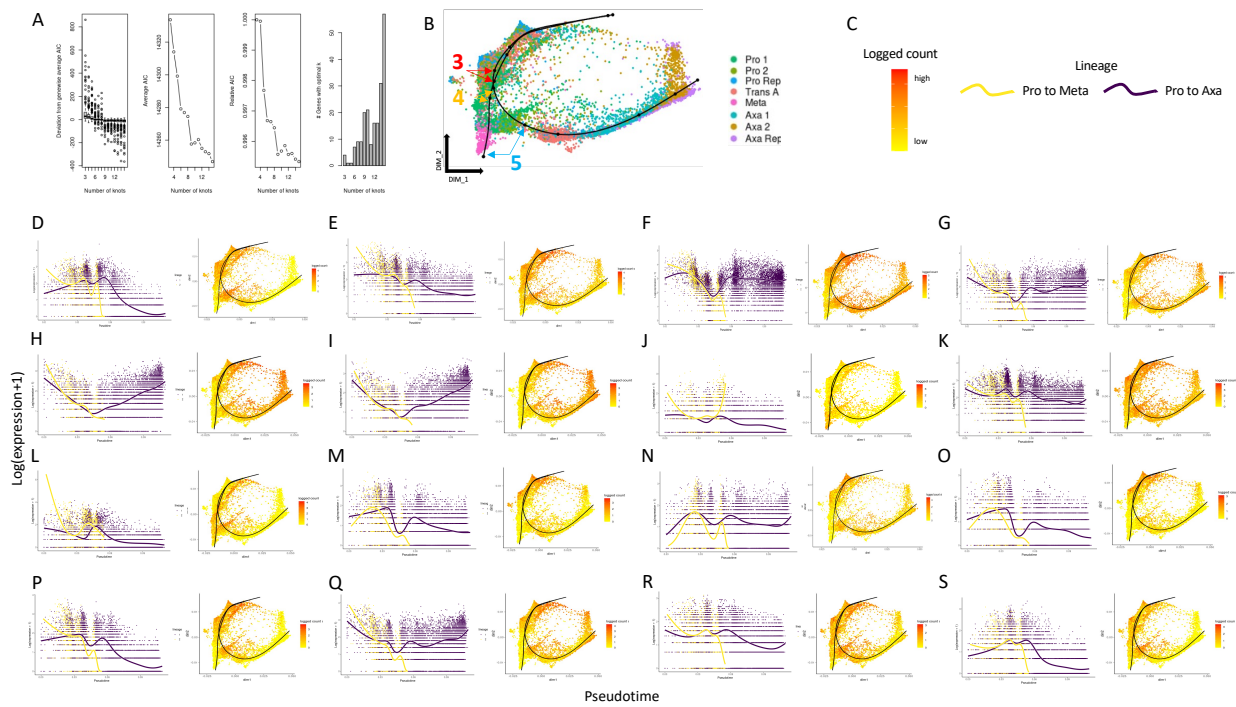


Figure 3-49 Top differentiated gene expression profiles across pseudotime, from start of trajectories to branching of trajectories.

TradeSeq negative binomial generalized additive model. The models are fit by running the **fitGAM** function. By default, the gene-wise NB-GAM estimates one smoother for every lineage using the negative binomial distribution and then run through diagnostic plots using the Akaike Information Criterion (AIC). **A)** Number of knots compared to deviation from genewise average AIC, Average AIC, Relative AIC and number of genes with optimal k. 9 knots are chosen as the best fit to optimally select for lowest number of knots and the greatest number of genes with optimal k. **B)** PHATE plot with trajectories drawn with a black line. Knots 3, 4 and 5 for each trajectory highlighted with arrows in red, yellow, and blue respectively. **C)** Expression colours for PHATE map markers, and trajectory colours for expression plot as shown in **D) – S)**. **D) – S)** Trajectory lineage pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of top differentiating markers with smoothed average expression of pseudotime 1 lineage indicated by yellow line and smoothed average expression of pseudotime 2 lineage indicated by purple line. Each point represents expression level for one cell's transcriptome coloured by cluster identity as their respective smoothed lineage colours. Included with each marker are logged marker transcript counts assigned to PHATE plots, coloured by over expression in red against cells not expressing in yellow, as in **C)**.

Shown in Figure 3-49 **A)** is the output of the **fitGAM** function from the tradeSeq package, as previously discussed in Figure 3-22. Here, 9 knots were selected for the trajectory across the PHATE map for the Pro - Axa trajectory, shown in **B)**, with 5 knots being drawn for the Pro - Meta Trajectory. In **C)** is a key for lineage smoothed expression plots and logged gene counts for the life cycle cluster marker plots shown below. Smoothed expression plots and their respective gene count plots were produced for the top marker genes expressed at the start of the trajectories, and between knots 3, 4 and 5, covering where the two trajectories branch. Some genes were repeated in the top markers between each knot, in which case the next overexpressing marker was selected. Shown in **D) - G)** are the top 4 differentiated markers from start of trajectories - **D)** NT1; LmxM.15.1240, **E)** Hypothetical protein; LmxM.30.0900, **F)** Alpha tubulin;

LmxM.13.0390, **G** Histone H1; LmxM.27.1190. **H - K** The top 4 differentiated markers from top 500 most differential genes expressed over pseudotime - **H** 60S ribosomal protein L21; LmxM.16.0460, **I** 60S ribosomal protein L28; LmxM.11.1130, **J** Promastigote surface antigen 38S; LmxM.12.0980, **K** Nuclear RNA binding domain; LmxM.31.0750. **L - O** The top 4 differentiated markers from between knots 3 and 4 (as shown in **B**) just as the two pseudotime trajectories separate - **L** META domain containing protein; LmxM.17.0890, **M** chaperonin HSP60, mitochondrial precursor; LmxM.36.2030, **N** ncRNA unspecified product; LmxM.23:ncRNA:rfamscan:218578-218718; **O** ADP/ATP translocase 1; LmxM.19.0210. **P - S** The top 4 differentiated markers from between knots 4 and 5 (as shown in **B**) after the two pseudotime trajectories separate and trajectory one ending in Meta cluster at knot 5 - **P** 3'-nucleotidase/nuclease; LmxM.30.2310, **Q** histone H3, putative; LmxM.16.0600, **R** fatty acid elongase; LmxM.14.0650; **S** Hypothetical protein, LmxM.30.0960.

3.4 Discussion

Currently, the timing and patterns of gene expression associated with life and cell cycle transitions seen in *L. mexicana* is not fully understood. Here, scRNA-seq was employed, so that the order of these transitional events and any potential stage specific patterns in gene expression could be identified, presenting a fuller understanding of the life cycle transitions. Currently, the application of scRNA-seq in the context of the *Leishmania* parasite is limited to analysis of *Leishmania* hybridisation *in vitro* (Louradour, et al., 2020), *L. major* infection in mouse models (Venugopal et al., 2022), and dual scRNA-seq of *L. donovani* infections isolated from lesions in mice (Karagiannis et al., preprint). Of note, is that the latter two studies were unable to isolate and describe *Leishmania* cells from infection samples. Additionally, analysis of cell development using pseudotime has not previously been applied in *Leishmania*.

Developmental progression through promastigote stages, starting in procyclic promastigote forms, and concluding in metacyclic promastigotes is a process termed metacyclogenesis. During this developmental process, intermediary stages are also present, but not limited to, nectomonad promastigotes and leptomonad promastigotes (Bates & Tetley, 1993a; D. Sacks & Kamhawi, 2001; D. L. Sacks & Perkins, 1984; Serafim et al., 2012). When analysed

transcriptomically by Inbar et al., (2017) and Coutinho-Abreu et al., (2020), distinct gene expression profiles were found to be associated with each stage. Additionally, investigations by Dillon et al., (2015) analysing *L. major* metacyclogenesis described 3,138 genes differentially expressed between forms, including GO terms associated with ATP synthesis being decreased between procyclic and metacyclic forms. While GO terms upregulated between stages were found to be associated with ATP binding and protein kinases (Dillon et al., 2015). Furthermore, proteomic analysis by Mojtabehi et al., (2008) and Amiri-Dashatan et al., (2020; 2020) recorded a rise in flagellum and cell motility proteins, associated with the increasing flagellum length in metacyclic forms, and a decrease in proteins associated with protein synthesis. Current thinking stipulates triggering events for metacyclogenesis being metabolically driven with environmental cues leading to stress induced differentiation (Bates, 2008; Bates & Tetley, 1993b; Kamhawi, 2006), although the full requirements and circumstances associated with the process have yet to be fully understood. Although questions in the literature have been raised as to how analogous *in vitro* cultures of promastigotes may be to vector derived samples (Alcolea et al., 2016), transcriptomic analysis by Inbar et al., (2017) discovered overall comparable gene expression. However, small variations found by Inbar et al., were described and associated with increased expression of autophagy associated protein turnover, metabolic processes, and amino acid transport (Inbar et al., 2017).

While overlaps between promastigote morphology and cell cycle stage have been previously postulated as being entwined *in vitro* (Bates, 1994; Rogers *et al.*, 2002; Wheeler *et al.*, 2011; Ambit *et al.*, 2011) the tools, until recently, have been lacking to analyse heterogeneous populations of cells progressing through life cycle and cell cycle phases at the single-cell resolution. When life cycle stages defined by morphology, as defined by Rogers *et al.* (2002) and Wheeler *et al.*, (2011), are compared to cell cycle stages as defined by number of nuclei and kinetoplasts, procyclic forms are found to resemble cells in G1 or post S-phase, nectomonads resemble cells in S-phase, and leptomonads resemble procyclic cell cycle stages but having divided with a longer flagellum than procyclics (Sunter and Gull, 2017). Transcriptionally, results in Figure 3-19 indicate procyclics to be overlapping with S phase, nectomonads with G2M, leptomonads with Early G1

and metacyclics in late G1, demonstrating a departure from morphological findings. This morphological and cell cycle overlap is further complicated when considering *Leishmania* cell division results in daughter cells with different morphologies (Wheeler et al., 2011; Wheeler et al., 2015). Timing of promastigote development was also investigated using pseudotime analysis (Figure 3-21 and Figure 3-22), where a total of 1318 markers were found across four promastigote life cycle stages (Figure 3-21). Additionally, transient expression of genes between non-replicating nectomonad, replicating leptomonad and the infective metacyclic stages were grouped into two further gene modules (Figure 3-22, D).

In this study, the disruptive technology of scRNA-seq was used to elucidate markers associated with life and cell cycle progression at the transcriptional level, to an extent not achievable with previous bulk RNA-Seq technologies (Fiebig et al., 2015). Investigation into promastigotes captured in the processes of differentiating into amastigotes was also explored (Figure 3-29), with 4393 markers associated with amastigote life cycle stages, 1937 potentially associated with promastigote cells still in the process of differentiating into amastigotes (Figure 3-29). For example, the analysis in this chapter reveals 269 promastigote markers and 267 amastigote markers not previously found using bulk RNA-Seq methodology (Figure 3-39) (Fiebig et al., 2015). Additionally, 207 markers were identified for the metacyclic promastigote stage previously unidentified by bulk RNA-Seq analysis, compared to 71 markers distinguished in bulk analysis (Figure 3-39) (Coutinho-Abreu et al. 2020). This higher resolution in marker analysis likely being found from deconvoluting heterogenous populations within the same samples and integrated samples, not possible in bulk RNA-Seq strategies and analysis. Where bulk RNA-Seq analysis can examine average total gene expression while scRNA-seq is able to examine and compare the transcriptome of up to 20,000 single cells concurrently (X. Li & Wang, 2021). Also of note, was the discovery of hypothetical protein expression associated with life cycle stages not previously identified by bulk RNA-Seq analysis. In total, marker analysis of integrated samples identified 331 hypothetical proteins potentially leading to new avenues of biology to be further investigated. Whether RNA levels are shared concomitantly with expressed protein levels is still to be fully understood

and is further considered in Chapter 4 with tagging strategies of 91 hypothetical proteins identified in this chapter from pseudotime analysis.

Known life cycle stage markers were used to distinguish cells from one life cycle stage from another (Figure 3-8 and Figure 3-9). In addition, cell cycle markers orthologues from *T. brucei* were used for labelling cells to discern the pattern of gene expression during mitotic growth of promastigote and amastigotes. Over three replicates, life cycle and cell progression were investigated, revealing a new 'transitional' cluster, which represents a cell population placed between promastigote and amastigote. Pseudotime analysis indicated this cluster is part of a pathway for progression from promastigote to axenic amastigote stages not dependant on moving through the infective metacyclic life cycle stage, which is unexpected as metacyclics are considered a key developmental intermediate between the fly and mammal (Gossage et al., 2003; Bates 2007; Giraud, et al., 2019). Speculation gives way to this cluster perhaps being an overlapping stage between differentially expressed polycistronic units between promastigote and amastigote stages, therefore indicating overlap in expression. This may also explain why a population that displays some overlap between promastigote and amastigote forms may have been missed in previous bulk transcriptional analysis, where the average expression may have been inadequate to isolate a stage between differentiating forms. This transitional cluster, named Trans A in this study, was seen both in individual replicates (Figure 3-10, C) and in the integrated datasets (Figure 3-33, A), and appeared to be more promastigote-like than amastigote-like both in clustering (Figure 3-10, D; Figure 3-33, D and in marker analysis (Figure 3-10, A); Figure 3-36A). Notably cluster Trans A displayed no cells differentially expressing metacyclic markers (Figure 3-40). Moreover, cluster Trans A cells appeared to be arrested in late G1 phase before committing to differentiation into amastigote forms (Figure 3-11, E); Figure 3-41, E). Further speculation as to what may progress the further development of this cluster into amastigote stages should first consider an environmental cue, as is well described for axenic amastigote differentiation in culture (Bates et al., 1992) which requires both an increase in temperature and acidic pH levels for committing to amastigote differentiation.

When examined using pseudotime analysis, two distinct trajectories were discovered. While both trajectories started in the promastigote clusters, when cells were ordered by gene expression profiles one trajectory ended in the metacyclic cluster and the other traversed through the Trans A cluster before ending in replicating axenic amastigotes. This transitional cluster was further corroborated when looking back at the life cycle markers used to distinguish clusters in early investigation, where promastigote markers (GP63, PFR2, GP63-2, and GT2) were shown to be under expression while passing through the Trans A cluster (coloured in red), and conversely, amastigote-associated genes (Tuzin, Amastin, A600-4, and Cathepsin) were seen to begin overexpression while passing through the Trans A cluster (Figure 3-48). Groups of genes that potentially regulate life cycle stage transitions were also identified in gene modules D (coloured in green) containing 293 genes and E containing 194 genes (coloured in pink) (Figure 3-47), with the former potentially being associated with metacyclogenesis and the later amastigogenesis due to the timing of expression patterns associated with the projected trajectory. A further 45 markers differentially expressed in the Trans A cluster identified for further investigation (Figure 3-37, B). As these 45 markers were differentially expressed in the Trans A cluster, they may serve as testable markers for the presence of expressed proteins *in vitro* cultures for comparison to *in vivo* cultures to identify if the Trans A cluster is solely an artifact of *in vitro* cultures. Additionally, for experimental consideration should be defining if this Trans A cluster may be a means to bypass the terminal metacyclic developmental stage, and thereby offer an alternative developmental path between promastigote and amastigote stages. Possibilities for what this Trans A cluster might represent include a potentially unexplored flexibility in the *Leishmania* life cycle, akin to what has been debated between Matthews and Engstler for the proliferative long slender bloodstream form (BSF) to proliferative procyclic forms (PCF) differentiation in *T. brucei* (Lisack et al., 2022; Matthews & Larcombe, 2022; Schuster et al., 2021).

Cell cycle labels using orthologues derived from Archer et al., (2011) for each of the replicates was also applied for Rep1 (Figure 3-11), Rep2_Pro (Figure 3-17), and Rep2_Axa (Figure 3-28) samples, and integrated samples (Figure 3-41), revealing hundreds of new markers associated with each cell cycle phase, where

global transcriptomic cell cycle analysis of *L. mexicana* is currently lacking in the literature. Additionally, cell cycle phases were associated with life cycle stages in promastigote forms (Figure 3-19), having been previously postulated and described by Wheeler et al., (2011), adding further evidence to the hypothesis that the proliferative cell cycle and differentiation are intrinsically linked as described above. Further investigations could consider methods of cell cycle arrest to disseminate enriched promastigote life cycle stages associated with each cell cycle stage, morphologically. Such as previously described applications for Artemisinin in cell-cycle arrest at the sub-G0/G1 phase (Sen et al., 2007), which may additionally be combined with marker expression analysis for the 45 identified Trans A genes.

Of note within these data is also the consistent finding of a non-coding RNA (ncRNA) in the top 10 markers of all sample analysed, being LmxM.23:ncRNA:rfamscan:218578-218718. This ncRNA was seen to be overexpressed in promastigote-like clusters and the Trans A cluster (Figures 3-10, 3-20, 3-29, and 3-36). ncRNAs have previously been reported in the literature as mediating gene expression by binding to RNA, RNA turnover, transcription, and translation (J. C. R. Fernandes et al., 2019; Hombach & Kretz, 2016; Mattick, 2001). Previous reports in bulk RNA-seq analysis have also found ncRNAs in top markers for *L. amazonensis* (Aoki et al., 2017). Additionally, ncRNAs have been described in their thousands in the *L. braziliensis* genome and validated transcriptomically (de Cássia Ruy et al., 2019; Torres et al., 2017). Further work for elucidating the interactome between RNA-binding proteins (RBPs) has recently been provided by de Pablos et al., (2019), where whole cell proteomes of *L. mexicana* procyclic promastigotes, metacyclic promastigotes and amastigotes were revealed. More recently work by Kalesh et al (2022), also provides evidence of coding and noncoding RNA-binding proteins, where 2,417 RNA-interacting proteins were identified by mass spectrometry analysis using tandem mass tag-labelling across promastigote and amastigote life cycle stages. This newly identified, promastigote specific, ncRNA in these data may present an intriguing addition to the previously described RBP population, which has previously been shown to be regulatory mechanism in *L. braziliensis* across developmental stages (Cristina Terrão et al., 2017; de Cássia Ruy et al., 2019). An additional example in *L. mexicana* for roles of RBPs includes a protective

response of RBPs being enriched in the nucleus following actinomycin D exposure (Názer & Sánchez, 2011). Lastly, roles for RBP proteins have been described by interaction of the methyltransferase Protein Arginine Methyltransferase 7 (PRMT7) in *L. major* by Ferreira et al., (2020) with proteins by arginine methylation of RBPs. Where 40 protein targets were identified, including 17 RBPs. Further investigation into the identified ncRNA, LmxM.23:ncRNA:rfamsan:218578-218718, could consider sequence analysis for potential binding sequences, purification for structural analysis by crystallography, alongside protein and mRNA cross-linking or quantitative mass spectrometry approaches to identify binding targets (Ramanathan et al., 2019).

In conclusion, demonstrated in this chapter is analysis and datasets of the life and cell cycle progression in *L. mexicana*, presented as a resource available for further investigation into identified markers and transcriptional expression of genes involved in differentiation and division.

3.5 Summary

- First use of scRNA-seq in *Leishmania mexicana* for analysis of life cycle and cell cycle transitions with pseudotime analysis.
- Shown to be able to map known classical life cycle markers to distinct clusters.
- More complexity to segregation of promastigote life cycle and cell cycling populations than expected.
- Identified a new transitional cluster, potentially a new life cycle stage in vitro for progression between promastigote and amastigote forms.
- Identified hundreds of new metacyclic markers.
- Identified the specific timing for life cycle progression both in promastigote only forms, and differentiation between promastigote and amastigote forms.

Chapter 4 Validation of transiently expressed hypothetical proteins by CRISPR-Cas9 mediated tagging

4.1 Introduction

In Chapter 3 *L. mexicana* transcriptional dynamics during life cycle progression were evaluated using scRNA-seq analysis of five time points covering *in vitro* promastigote to amastigote stages. In this chapter expression levels of 91 hypothetical proteins, 3 promastigote life cycle stage markers, and 2 cell cycle phase markers identified in Chapter 3 were validated using a high-throughput CRISPR-Cas9 tagging method, developed by the Gluenz laboratory (Beneke et al., 2017; Beneke & Gluenz 2019), and fluorescent microscopy. A hypothetical protein was also selected for being the top marker of the Trans A cluster, identified in Chapter 3. Comparisons between median fluorescent expression of tagged proteins and pseudotime gene expression were then made as an initial assessment into how gene expression identified by scRNA-seq might be representative of protein levels.

4.1.1 CRISPR tagging

The now widely used genome editing tool, termed Clustered Regularly Interspaced Palindromic Repeats (CRISPR) for its repetitive copies of palindromic 30 bp sequences, was first identified in prokaryotic adaptive immune systems (Mojica et al., 1993). A transformative step in its application as a tool for genome editing was the description of specific CRISPR-associated (Cas) genes adjacent to CRISPR sequences (Jansen et al., 2002). Early applications of CRISPR were described by (Barrangou et al., 2007) during their study of bacteriophage infections in *Streptococcus thermophilus*, used in cheese production. In their work, resistant strains of *S. thermophilus* were found to have newly inserted phage specific sequences at CRISPR loci (Barrangou et al., 2007). However, it would be 10 years before the pseudo-adaptive prokaryote immune system would be repurposed into a DNA editing tool by Doudna and Charpentier (Jinek et al., 2012). Doudna and Charpentier modified a Cas9 protein isolated from *Streptococcus pyogenes* and designed guide RNA to achieve Cas9-mediated cuts of purified DNA *in vitro* (Jinek et al., 2012). Additionally, a single guide RNA composed of several RNA subunits were identified in *S. pyogenes* (Jinek et al., 2012). The guide RNA is made up of a spacer sequence, forming the targeted sequence for cleavage, and repeated sequences, which forms a hairpin structure that allows binding with Cas9 proteins. Combined, this system provides a

programmable gene editing mechanism capable of providing double-stranded cleavage of targeted DNA sequences. The double strand break repair can be mediated by non-homologous end joining or by homology directed end joining, the latter being the more reliable repair strategy (Jiang & Doudna, 2017). Additionally, DNA target sequences matched to either side of the DNA cleavage site may contain sequences for insertion, thereby providing gene editing capability. To explore the biology of resultant mutants, sequences can be added to enable selection for antibiotic resistance of modified cells and tagging of proteins for isolation, modification, or quantification (Sansbury et al., 2019).

Applications of these systems have now evolved into high-throughput methodologies using modular plasmid-based systems. Such applications have been used in parasites such as *Plasmodium falciparum* (Nishi et al., 2021), *Toxoplasma gondii* (Young et al., 2019) and *L. mexicana* (Beneke et al., 2017; Beneke & Gluenz 2019). The latter uses cells expressing Cas9 and T7 RNA polymerase to which templates for single guide RNA and drug-selectable editing cassettes set in plasmids are easily generated and amplified via PCR for transfections of gene knockouts or tagging methodologies (Beneke & Gluenz 2019). An online platform, also generated by Beneke et al. (2017) (<http://leishgedit.net/>), provides automated primer design for genes of interest. This method improves on previous methodology using stepwise cloning of DNA sequences into plasmids for DNA constructs to induce homologous recombination mediated genome edits (Cruz & Beverley, 1990). Untemplated repair of CRISPR-Cas9 DNA double strand breaks in *L. mexicana* uses microhomology-mediated end joining instead of nonhomologous end joining typically used in vertebrates (Zhang & Matlashewski, 2015). Additionally, homology flanks placed in repair templates can be as short as 24 nucleotides in length and ensure accurate modification of intended targets (Beneke et al., 2017).

The Gluenz CRISPR-Cas9 system can be applied to 96-well based transfection assays allowing the generation of modified cell lines in multiplex. The Gluenz method was employed in this chapter to tag 96 proteins of interest identified through scRNA-seq marker and pseudotime analysis presented in Chapter 3. 91 of the selected proteins were hypothetical proteins not previously described in the

literature. Via the modification of 96 proteins of interest with a mNeonGreen tag at their C-terminal, we aimed to compare pseudotime gene expression profiles with expressed protein through promastigote and amastigote stages by fluorescent microscopy.

4.2 Aims

- Identify expression of hypothetical proteins discovered during *L. mexicana* life cycle development.
- Assess if tagged protein expression matches described mRNA expression identified by scRNA-seq pseudotime analysis of life cycle development.

4.3 Results

4.3.1 Generation of 96 tagged cell lines by CRISPR-Cas9

Hypothetical proteins were selected for displaying variation in expression along pseudotime trajectories drawn over life cycle progression. Following identification of 96 genes with varying expression across pseudotime (Chapter 3) and selecting for those with only a single 3' single guide RNA (sgRNA) target via LeishEdit, primers were designed for C-terminal tagging with mNeonGreen using LeishEdit (<http://leishgedit.net/>). The C-terminus was selected as commonly being a more efficient target over the N-terminus in *Leishmania*, as described by Yagoubat, et al., (2020). Of the 96 selected proteins for tagging, 91 were hypothetical proteins, 3 were markers for proteins expressed in the promastigote life cycle stage, and 2 selected for proteins expressed for cell cycle progression and regulation. For the 3 promastigote stage markers, paraflagellar rod protein 2 (PFR2), 3'nucleotidase/nuclease (3'-Nuc), major surface protease gp63 (MSP) were selected as these genes are known to possess only a single 3' sgRNA target, limiting off target tagging, and are well described as over-expressed in promastigote stages compared to amastigote (Guimarães-Costa et al., 2014; Holzer et al., 2008; Marshall et al., 2018; Mishra et al., 2003; Wheeler et al., 2015). The DOT1A and DOT1B proteins were selected for DNA replication and cell cycle progression indicators. Previous studies described the homologue in *Trypanosoma cruzi* (*T. cruzi*) of DOT1A as adding mono- and

dimethylation of lysine 76 (H3K76) during late G2 and mitosis, while DOT1B trimethylates lysine 76 in all stages of the cell cycle (Nunes et al., 2020). Over expression of DOT1A homologues in *Trypanosoma brucei* (*T. brucei*) caused early H3K76 methylation of cells already in S phase instead of a general increase of methylation (Gassen et al., 2012). *T. brucei* cells were also shown to demonstrate over-replication when DOT1A was over-expressed, and conversely when inhibited by RNAi, DNA replication was halted (Gassen et al., 2012). DOT1B has also been linked with antigenic variation and developmental differentiation (Figueiredo et al., 2008, 2009). The hypothetical protein LmxM.08.0810 was selected as a differentially expressed marker of the Trans A cluster identified in Chapter 3. This hypothetical protein has orthologues for 19 amastin-like proteins in *L. mexicana* with amastin-like orthologues in at least 10 other *Leishmania* species, and 2 amastin surface glycoprotein orthologues in *T. cruzi*, identified via TriTrypDB.

Primers were designed for sgRNA for each gene for targeting with a drug selectable editing cassette template using the modular pPLOT plasmid, along with the common sgRNA scaffold primer G00, as described in Beneke et al., (2017) and Beneke & Gluenz (2019). Here the plasmid template pPLOT-mNeonGreen blast-blast, kindly donated by the Gluenz group, was used for drug selection using blasticidin. Following plasmid amplification in *E. coli*, an initial restriction enzyme digest test was performed to ensure plasmid purification and validation, as shown in Figure 4-1.

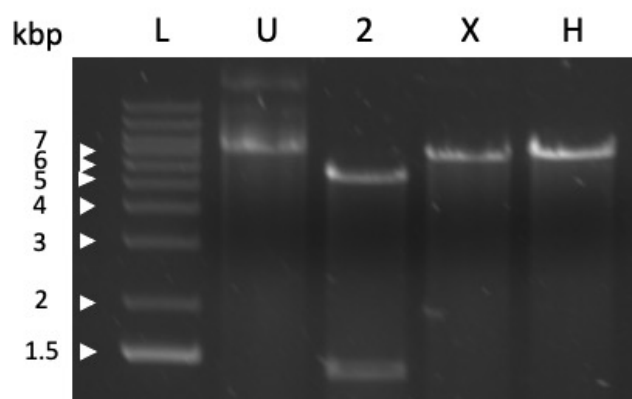


Figure 4-1 pPLOT-mNeonGreen blast-blast plasmid restriction enzyme digest.

Restriction enzyme digest using XcmI and HindIII restriction enzymes. Lanes annotated for 1 Kb plus ladder (L), uncut plasmid control (U), double cut using XcmI and HindIII restriction enzymes (2), and single cuts of XcmI (X) and HindIII (H) restriction enzymes.

Amplification of 96 DNA targets for downstream forward and reverse primers with 3' tagging cassette for the pPLOT plasmid donor fragment was performed in a 96-well plate by PCR and validated by agarose gel electrophoresis shown in Figure 4-2.

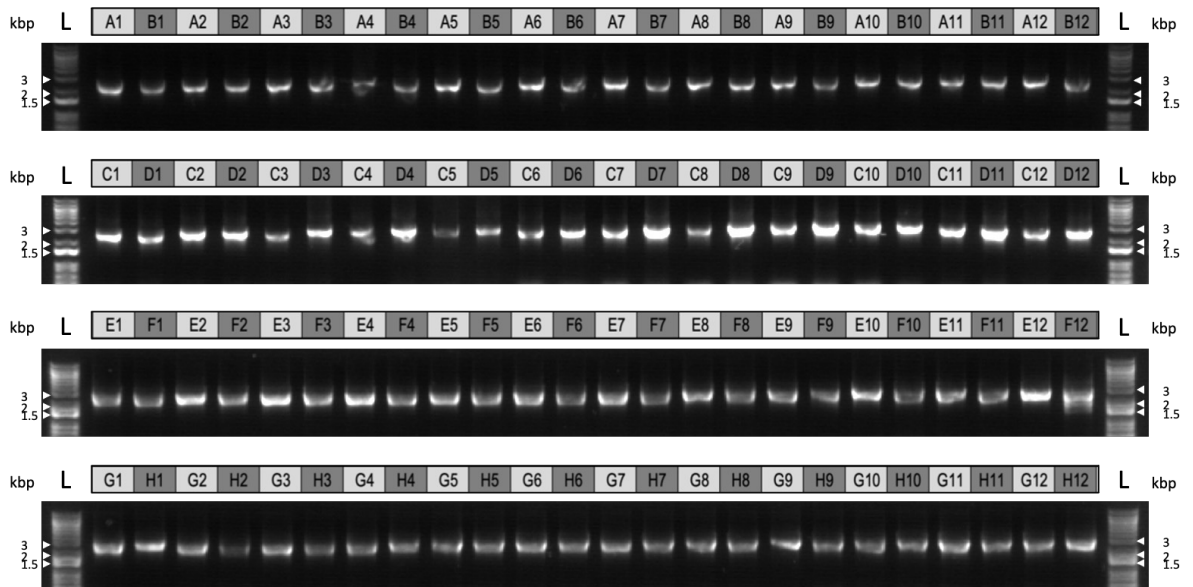


Figure 4-2 PCR products for DNA amplification of 3' tagging cassette for pPLOT mNeonGreen plasmid donor fragments for tags of hypothetical and life cycle stage marker proteins.

PCR amplification of 3' tagging cassette for pPLOT mNeonGreen plasmid donor fragments for 90 hypothetical proteins with 4 life cycle stage markers and 2 cell cycle regulation markers.

All 96 PCRs were successful to generate donor DNA for the blasticidin pPLOT plasmid. To generate sgRNA templates in a 96 well plate PCR was again used to amplify templates, the products of which were also run in agarose gel electrophoresis, shown in Figure 4-3

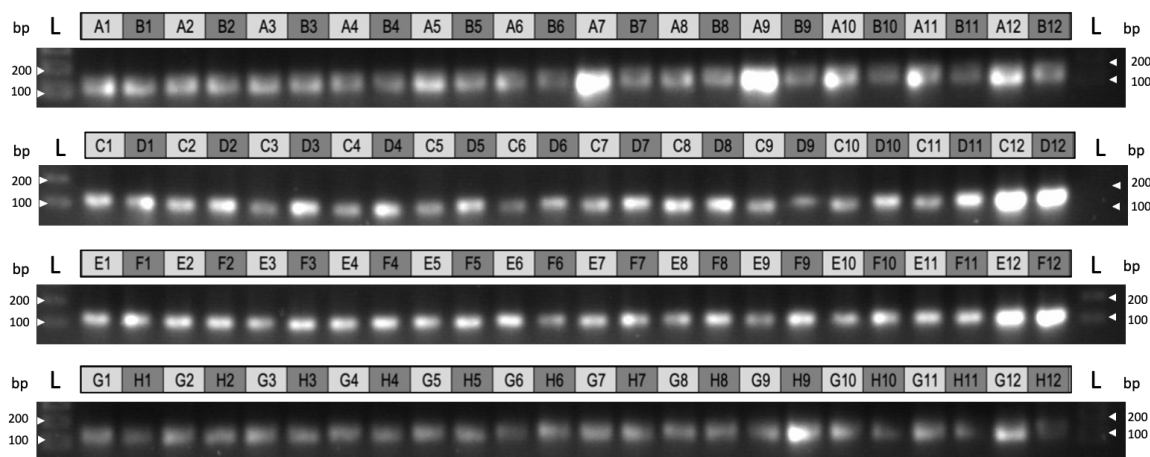


Figure 4-3 PCR products for DNA amplification of downstream guide RNAs for tags of hypothetical and life cycle stage marker proteins.

PCR amplification of downstream guide RNAs for 90 hypothetical proteins and 6 life cycle stage marker tags.

All 96 sgRNA templates were successfully generated and products for each were pooled with the matching donor fragments, heat sterilised and used for transfections with *LmexCas9T7* cells expressing Cas9 and T7 RNAP, also kindly donated by the Gluenz laboratory. *LmexCas9T7* cells were grown in M199 media with Cas9 and T7 RNAP lines carrying a Nourseothricin (*NTC*) and Hygromycin resistance cassette and selected for using Hygromycin at 32 $\mu\text{g}/\text{mL}$ and Nourseothricin at 50 $\mu\text{g}/\text{mL}$. Transfections are described in detail in Section 2.1.10. Following transfections, tagged cell lines were selected by blasticidin resistance, at a concentration of 11 $\mu\text{g}/\text{mL}$, until drug resistant populations emerge at three weeks post-transfection. To increase transformation efficiency, cells were first recovered in ‘Magic’ M199 media before moving to HOMEM, as described in Chapter 2. Antibiotic resistance clones, putatively expressing the tagged proteins, were adapted to complete HOMEM at 25°C and cultured in plates sealed with parafilm for comparison to expression levels of the *L. mexicana* scRNA-seq analysis, removing potential differences in expression due to variations in growth conditions.

4.3.2 Comparing pseudotime gene expression with tagged fluorescent expression of proteins.

Fluorescence microscopy of each tagged cell line stained with Hoechst 33342 was performed using a 20x objective and images of promastigote and axenic amastigote stages taken. Timepoints selected were for logarithmic growth in both stages, and due to time constraints, one timepoint for each life cycle stage

was used to image tagged protein expression. Growth phase timepoints were taken at 48 h for promastigotes and 240 h for amastigotes, starting at a density of 1×10^5 cells/mL to match timepoints taken for scRNA-seq samples, described in Chapter 3. At least 100 cells for each timepoint were imaged and quantified for expression of mNeonGreen using Fiji software analysis (Schindelin et al., 2012), where thresholds and fluorescence were assessed, quantified, and calculated. All images were acquired using the same settings and exposure times on a Leica DMi8 S platform. Mean grey value and medians were selected to assess intensity per sample timepoint for cells expressing mNeonGreen. To compare predicted expression dynamics inferred by scRNA-seq pseudotime analysis (Chapter 3) with expression of selected labelled markers genes (this Chapter), medians were then used for plotting relative expression between promastigote and axenic amastigote samples (Figure 4-4). Medians of expression were selected to limit any effect of outliers.

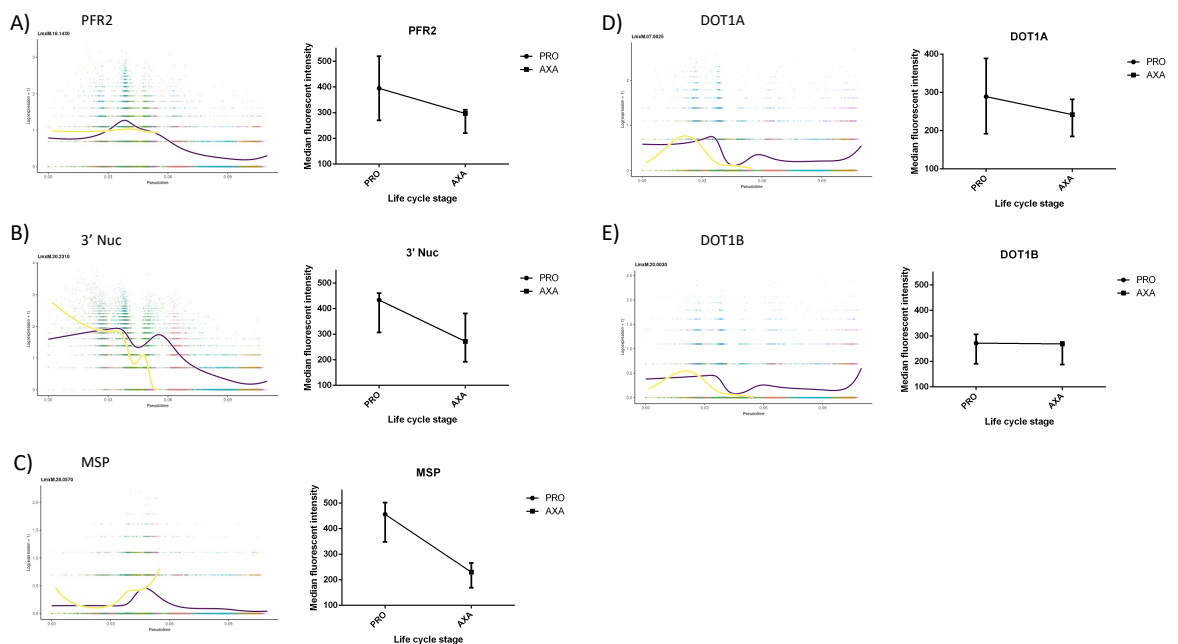


Figure 4-4 Expression of selected life cycle stage markers across pseudotime and fluorescent tag expression in promastigote and amastigote timepoints.

Trajectory lineage pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of selected life cycle markers with smoothed average expression of pseudotime 1 lineage indicated by yellow line and smoothed average expression of pseudotime 2 lineage indicated by purple line. Each point represents expression level for one cell's transcriptome coloured by cluster identity as in Figure 3-41. Selected markers for promastigote stage were paraflagellar rod protein 2 (PFR2) in **A**), 3'nucleotidase/nuclease (3'-Nuc) in **B**), major surface protease gp63 (MSP) in **C**) and two markers for cell cycle progression and replication in DOT1A and DOT1B (**D**) and **E**), respectively). Included with each pseudotime trajectory plot is the respective markers median fluorescence expression from promastigote and axenic amastigote timepoints as calculated by Fiji image analysis.

Promastigote markers PFR2, 3'Nuc and MSP showed higher expression according to mNeonGreen intensity quantification in the promastigote stage compared to the amastigote stage, with expression being reduced by 20%, 19.4%, and 49.6% respectively. This pattern broadly matched the transcript levels changes across pseudotime observed in Chapter 3. DOT1A was seen to be more expressed in promastigote stages by mNeonGreen quantification - its expression was reduced by 17.2% in amastigote forms, where interestingly a large expression of DOT1A was also seen in pseudotime analysis in promastigote stages compared to amastigote. However, of note are that error bars indicate wide variation in analysed promastigote cells. Further investigation would benefit from analysis of additional timepoints to assess if high variation is consistent in promastigote forms. Alternatively, cell cycle synchronisation could be considered to determine DOT1A expression during cell cycle progression. DOT1B displayed the least change between promastigote and amastigote stages, only changing by 1.1% in mNeonGreen expressing cells, matching the pseudotime expression profile. The hypothetical protein LmxM.08.0810, which was selected as a Trans A cluster marker from scRNA-seq analysis in chapter 3, was then compared to mNeonGreen expression in Figure 4-5.

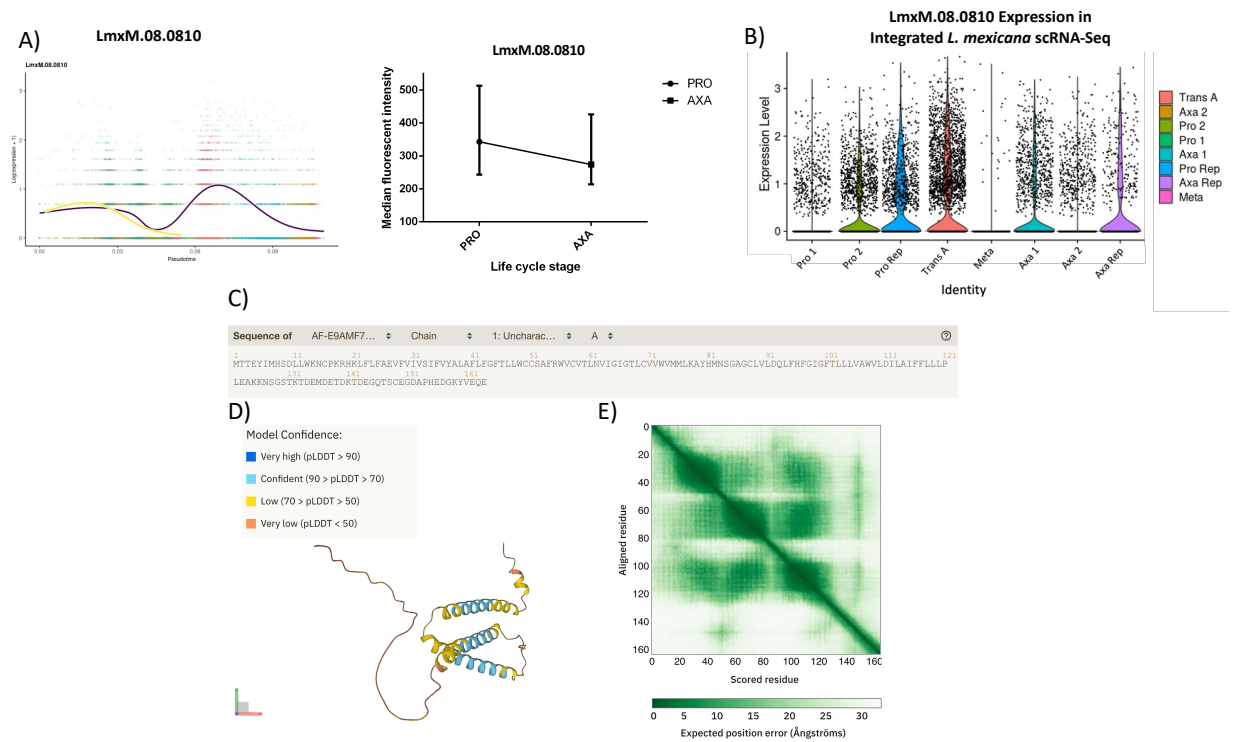


Figure 4-5 Expression of the hypothetical protein LmxM.08.0810 across pseudotime and fluorescent tag expression in promastigote and amastigote timepoints.

A) Trajectory lineage pseudotime (x-axis) against expression levels (y-axis; $\log_2(\text{expression} + 1)$) of selected life cycle markers with smoothed average expression of pseudotime 1 lineage indicated by yellow line and smoothed average expression of pseudotime 2 lineage indicated by purple line. Each point represents expression level for one cell's transcriptome coloured by cluster identity as in Figure 3-41. Here, expression of LmxM.08.0810 is seen to increase in the Trans A clustered cells coloured in red. Included with the pseudotime trajectory plot is the respective median fluorescence expression from promastigote and axenic amastigote timepoints for hypothetical protein LmxM.08.0810. **B)** Violin plot for LmxM.08.0810 expression over clusters in integrated data set. **C)** Amino acid sequence for LmxM.08.0810. **D)** AlphaFold structure prediction for LmxM.08.0810, with per-residue confidence score indicated by colour, with 90 to 70% confidence coloured in turquoise, 70-50% confidence coloured in yellow, and less than 50% confidence plotted in orange. **E)** Predicted aligned error coloured at position (x, y) indicating AlphaFold's expected position error at residue x, when the predicted and true structures are aligned on residue y (Jumper et al., 2021).

Despite the hypothetical protein LmxM.08.0810 having multiple orthologues in amastin like proteins in other *Leishmania* and *Trypanosoma* species, expression intensity of the mNeonGreen tag was seen to be reduced by 20.1% in the amastigote stage, however, large variation in expression was again seen in this cell line. This lowering of LmxM.08.0810 expression in amastigote forms is, however, consistent with previous observations at low-resolution cluster analysis described in Chapter 3 (Figure 3- 33, D). Furthermore, shown in B is the expression profile for LmxM.08.0810 from scRNA-seq integrated samples analysed in Chapter 3, drawn by Violin plot. This demonstrates the highest expression of LmxM.08.0810 in the Trans A cluster, with high expression also present in proximal promastigote and amastigote cluster to Trans A, displaying a slightly higher enrichment of expression in promastigote clusters. Additionally,

the Trans A cluster expression was found to be closer in expression profiles to promastigote associated markers. This indicated that the parasites represented in the Trans A cluster, and identified here through this marker genes, may occur more frequently in promastigote-rich populations compared to amastigote-rich populations, and have promastigote-like morphology. For further consideration is the overlap with similar expression profiles for proximal promastigote and amastigote clusters. However, further investigations and validations at both protein and gene expression levels are required. The predicted structure of LmxM.08.0810 was also analysed by AlphaFold (a recently developed open access predictive protein structure database developed by Google's DeepMind team (Jumper et al., 2021; Varadi et al., 2021), in Figure 4-5 D. Of note was the generally low confidence scores associated with residues aside from helix domains, resembling amastin orthologues. In E is displayed the expected distance error in Ångströms, coloured in green, where dark green indicates a low error for the predictive model. The colour along x,y corresponding to the expected distance error in residue x's position, when the prediction and true structure are aligned on residue y. The three darker squares along x,y indicated the confident alignment scores of the helix domains.

A comparison between pseudotime gene expression and mNeonGreen protein tag expression was then made for all 96 tagged cell lines in Figure 4-6.

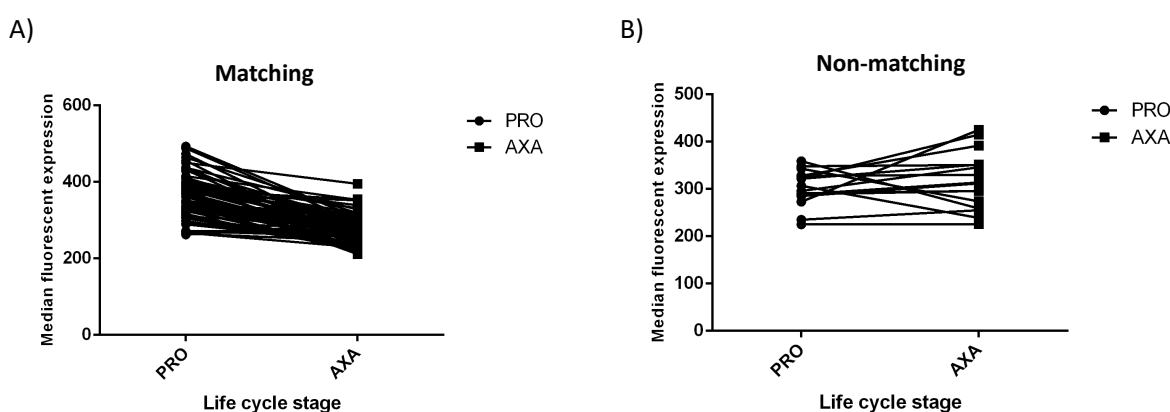


Figure 4-6 Median fluorescence expression from promastigote and axenic amastigote timepoints for all 96 tagged proteins separated by those matching pseudotime expression at respective life cycle stages and non-matching.

A) Plots of 80 matching tagged protein expression scores by median fluorescence to their respective pseudotime expression at respective life cycle stage time-points. B) the same as A) but plotting 16 tagged proteins whose median fluorescence did not match their pseudotime expression plots.

Comparing protein levels for all 96 mNeonGreen expressing cell lines to gene expression by pseudotime, 80 out of 96 tagged proteins matched pseudotime expression values, indicating pseudotime expression could be a good representation of protein levels. Lists of the 80 matching and 16 non-matching expressions between pseudotime and fluorescent protein quantification can be found in Appendix IV. Of note, is the overall downward trend in expression in matching protein expressions seen in amastigote timepoints.

4.4 Discussion

To compare pseudotime gene expression inference with transcribed protein levels in *L. mexicana*, a high-throughput CRISPR/Cas9 methodology was employed to tag 96 proteins of interest with mNeonGreen. Here is presented an assessment between gene expression and protein levels by fluorescence microscopy as an initial investigation, further experiments being limited by time constraints. Although 80 out of the 96 proteins (83.3%) generally matched pseudotime expression profiles, analysis was limited to one time point in each life cycle stage containing heterogenous populations of cells and limited to basic analysis by microscopy of fluorescent expression. Expression of promastigote markers by scRNA-seq pseudotime analysis matched tagged mNeonGreen expression, with greater inconsistencies seen in cell cycle progression markers DOT1A and DOT1B. The hypothetical protein LmxM.08.0810, was selected as the top marker for the Trans A cluster, described in Chapter 3. While large variation in protein expression was found, the general trend of decreased mRNA expression in trajectories was seen in protein levels. Perhaps more interestingly, tagging of this hypothetical protein produced here does provide opportunities for further investigation into where and how the Trans A cluster may be represented at the protein level, and if this cluster may represent a novel life cycle stage. However, this investigation remains in very preliminary stages, requiring further study to assess if analysis of scRNA-seq mRNA expression by pseudotime may more accurately infer protein expression.

4.4.1 mRNA and protein expression in *Leishmania*

The dogma of DNA to RNA to protein is still being explored in numerous biological systems (G.-W. Li & Sunney Xie, 2011; Schneider-Poetsch & Yoshida,

2018). However, mechanisms for transcription and translation are convoluted and a correlative relationship between each is often not easily described (Buccitelli & Selbach, 2020; Taylor, 2006). While the majority of eukaryotes use a monocistronic mRNA expression process (Cramer, 2019; Panigrahi & O'Malley, 2021), *Leishmania* and other kinetoplasts use a polycistronic expression process, which instead relies on post-transcriptional regulation to modify expression profiles (Clayton, 2019; Martínez-Calvillo et al., 2004). Thus, transcript abundance levels, as measured here by scRNA-seq, are the result of post-transcriptional mechanisms, such as mRNA maturation and decay rates. Still, the lack of specificity of regulation at the transcriptional level poses increased quandaries on how it may be possible to infer protein expression from mRNA. Of note, is the large proportion of RNA binding proteins found in proteomic analysis in trypanosomatids, which are thought to act as gene regulators (de Pablos et al., 2019). The polycistronic segments observed in trypanosomatids have previously been shown to be comprised of several open reading frames (Martínez-Calvillo et al., 2004; Myler et al., 1999), mainly containing no introns (Clayton, 2019). Polycistronic segments, that can be up hundreds of genes in length, are transcribed into precursor mRNAs (Ivens et al., 2005) and are transcriptionally processed to form individual mature mRNAs for each gene located in these regions (Liang et al., 2003; Martínez-Calvillo et al., 2004). Currently, limited association between gene function and location of polycistronic loci have been described in the literature (Beverley, 2002; Ivens et al., 2005). Mature mRNA is generated through trans-splicing of the of 39 bp mini-exon splice leader mRNA cap at the 5' (Matthews et al., 1994; Liang et al., 2003; Kramer and Carrington, 2011). In the same step of mature mRNA processing, polyadenylation also occurs to add stability to mRNAs (Agabian, 1990; E. J. Bates et al., 2000; Decuyper et al., 2005; X. H. Liang et al., 2003). It has also been described that balance between mRNA degradation mediated by maturation processes can affect transcription abundance, suggesting a function for mRNA degradation in defining mRNA levels (S. Archer et al., 2008; Fadda et al., 2014). Furthermore, knockouts of individual translation initiation proteins have been described as reducing global protein levels with knock-on effects on morphology and parasitaemia (Baron et al., 2021; Fadda et al., 2014). Epigenetic regulation has also recently been shown to affect regulation of transcription in *Leishmania* when bromodomain factors associated with histone acetylation were deleted by

CRISPR/Cas9 with deletions reducing global RNA polymerase II transcription (Jones et al., 2022). Additional considerations should also be given to how protein folding mechanisms may contribute to post-translational levels of proteins required at specific life cycle stages (Clayton & Shapira, 2007).

Recently, a review by Cortazzo da Silva et al. (2022) compared transcriptomic and proteomic data for differentiation into metacyclics and amastigotes forms across several *Leishmania* species. Although individual data was taken from different studies, each using their own statistical methodology, a proof-of-concept was found in forty-six common genes across the studies. For comparison between transcriptomic and proteomic results when analysing metacyclogenesis (Alcoleaid et al., 2019; Serafim et al., 2012), six studies were considered; three for transcriptomic data from Almeida et al., (2004), Dillon et al., (2015), and Inbar et al. (2017), and three for proteomic data sets; Mojtahedi et al., (2008), Amiri-Dashatan et al., (2020) Amiri-Dashatan et al., (2020). Of the forty-six common genes found in all studies, 28 (60.9%) were found to have a positive correlation between transcriptomic and proteomic levels during metacyclogenesis. The most consistent results were found in gene-expression related genes, where a consistent down-regulation was found (Cortazzo da Silva et al., 2022); *i.e.*, a decrease in transcription and translation described in metacyclic forms of the promastigote by Kloehn et al. (Kloehn et al., 2015). These results give some credence to the theory that despite constitutive expression, RNA levels are reflected at the protein level, and can lead to comparative differences represented at the proteomic level (Cortazzo da Silva et al., 2022).

Furthermore, when the same review of transcriptomic and proteomic dataset was performed by Cortazzo da Silva et al. (2022) for amastigogenesis in forty-three common genes, thirty (69.8%) were found to correlate using similar comparative methods for the metacyclogenesis comparisons. In these analyses eight data sets were analysed, five using transcriptomic data (Alcolea et al., 2010; Almeida et al., 2004; Holzer et al., 2006; Leifso et al., 2007b; Saxena et al., 2007) and three proteomic data (Walker et al., 2006; Leifso et al., 2007; Brotherton et al., 2010). Of note is that amastigotes were sourced from various differentiation processes; Almeida et al., (2004); Holzer et al., (2006), and

Leifso et al., (2007) generating amastigotes from BALB/c mice lesions, Alcolea et al., (2010) isolated amastigotes from macrophage lysis, while Walker et al., (2006), Saxena et al., (2007) and Brotherton et al., (2010) used axenic amastigote cultures. While these environmental aspects should be kept at the forefront for deliberation, there remains significant similarities, not least due to constitutive expression for the majority of genes (Cohen-Freue et al., 2007; Shapira et al., 1988). For the thirty genes that were found to correlated from the forty-three common genes analysed, twenty-six displayed a decrease in expression, matching the overall trend displayed in protein expression shown in Figure 4-6. This finding is also consistent with previously reported results in a reduction of mRNA and protein found in amastigotes (Shapira et al., 1988; Coelho et al., 2012). This general trend in mRNA levels conveying protein levels is seen in this initial assessment of fluorescently tagged hypothetical proteins, where 83.3% of cell lines expressing mNeonGreen matched scRNA-seq pseudotime gene expression profiles.

However, contrary evidence in the literature does exist for mRNA levels being poor indicators of proteomic levels represented in life cycle stages of *Leishmania*. Examples include Lahav et al., (2011) where dynamic changes between mRNA and protein levels, globally, may be responsive to environmental cues when examined in *L. donovani*. While Lahav et al., (2011) further describe mRNA levels being most influential in early differentiation stages, translational and post-translational processes were described as more significantly deterministic of protein levels than mRNA levels. Of note was a finding of only 20-30% were found to be correlated between mRNA and protein expression levels during amastigogenesis (Lahav et al., 2011). More recently, mRNA levels have been described as being poor indicators of whole cell protein levels in *L. mexicana* by de Pablos et al., (2019), when comparing whole cell proteomic data to bulk RNA-Seq of the same life cycle stages by Fiebig et al., (2015). Of interest would be a comparison between recorded whole cell protein levels by Pablos et al. and scRNA-seq expression levels as described in chapter 3 here. However, further considerations should also be given to examples of stage specific protein expression associated with survival strategies. One example being the expression of lipophosphoglycan (LPG) on the surface of promastigotes; where protein expression to allow attachment to the sand-fly midgut (Alcolea et al., 2014a;

Wilson et al., 2010) may indicate an underlying regulation of mRNA and protein expression control mechanisms. Further examples include, but are not limited to, amastin mRNA expression associated with amastigote membrane formation during macrophage phagocytosis (Alcolea et al., 2014; Fiebig et al., 2015).

Further investigation would certainly benefit from more timepoints and possible synchronisation of cells to compare with the 5 timepoints used here for sample collection across promastigote to amastigote life cycle development. Of note is also the high variance seen in expression between fluorescent intensity in samples, likely due to this heterogenous mix of cells' hence, only general trends in expression are inferred here as an initial investigation. Additionally, further methods of protein quantification should be considered with these 96 cell lines. For instance, western blots could be performed using the MYC epitope tag inserted adjacent to the mNeonGreen tag as part of tagging procedure. Quantification by western blot could also be performed over 5 timepoints to assess temporal fluctuations in protein expression as the parasite develops through life cycle stages. Alternative high-throughput quantification of the mNeonGreen tag could be performed in 96-well based assays using flow cytometry over the five timepoints used for scRNA-seq sample collection; in fact, this was attempted, but requires optimisation of detection ranges between all 96-samples.

In this preliminary investigation, by focusing on mainly hypothetical proteins, with little or no previous data on their expression before this thesis, we provide strong evidence that changes in RNA levels detected by scRNA-seq are reflective of changes in protein levels. Hence, scRNA-seq provides a strong route to infer genes whose changing expression drives phenotypic changes.

4.5 Summary

- Tagging of 96 proteins for initial comparison with pseudotime gene expression revealed matching expression for 80 proteins through preliminary assessments by fluorescent microscopy.
- Further validation required to compare protein expression levels with gene expression by pseudotime trajectory inference.

Chapter 5 Single cell RNA-sequencing analysis of *Leishmania* infecting human macrophages

5.1 Introduction

The host's response has a critical role to play in ablating parasite infection outcomes (Dos et al., 2019). As such it is important to understand how host cell gene expression varies in response to infection as well as how host gene expression may be driven by the parasite. Parasites may modify host cell signalling or modulate transcriptional responses to ablate the host immune response toward the infection, as well as salvage host cell nutrients to promote parasite proliferation (Schmid-Hempel, 2009). Macrophages differentiate from myeloid precursors, most typically derived from monocytes (Geissmann et al., 2010). Macrophages are professional phagocytotic cells and are generally considered as the main host cell for *Leishmania* to reside within during mammalian infections, where the parasite differentiates into amastigote forms within the phagolysosomes and vacuoles (Kaye & Scott, 2011). In addition to macrophages, dendritic cells and neutrophils have also been described as targets for *Leishmania* (Peters et al., 2008; Ng et al., 2008; Charmoy et al., 2010). Previous examples have described *Leishmania* modulating inflammation responses in macrophages: gamma interferon (IFN- γ) production has been shown to be attenuated, as well as inhibiting the major histocompatibility complex class II molecules (Geissmann et al., 2010; Proudfoot et al., 1995), the Janus kinases/Stat1 axis (Nandan & Reiner, 1995b), and Mitogen-activated protein kinases (MAPKs) (Nandan et al., 1999b).

Previous studies of the transcriptomic changes associated with macrophages infected with *Leishmania* by Fernandes et al. (2016) were undertaken to investigate host-cell and parasitic interactions. Fernandes et al., (2016) used bulk RNA-Seq dual profiling of two cutaneous leishmaniasis causing parasitic species, *L. major* and *L. amazonensis*, in combinations with human and mouse macrophages. Great similarities were discovered in expression of orthologous gene sets between the two parasite species, in the respective host responses during infections (Fernandes et al., 2016). Common *Leishmania* transcripts found to be down-regulated in human macrophages were META domain containing protein (META1), zinc-metalloprotease GP63, and Surface Antigen Protein-2 (PSA-2); all being metacyclic promastigote stage markers. Conversely, these markers were all up regulated in mouse models (Fernandes et al., 2016). *Leishmania* transcripts up-regulated in human cells were RNA binding protein 5

(RBP5) and cathepsin L-like proteases in human macrophages, which in mouse models were seen to be down-regulated. Differences were found in infection responses between human and murine macrophages, however, where 1,133 host transcripts were upregulated and 754 down-regulated in human cells after *L. major* infection, with murine cells upregulating 862 transcripts and 764 down-regulated.

In the human macrophages analysed by Fernandes et al. (2016), of the 328 unique transcripts described, many cytokines associated with immune cell responses were upregulated, including tumour necrosis factor (TNF), interleukin-6 (IL-6), and interleukin-1 β (IL-1 β). Also of note was the transitory nature of the transcriptomic response, where differential expression analysis displayed an increased host response only 4 hours post-infection, while 24 h - 72 h timepoints showed fewer significant differences in transcriptomic expression of the number of differentially expressed genes for these pathways when compared to uninfected controls. This transient response to infection is potentially a novel feature of *Leishmania* infections in comparisons to other macrophage dependant infection responses, such as *Mycobacterium tuberculosis* (Sasindran & Torrelles, 2011; Chandra et al., 2022), which produces a sustained inflammatory response.

Similar studies of dual transcriptome analysis in *L. major* and murine macrophages were described by Dillon et al., (2016), where investigation into differentially expressed gene and enrichment in Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis demonstrated overlap in TNF pathway enrichment, Jak-STAT, and MAPK pathways, as described by Fernandes et al. (2016). Conversely, anti-inflammatory, tissue growth, and repair pathways were enriched 4 h post-infection. Dillon et al., (2016) noted similarities in such responses to *L. major* infections in murine macrophages and those described by Fleming et al., (2015) using macrophages exposed to lipopolysaccharides. Consistent with the study by Fernandes et al. (2016) was the transient expression of the transcriptomic immune responses in murine macrophages, having a significant response 4 h post-infection with significantly reduced expression in subsequent timepoints, with the exception of ATP production and other glycolytic pathways, which were also enriched 24 h post infection but not thereafter (Dillon et al., 2016).

In addition to the above approaches, Fiebig et al., (2015) used bulk RNA-Seq in *L. mexicana* over promastigote, axenic amastigote and intracellular amastigotes infecting mouse bone derived macrophages, to compare the transcriptomes of the different life cycle forms, excluding the host. Differential expression comparison of promastigote and intracellular amastigotes provided 3,832 genes between the two life cycle stages. Amongst the genes downregulated in amastigote forms were transcripts associated with flagellum production, whereas genes found to be upregulated included peptidases, transporters, and membrane bound surface proteins, with 936 differentially expressed genes identified.

Currently, examples of published scRNA-seq studies in *Leishmania* infections are limited to a single study by Venugopal et al., (2022), who used *in vivo* murine infections. Both bulk and scRNA-seq analysis was employed to investigate gene expression changes from cells isolated from infection sites. In bulk RNA-Seq analysis several immune response markers were identified as being up-regulated after infection, including Regulatory Factor X5 (Rfx5), CD8 antigen beta chain 1 (Cd8b1), and cluster of differentiation 4 (Cd4). Chemokines associated with infections were also found: Chemokine (C-C motif) ligand 5 (Ccl5) and Chemokine (C-X-C motif) ligand 9 (Cxcl9). Down-regulated transcripts featured mostly ribosomal pathways. scRNA-seq analysis revealed a large variation in interferon-induced GTPases between samples, and antigen presentation molecules discovered in neutrophils, inflammatory monocytes, and monocyte-derived macrophages recruited to isolated lesions. Previous examples of the heterogeneity of cells involved in the host immune response to *Leishmania* have also been described, as reviewed by Sacks & Noben-Trauth (2002) and Scott & Novais (2016). Such heterogeneity is something scRNA-seq is well positioned to deconvolute. Interestingly, Ingenuity Pathway Analysis (IPA) employed by Venugopal et al., (2022) in scRNA-seq samples found a downregulation of Ribosomal protein S6 kinase beta-1 (eIF4/p70S6k), Eukaryotic translation initiation factor 4, mammalian target of rapamycin (mTOR) pathways, and Eukaryotic Initiation Factor 2 (EIF2) signalling, in various cell types, including macrophages. Despite the extensive murine cell dataset collated by Venugopal et al., (2022) no clearly detectable *L. major* cells were described in their samples. Some limited transcripts were mapped to the *L. major* genome

originating from within multiple murine cell types, but no parasite cell population for cluster or marker analysis was recovered.

In the current study, human macrophages derived from *in vitro* differentiation of THP-1 cells were used as a model for infection by *L. mexicana*. As the macrophage transcriptomic response is thought to be a transient expression limited to the first 24 hours post infection in bulk RNA-Seq dual transcriptomes (Fernandes et al., 2016), time points for infection were selected at 4 hours and 24 hours post infection, aiming to capture this transitory response. Additionally, early time points for infection would most likely provide as close to a 1:1 ratio of host cell to parasite as possible, not only to ensure all macrophages were infected, but also to increase the chances of recovering both host and parasite cell transcriptomes from the data at a single-cell resolution. The experimental protocol is defined below.

Here, we used scRNAseq to explore host and parasitic transcriptomic responses to infection a macrophage model was employed to culture cells infected with *L. mexicana* so that samples may be taken for scRNA-seq and analysed for differential expression of markers between uninfected controls and infected macrophages. Additionally, any captured *L. mexicana* cells may reveal transcriptomic responses closer to changes in gene expression responses associated with host cell phagocytotic infections.

5.2 Aims

- To use single cell transcriptomic data to identify markers associated with human macrophage infections with *L. mexicana*.
- To identify genes differentially expressed in *L. mexicana* during macrophage infections.
- To combine all experimental single-cell experiments to generate a foundational cell atlas of life cycle and cell cycle progression of *L. mexicana*.

5.3 Results

5.3.1 Single cell RNA-sequencing sample culture and preparation for macrophage infections

To investigate the host response during *L. mexicana* infection and compare amastigote differentiation inside the host cell with amastigotes grown *in vitro*, samples of macrophages were first induced from THP-1 cells. Macrophage samples were then separated into a non-infected control samples (CTL) and an infected sample with the addition of stationary phase grown *L. mexicana* (INF) before collection for scRNA-seq samples. Macrophages were grown at a density of 3×10^5 cell per mL, and in the INF sample, infected at a ratio of 1:10 with *L. mexicana* for 4 h before washing to remove any *L. mexicana* cells that had not infected the macrophages. The ratio of 1:10 macrophage to *L. mexicana* cells, derived from tests and previously used by Fernandes et al., (2016) was chosen to aim for a 1:1 ratio of human to *L. mexicana* cells for scRNA-seq analysis between the 4 h and 24 h in INF timepoints. To assess infection rates in the INF sample, *L. mexicana* were stained with the cytosol dye carboxyfluorescein diacetate succinimidyl ester (CFSE) to help visualise *L. mexicana* infecting macrophages via microscopy. The results of infection assays done in triplicate can be found below in Figure 5-1 A and B, with sample data for scRNA-seq analysis.

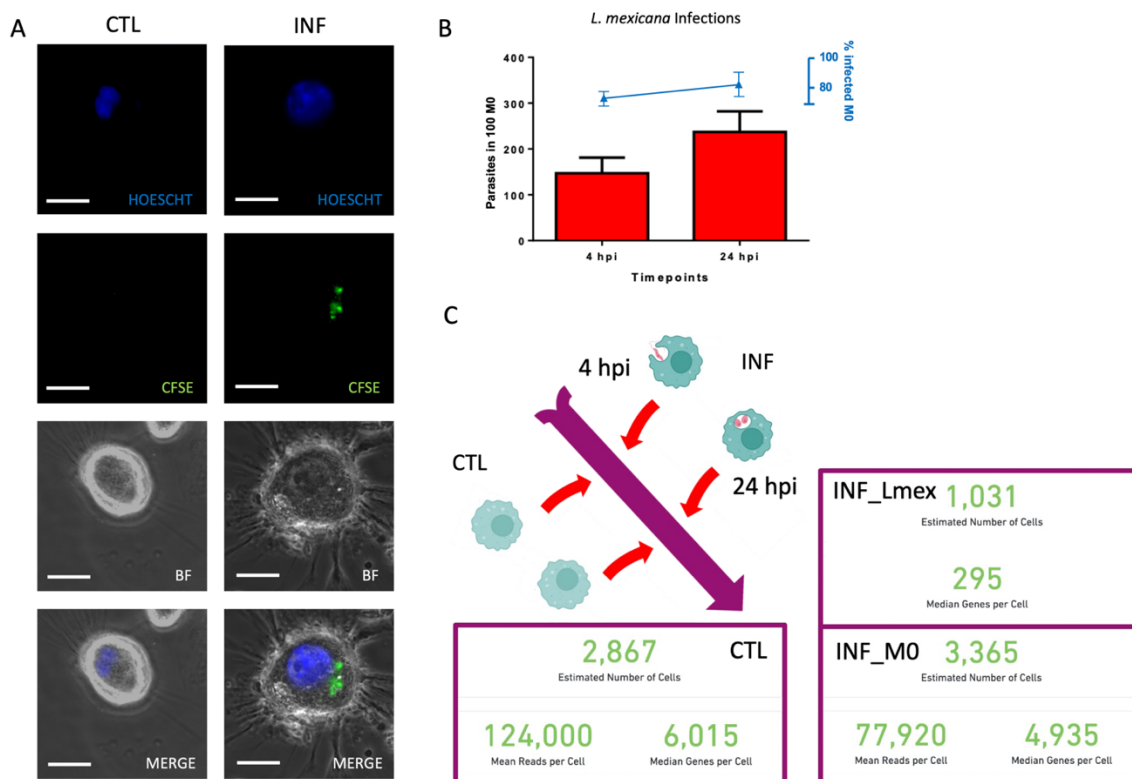


Figure 5-1 Experimental sample collection and sample information single cell RNA-sequencing experiment using macrophages infected with *L. mexicana*.

Experimental set-up for *Leishmania* sample run in single cell RNA-Sequencing containing 2 timepoints across the first 24 hours of macrophage infections. **A)** *L. mexicana* were stained with carboxyfluorescein diacetate succinimidyl ester (CFSE, coloured in green) in late log-phase promastigote stages before infecting macrophages at a 10:1 ratio. Samples were taken at 4 and 24 timepoints and macrophages stained with Hoescht (coloured in blue) before being imaged at 60x magnification. Pictured is uninfected control (CTL) and infected (INF) macrophages taken from 24 hours post infection (hpi). Scale bars are 10 μ m. **B)** The number of internalised *L. mexicana* was determined via microscopy (red bar plots), with the number of parasites counted in 100 macrophages and calculated as a percentage of macrophages infected (blue line). Counts run in triplicate for both timepoints. **C)** A ratio of 10:1 *L. mexicana* to macrophages was used for infections (aiming for a 1:1 ratio after infection) and samples from CTL and INF at 4 h and 24 h collected. In total, 2,867 cells were sequenced for the CTL sample with mean reads per cell of 124, 000 and median genes per cell of 6,015. For the INF sample containing macrophages infected with *L. mexicana*, 3,365 cells were sequenced with 77,920 mean reads per cell and 4,935 median genes per cell.

The combined samples of uninfected macrophages and macrophages infected with *L. mexicana* at 4 h and 24 h timepoints contained an estimated 6,232 human cells in total. The uninfected macrophage control (CTL) containing 2,867 cells produced mean reads per cell of 124,000. The infected human cells (INF) also containing *L. mexicana* within the macrophages had 3,365 human cells in total and produced mean reads per cell of 77,920. Analysis for the macrophage samples is shown below in Section 5.3.2. The *L. mexicana* cells counts in the sample were estimated at 1,031, and median genes per cell of 295. And individual *L. mexicana* analysis free from macrophage transcripts were analysed below in Section 5.3.3.

5.3.2 Single cell RNA-sequencing analysis of infected macrophages

Mapping of reads generated for the macrophage samples, as described in Chapter 3, were assessed by Cell Ranger outputs, generated by Pawel Herzyk of Glasgow Polyomics. Subsequently, quality control of the samples is applied to limit the impact of noise generated due to biological and technological variations (Kolodziejczyk et al., 2015).

5.3.2.1 Quality control filtering of single cell RNA-sequencing for combined uninfected and infected macrophage samples

Sequencing data was first mapped to the human reference genome GRCh38, combined with the custom *L. mexicana* reference genome as in Chapter 3, by Dr. Pawel Herzyk of Glasgow Polyomics, the output of which is shown below in Appendix I.

Shown in Figure 3-3 below are the quality control filters considered, which were number of unique features, UMIs, and mitochondrial DNA (MT) to exclude cells with low reads, as performed for previous samples.

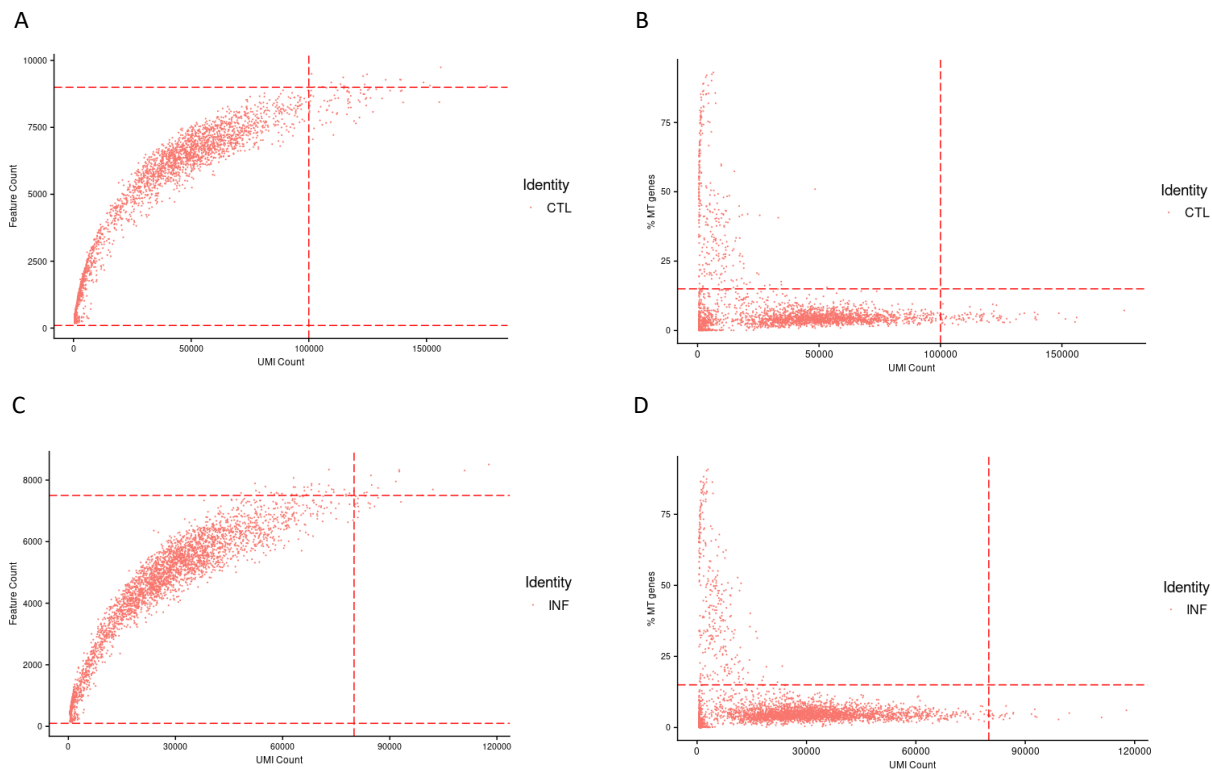


Figure 5-2 Quality control and filtering of transcriptomes in macrophage samples.

Scatter plots of quality control and filtering cut-offs. Each dot representing one captured transcriptome. Unique Molecular Identifiers (UMI) are plotted against **A**) gene counts (features) with UMI counts for the control sample (CTL) set at <100,000, feature counts >200 and <9000 to exclude for multipllets. **B**) Percentage of mitochondrial reads (MT) included with CTL mapping reads < 15% to exclude lysed or apoptotic cells. **C**) Gene counts (features) with UMI counts for the infected sample (INF) set at <80,000, feature counts >200 and <7500 to exclude for multipllets. **D**) Percentage of mitochondrial reads included with INF sample mapping reads < 15% to exclude lysed or apoptotic cells.

In Figure 3-3 **A** and **B** the quality control cut-offs for the uninfected sample were chosen as being 100,000 as a upper limit for UMI counts, with feature count exclusions set at >200 and <9000, with a mitochondrial cut-off of 15%, representative of the typical mitochondrial DNA expected in macrophages (Osorio and Cai, 2021) and also used for the infected sample **D**. For the infected sample, cut-offs were selected for UMI counts above 80,000 **C**, feature counts >200 and <7,500, which were lower than the control parameters, possibly indicating cells with reduced viability due to infection with *L. mexicana*. Quality control cut-offs are increased here to account for the greater range and number of gene expression in human macrophage cells compared to *L. mexicana*. Quality control cut-offs for the *L. mexicana* only sample derived from the infected macrophages represent a closer range to those previously discussed above for the the in vitro only *L. mexicana* samples.

Variable genes were then selected, as described previously, with the top ten for each plotted by scatter plot below in Figure 3-4.

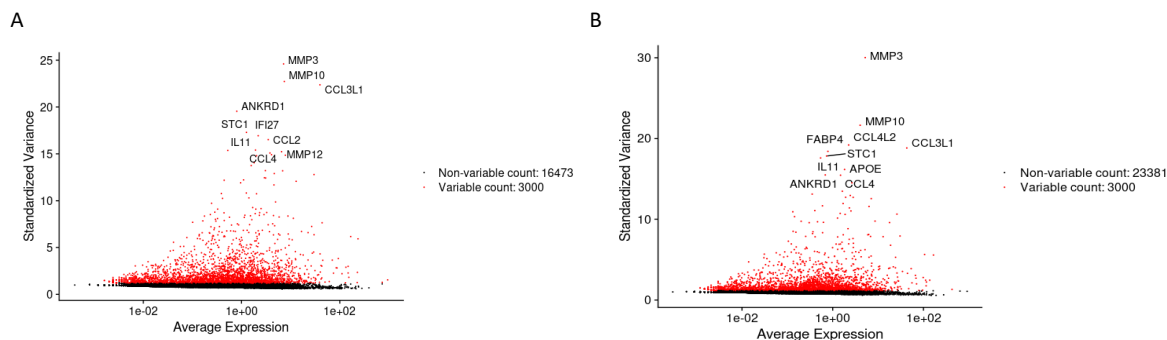


Figure 5-3 Scatter plot of top 3,000 variable features in macrophage datasets.

Scatter plot of top 3,000 variable features (red) selected for downstream analysis, in the control sample **A**) and infected sample **B**), plotting average expression of features against standard variance. Top ten variable features for each labelled.

Overlap between the two samples were found in seven of the top ten most variable genes: *MMP3*, *MMP10*, *CCL3L1*, *ANKRD1*, *STC1*, *IL11* and *CCL4*. Three genes were only seen the top ten of the CTL sample: *IFI27*, *MMP12* and *CCL2*. *IFI27* is associated with RNA polymerase II-specific DNA-binding transcription factor binding activity and innate immune system interferon gamma signalling (Gao et al., 2021), *MMP12* is known to be involved in the breakdown of extracellular matrix (Gao et al., 2021), and *CCL2* is a cytokine involved in immunoregulatory and inflammatory processes (Gschwandtner et al., 2019). The three genes only seen in the INF top ten were *CCL4L2*, *FABP4* and *APOE*. *CCL4L2* is a chemokine associated with cell entry (Artigas-Jerónimo et al., 2021), *FABP4* encodes for a fatty acid binding protein required for neutrophil recruitment (X. Liang et al., 2019), and *APOE* has been shown to prompt anti-inflammatory responses in macrophages (Baitsch et al., 2011).

As in previous samples, linear dimension reduction was performed by PCA on the integrated CTL and INF samples with scaling on the top 3,000 features, shown in Figure 3-5.

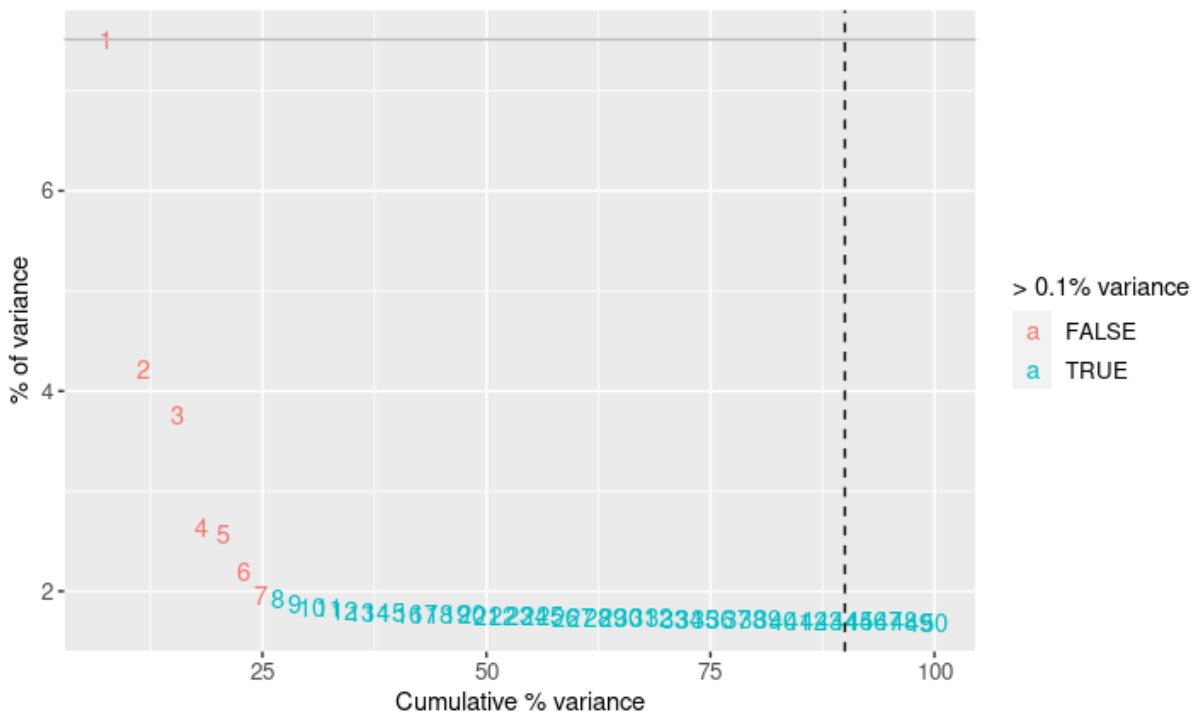


Figure 5-4 Elbow plot of variance found in principle components observed in integrated macrophage samples.

Elbow plot to rank principle components (PC) found in integrated uninfected and infected macrophage samples assessing the percentage of variance explained by each PC. In the integrated samples, an “elbow” is seen around PC 7, suggesting that the majority of variance is captured in the first 7 PCs.

In Figure 3-5 an ‘elbow’ was present around PC 7 -8, which suggested most of the significant variance was captured in the first 7 PCs. 7 PCs was therefore chosen for the number of dimensions in the dataset for further downstream analysis. Resolution of the clustering was determined by Clustree as described previously, and shown below in Figure 3-6

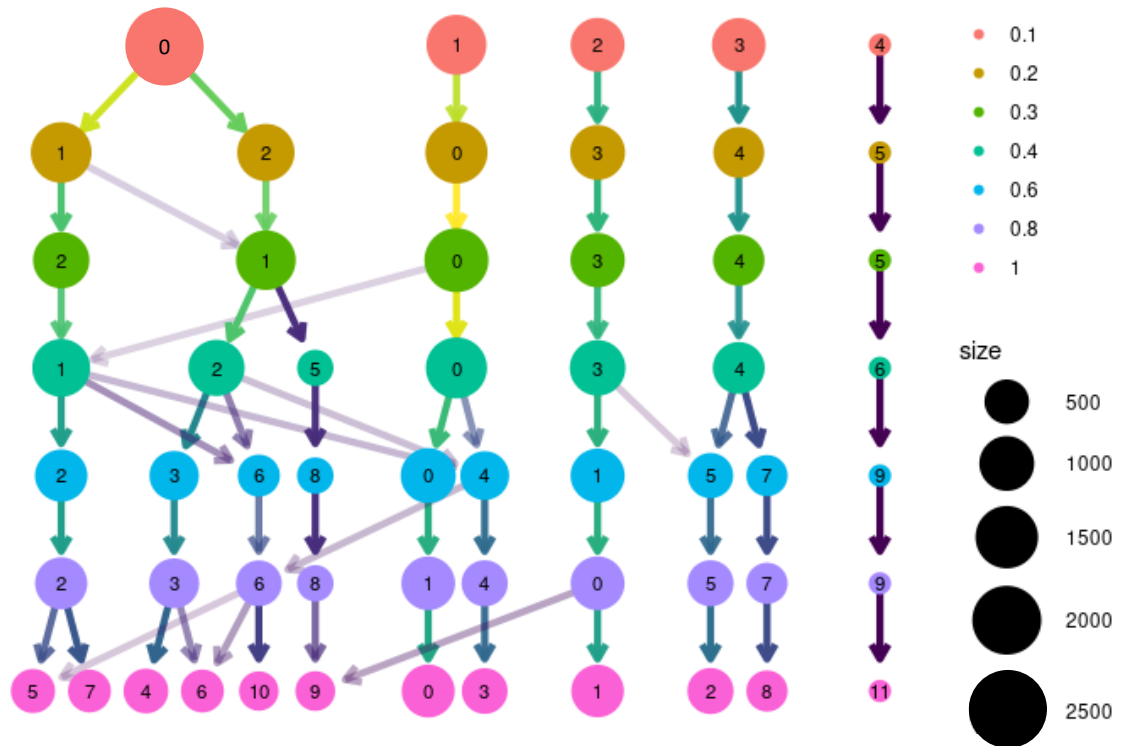


Figure 5-5 Clustree plot demonstrating integrated macrophage sample movement with increasing cluster resolution.

Clustree plot for integrated macrophage samples showing changes in sample movement as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells partitioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.2 (coloured in ochre) is chosen resolving into 6 clusters for further clustering analysis.

A resolution of 0.2 (coloured in ochre) partitioned the dataset into 6 distinct clusters. A resolution of 0.2 was selected for further clustering analysis, explored below in Section 3.3.2.2

5.3.2.2 Clustering analysis of single cell RNA-sequencing for combine macrophage and *L. mexicana* samples.

Clustering was visualised in a UMAP, as for previous samples, in Figure 3-7.

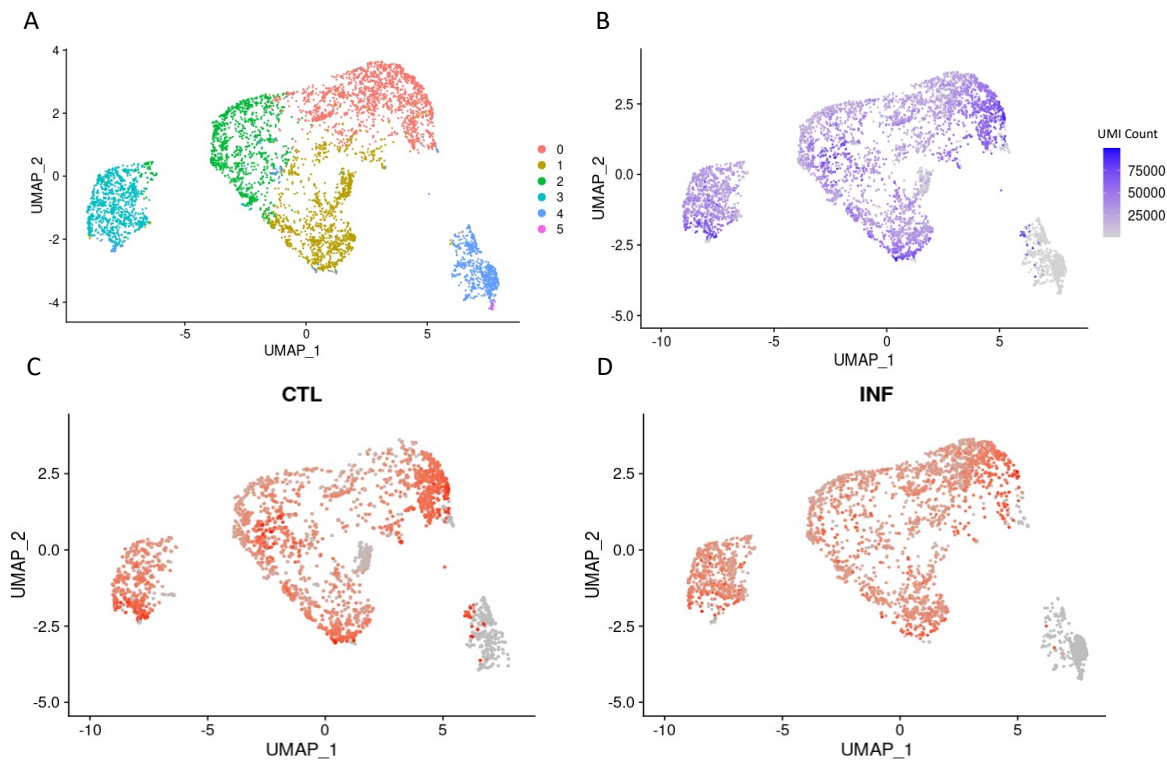


Figure 5-6 Clustering of human macrophage and *L. mexicana* transcriptomes across two timepoints post infection.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in integrated samples to visualise relative relationships between individual transcriptomes. **A)** UMAP of integrated samples at a resolution of 0.2 reveals 6 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** UMAP of uninfected control (CTL) only to show cell overlap between samples, with total coloured in red. **D)** UMAP of infected sample (INF) only to show cell overlap between samples, with transcripts coloured in red, showing cluster 5 (coloured in pink) only present in INF sample and with low expression compared to the rest of the cells in sample.

Clustering revealed large overlaps between CTL and INF samples, indicating a majority of cell types found were present in both uninfected and infected samples. Cluster 4 and 5 (coloured in blue and pink) showed low total RNA when compared to the other clusters, with cluster 5 being the smallest cluster and only represented in the INF sample shown in Figure 3-7 D.

5.3.2.3 Marker analysis of integrated macrophage samples

To further associate captured transcripts with infection by *L. mexicana*, marker analysis was undertaken as in previous samples, and shown in Figure 5-7.

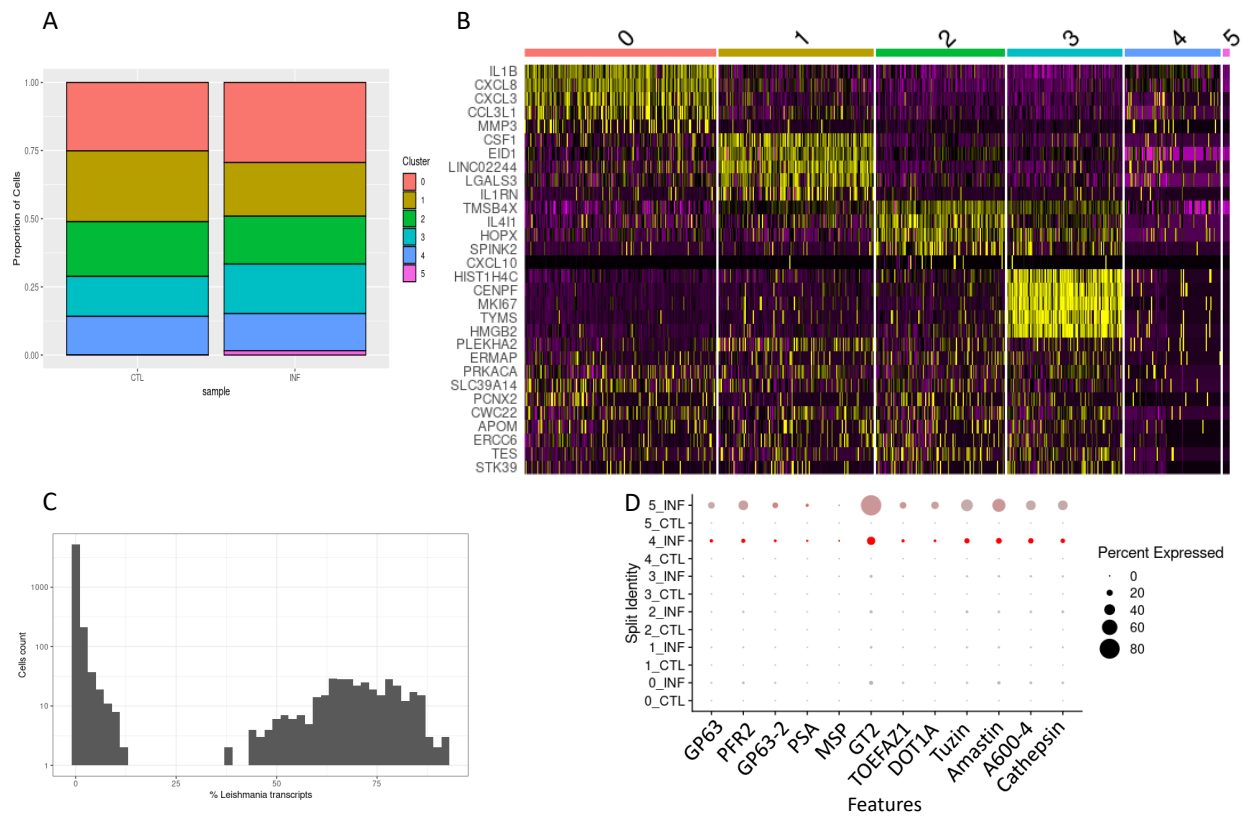


Figure 5-7 Clustering and top differentiated marker analysis by heatmap with dot plot of *L. mexicana* transcripts in integrated macrophage samples.

A) Proportional groupings of cells in each cluster and coloured by cluster, as in Figure 3-7. **B)** Heatmap of top five distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster. **C)** Percentage of *L. mexicana* transcripts in cells across combined macrophage samples. **D)** Dot plot of marker genes for life cycle specific stages, as used in chapter 3. ordered from left to right for promastigote to amastigote stages, with clusters separated by control (CTL) and infected macrophage samples (INF). *L. mexicana* transcripts can be seen as enriched in clusters 4 and 5 and only in INF cells. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot.

The proportion of cells in each cluster was calculated and plotted in Figure 5-7 A), showing that cluster 5 is indeed only present in the INF sample, representing 2.6% of total cells present. Other clusters were closely matched between the two samples, and further indicating clustering is unlikely to be based on infection response. However, cluster 0 (coloured in red) and cluster 3 (coloured in turquoise) were larger in the INF sample by 4.8% and 3.7%, respectively. Expression of the top 5 markers for each cluster were plotted in B with notably reduced expression of marker genes in clusters 4 and 5, which lack high expression of any specific markers compared to other clusters. Thus, these are largely defined on a lack of expression of markers associated with other clusters. To assess the number of *L. mexicana* transcripts in the combined samples, proportions of *L. mexicana* transcripts were calculated against cell counts in the sample shown in C. *L. mexicana* transcripts were not seen to be present in a

majority of cells; however, 14.1% of cells did show *L. mexicana* transcripts present in the combined samples, suggesting 14.1% of macrophages captured are infected with *L. mexicana*. Plotting previously used life cycle and cell cycle stage markers for *L. mexicana* samples in **D** revealed *L. mexicana* transcripts are present in clusters 4 and 5 and only in INF samples cells, with 79% of the GT2 marker for metacyclics present in cluster 5 in INF cells. Cluster 4 and 5 are defined by low expression of other cluster markers and presence of *L. mexicana*.

To further illustrate these data, feature plots and violin plots of top differentially expressed markers for each genome were also produced to give expression probability distributions across clusters in Figure 5-8.

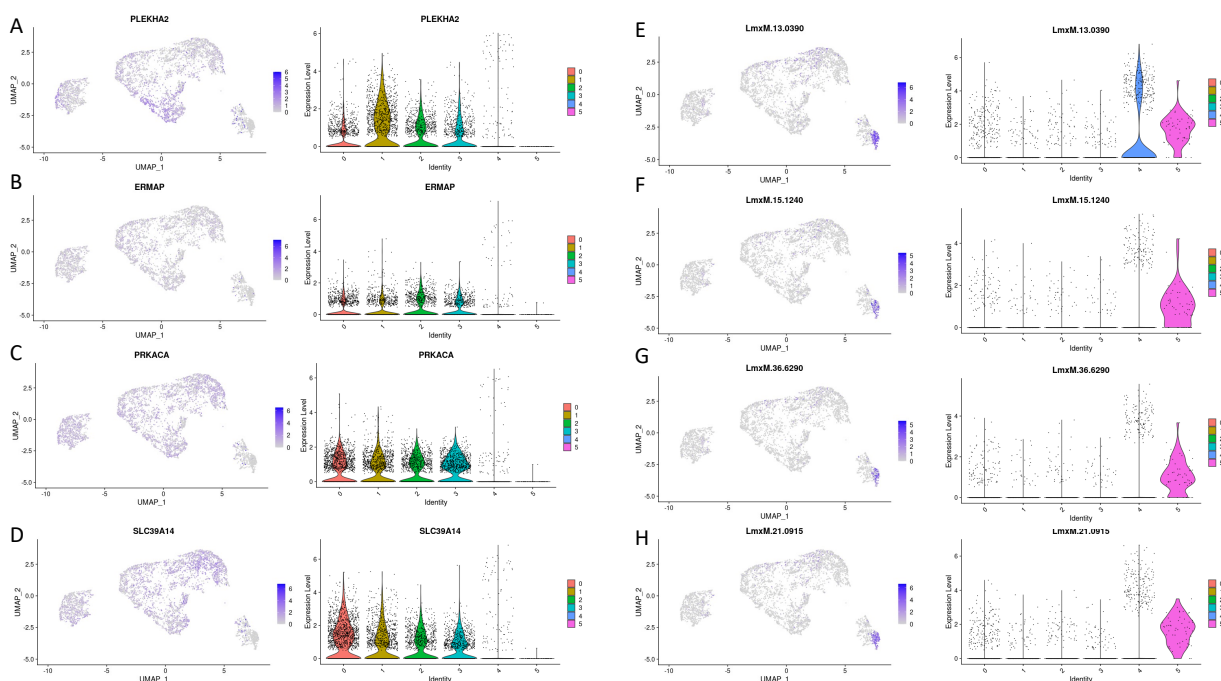


Figure 5-8 UMAP feature plots and violin plots in integrated macrophage scRNA-seq samples for top 4 human and *L. mexicana* markers.

Uniform Manifold Approximation and Projection (UMAP) of integrated macrophage samples, with infecting *L. mexicana*, and their respective violin plots for expression across clusters coloured by transcript counts per cell. The top 4 markers for macrophage and *L. mexicana* cells are shown here. Markers **A** - **D**) (*PLEKHA2*, *ERMAP*, *PRKACA*, *SLC39A14*) are used as markers for human macrophage cells, with **A** – **B**) being the top 2 markers for uninfected control cells, and **C** – **D**) the top 2 markers for infected macrophage cells. Markers **E** – **H**) (alpha tubulin; LmxM.13.0390, glucose transporter 2; LmxM.15.1240, histone H2A; LmxM.36.6290, nucleoside transporter 1; LmxM.21.0915) being the top 4 markers for *L. mexicana* cells found in the infected macrophage samples. **E** – **F**) represent the top 2 markers for the metacyclic like cells and markers **G** – **H**) represent the top 2 markers for the amastigote cells.

Top differentially expressed markers for the uninfected control sample in **A** and **B** showed an even spread of expression across cluster 0 - 3, with increased expression in cluster 1 (coloured in ochre), also seen in markers **C** and **D** for the top differentially expressed markers from the infected macrophage samples,

where some increased expression was seen in cluster 0 (coloured in red). This further indicated an overall similarity in gene expression between clusters, meaning that differences in expression caused by *L. mexicana* infection may be limited to a minority of genes rather than a global change in gene expression. Top markers for *L. mexicana* infected cells, shown in E - H, showed increased expression in clusters 4 and 5 (coloured in blue and pink, respectively) with marker E, alpha tubulin, being selected to indicate possible enrichment for cells mid transition between metacyclic stages and amastigote stages due to being expressed more in promastigote stages (Ramírez et al., 2013), and thus likely indicating metacyclic cells captured at the 4 h timepoint sample. Higher expression for *L. mexicana* transcripts was seen in cluster 5, which indicated a population of *L. mexicana* cells captured in the infected sample for further analysis.

5.3.2.4 Marker analysis of uninfected control macrophage sample

To investigate differences in gene expression between samples, each experiment was analysed individually to identify clusters present in each and their related markers for comparison between conditions. First, the uninfected control sample was analysed to generate markers for comparison to the infected sample. Dimensions and resolution for the control sample shown are in Figure 3-10.

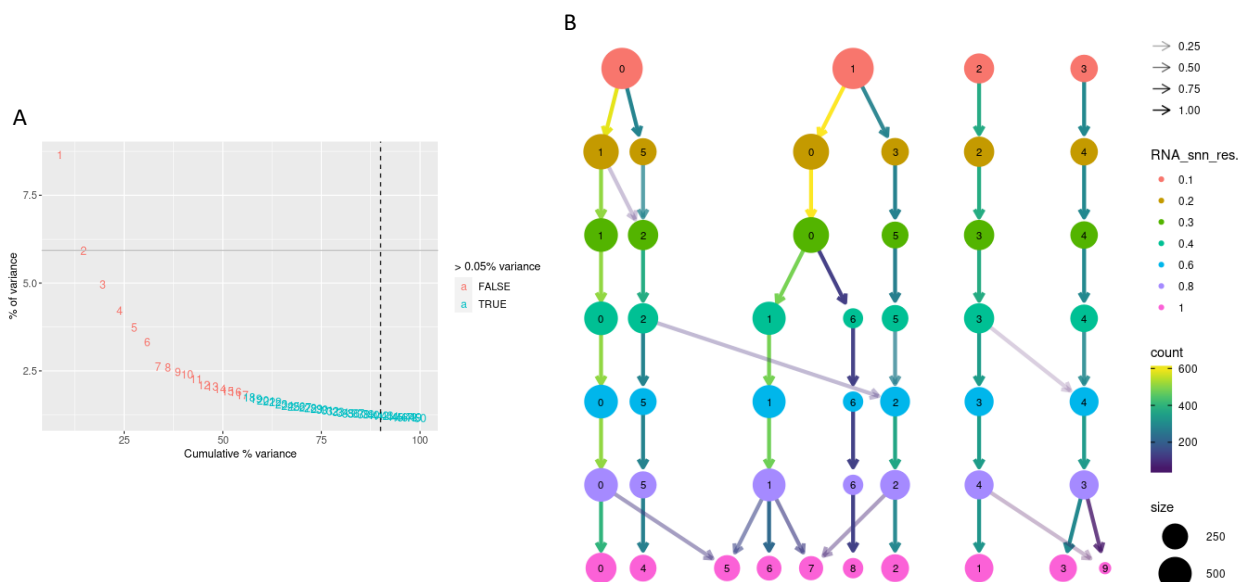


Figure 5-9 Elbow plot of variance found in principle components observed in uninfected control macrophage sample and Clustree plot demonstrating movement with increasing cluster resolution.

A) Elbow plot to rank principle components (PC) found in uninfected control sample assessing the percentage of variance explained by each PC. In the control sample, an “elbow” is seen around PC 17, suggesting that the majority of variance is captured in the first 17 PCs. **B)** Clustree plot for integrated macrophage samples showing changes in sample movement as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells portioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.1 (coloured in red) is chosen resolving into 4 clusters for further clustering analysis.

17 PCs were selected and a resolution of 0.1 selected to give 4 clusters across two timepoints shown in Figure 5-10.

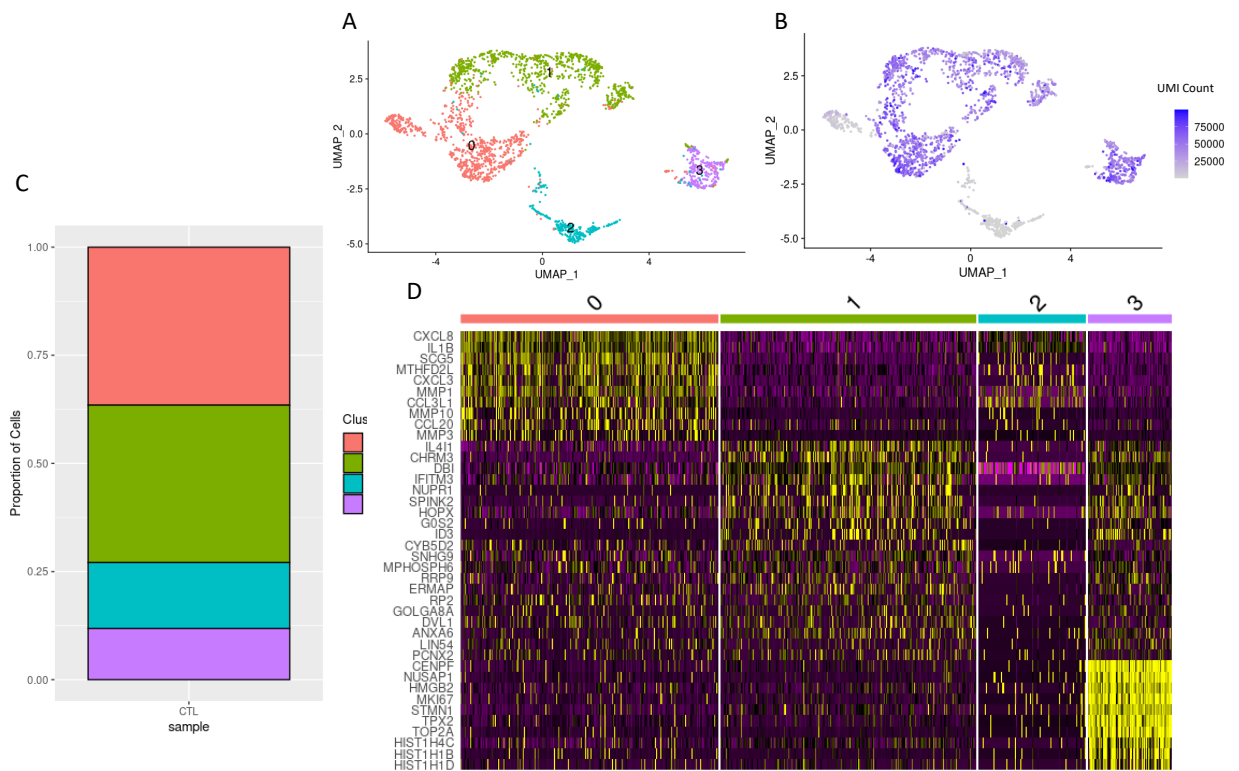


Figure 5-10 Clustering and top differentiated marker analysis by heatmap in uninfected control macrophage sample.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in uninfected control (CTL) to visualise relative relationships between individual transcriptomes. **A)** UMAP of CTL sample at a resolution of 0.1 reveals 4 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster.

In total, 10,056 human cluster markers were found in the uninfected control sample for comparison to the infected sample.

5.3.2.5 Marker analysis of infected macrophage sample

The infected macrophage sample, containing *L. mexicana* cells, was analysed to generate markers for comparison to the infected sample. Dimensions and resolution for the control sample are shown in Figure 3-10.

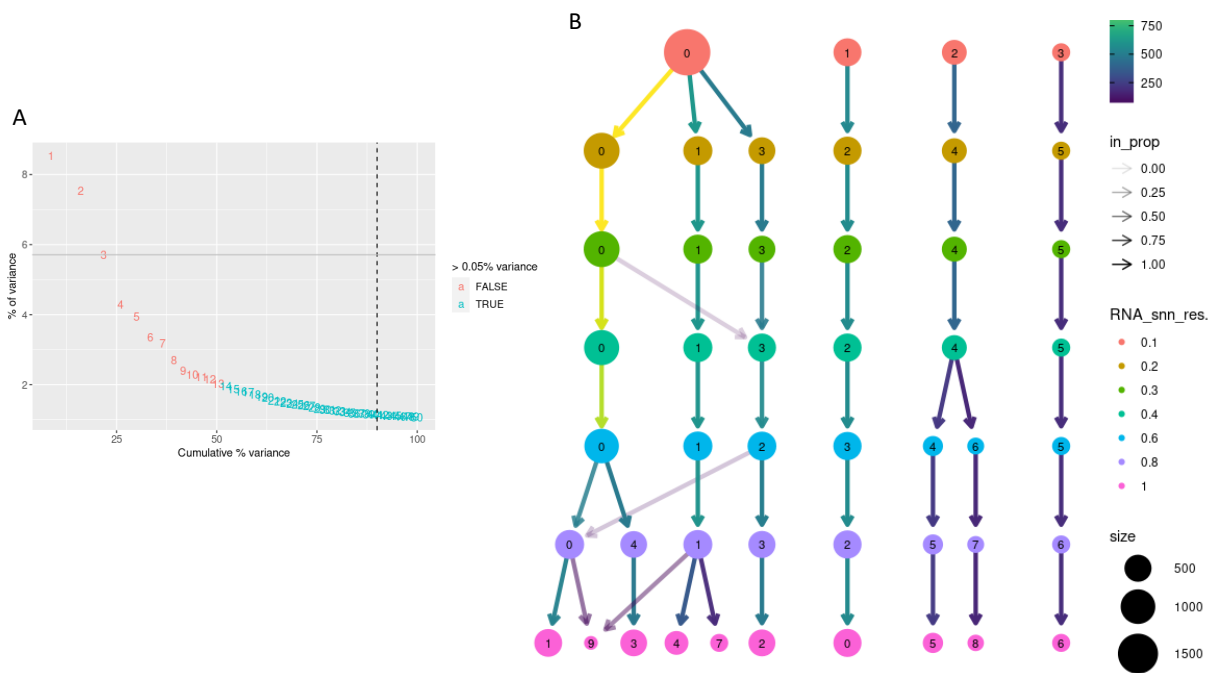


Figure 5-11 Elbow plot of variance found in principle components observed in infected macrophage sample and Clustree plot demonstrating movement with increasing cluster resolution.

A) Elbow plot to rank principle components (PC) found in uninfected control sample assessing the percentage of variance explained by each PC. In the infected sample, an “elbow” is seen around PC 13, suggesting that the majority of variance is captured in the first 13 PCs. **B)** Clustree plot for infected macrophage sample showing changes in sample movement as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells portioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.1 (coloured in red) is chosen resolving into 4 clusters for further clustering analysis.

13 PCs were selected and a resolution of 0.1 selected to give 4 clusters across two timepoints, as shown in Figure 5-12.

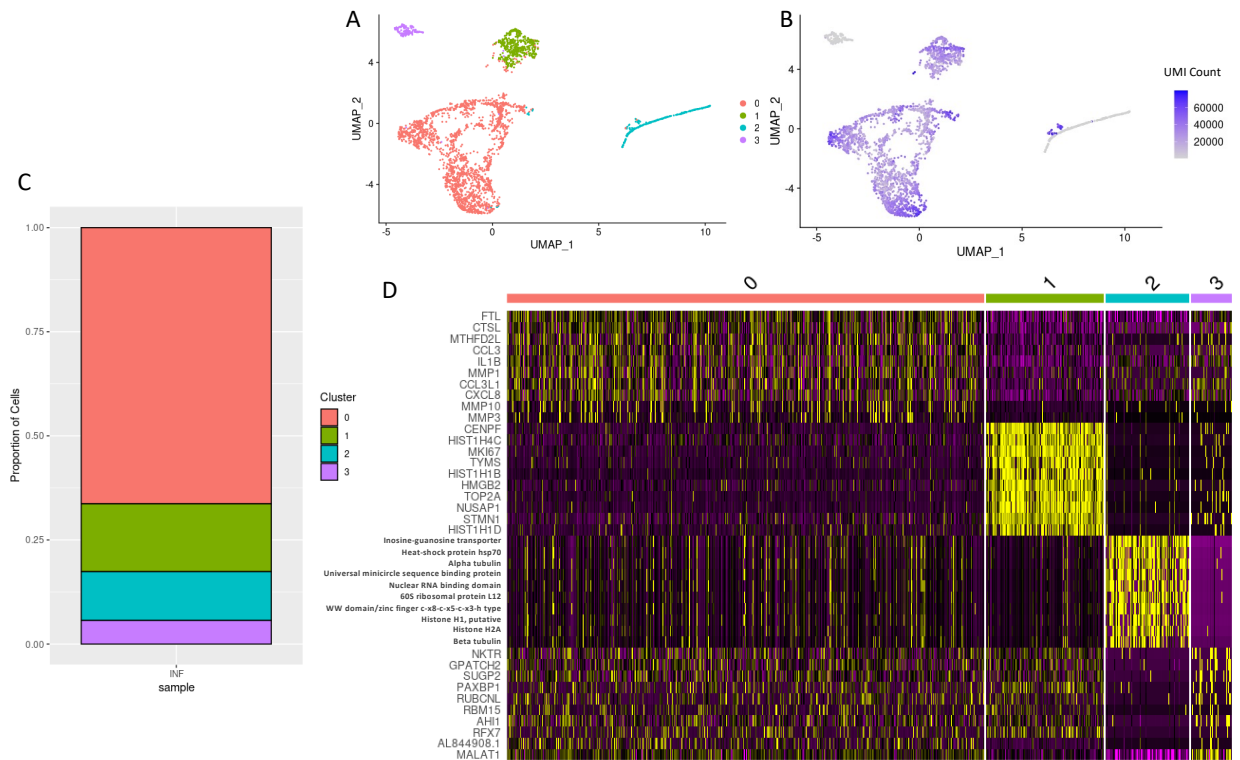


Figure 5-12 Clustering and top differentiated marker analysis by heatmap in infected macrophage sample with *L. mexicana*.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in infected macrophage and *L. mexicana* sample (INF) to visualise relative relationships between individual transcriptomes. **A)** UMAP of INF sample at a resolution of 0.1 reveals 4 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster.

In total, 9,405 human cluster markers were found in the infected macrophage sample for comparison to the uninfected control sample. Also found within the analysis of the isolated infected sample was a cluster comprising *L. mexicana* cells (cluster 2, turquoise, Figure 5-12), which represented 12.2% of the cells in this sample. Marker analysis for cluster 2 produced 193 *L. mexicana* genes for comparison to axenic amastigote markers (analysed previously, Chapter 3).

5.3.2.6 Marker comparison analysis between uninfected control and infected macrophage samples

The markers derived from individual sample analysis were next used to compare expression profiles between the uninfected and infected macrophage samples; Figure 5-13.

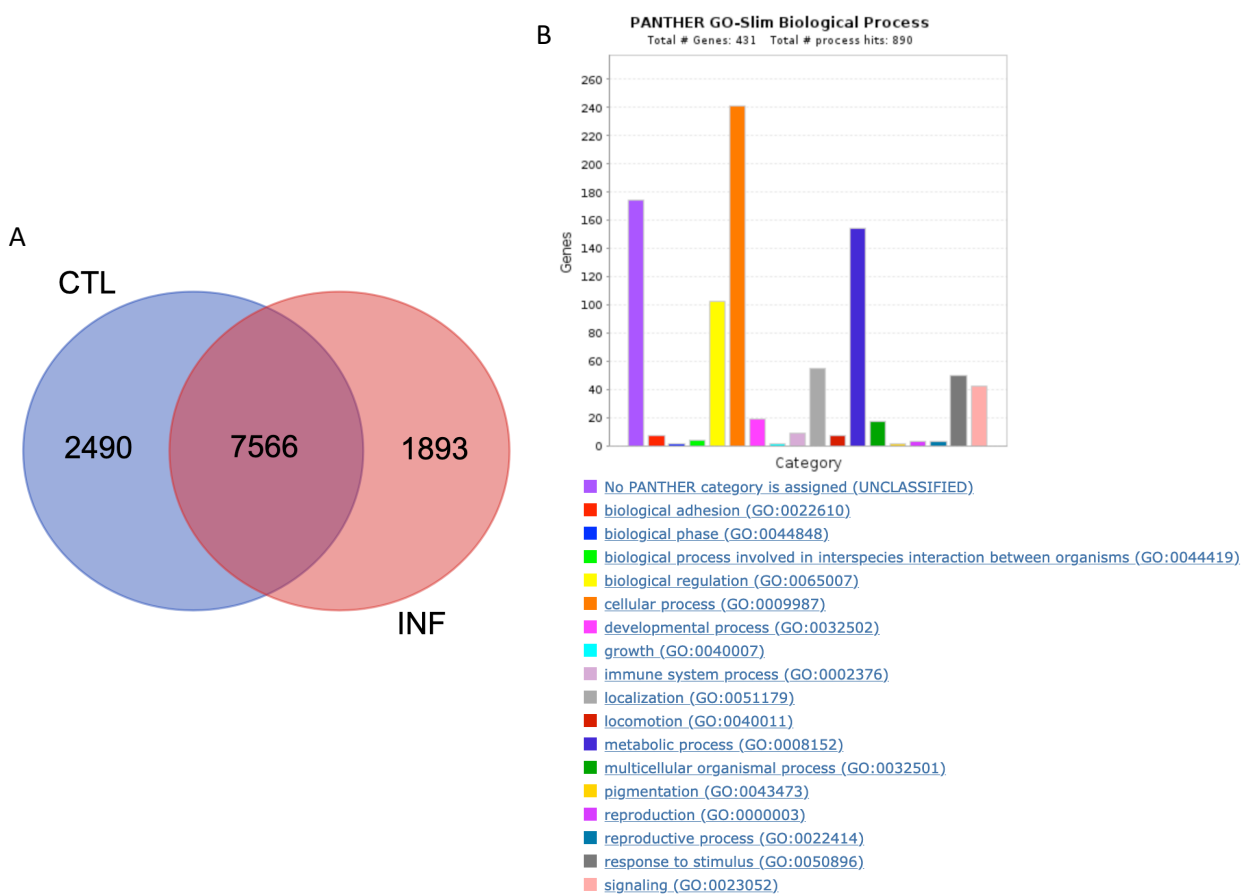


Figure 5-13 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage sample.

A) Venn diagram for uninfected macrophage control (CTL) and infected macrophage sample (INF) markers following individual analysis demonstrating overlap between samples. 379 markers were found to be unique to the INF sample. **B)** PANTHER gene ontology (GO) enrichment for biological processes found in 379 unique markers for infected macrophage sample, with 606 process hits. Notable GO terms include biological processes involved with interspecies interactions between organisms and immune system processes.

Comparing the 11,949 marker genes found between the two samples revealed an overlap of 7,566 genes, constituting 63.3% of the total differentially expressed markers between both samples. 1,893 genes (15.8% of the total genes between the two samples) were unique to the infected macrophage sample. Using PANTHER gene list analysis (Mi et al., 2013) for GO term enrichment revealed 18 GO terms. Interestingly, included in the GO terms was ‘biological process involved in interspecies interaction between organisms’ (GO:0044419), containing 4 markers and, ‘immune system process’ (GO:0002376), containing 9 markers of which 4 were also found in GO:0044419 (shown in Table 1-1).

GO Term	Mapped Gene ID	PANTHER Family/Subfamily	PANTHER Protein Class
Interspecies interaction between organisms & Immune system process	RNASE2	Non-secretory ribonuclease	Endoribonuclease
Interspecies interaction between organisms & Immune system process	SRC1	Proto-oncogene tyrosine-protein kinase Src	Non-receptor tyrosine protein kinase
Interspecies interaction between organisms & Immune system process	MYD88	Myeloid differentiation primary response protein MyD88	Scaffold/adaptor protein
Interspecies interaction between organisms & Immune system process	RELB	Transcription factor RelB	Rel homology transcription factor
Immune system process	CTSH	Pro-cathepsin H	Cysteine protease
Immune system process	HLA-B	HLA class I histocompatibility antigen, B alpha chain	Major histocompatibility complex protein
Immune system process	CD99	CD99 antigen	—
Immune system process	CXCL16	C-X-C motif chemokine 16	Chemokine
Immune system process	FCGRT	IgG receptor FcRn large subunit p51	Major histocompatibility complex protein

Table 5-1 Infected macrophage markers with gene ontologies of interest for *L. mexicana* infections.

A) Venn diagram for uninfected macrophage control (CTL) and infected macrophage sample (INF) markers following individual analysis demonstrating overlap between samples. 379 markers were found to be unique to the INF sample. **B)** PANTHER gene ontology (GO) enrichment for biological processes found in 379 unique markers for infected macrophage sample, with 606 process hits. Notable GO terms include biological processes involved with interspecies interactions between organisms and immune system processes.

Additionally, a comparison of infected human cells using bulk RNA-sequencing markers was made using data generated by Fernandes et al., (2016), with macrophage transcripts from *L. major* infections at the 4 h timepoint. Human markers were compared to controls and only genes unique to infected samples used to compare to the scRNA-seq markers generated here as unique to the infected macrophage sample; Figure 5-14.

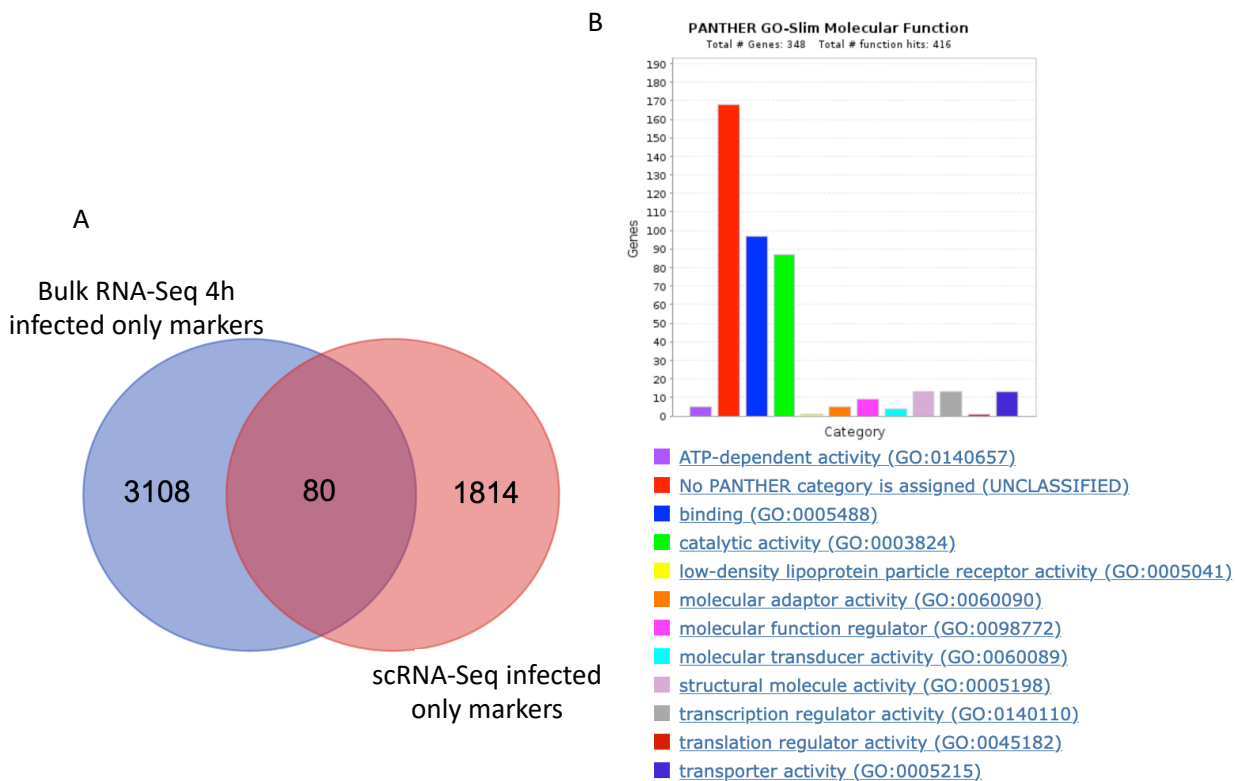


Figure 5-14 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage samples by scRNA-seq.

A) Venn diagram for infected macrophage markers unique to macrophages infected with *L. major* at 4 hours post infection (Bulk RNA-Seq 4h infected only markers by Fernandes et al., 2016) and infected macrophage sample from scRNA-seq identified here (scRNA-seq infected only markers) using markers only identified in infected macrophages 1814 markers identified unique to scRNA-seq analysis with GO terms identified in **B**). **B)** PANTHER gene ontology (GO) enrichment for biological processes found in 1814 unique markers for infected macrophage sample, with 705 process hits.

In total, 1814 markers were found unique to infected macrophages generated here and analysed by scRNA-seq. Interestingly, the overlap found was small, perhaps due to the variation in macrophage sources; in this study THP-1 cell lines were used to differentiate into macrophages, while in Fernandes et al., (2016) macrophages were derived from CD14⁺ monocytes. Amongst these scRNA-seq markers were the GO terms biological process involved in interspecies interaction between and immune system process, as previously identified, Table 1-1, were again detected.

5.3.3 Single cell RNA-sequencing analysis for *L. mexicana* within infected macrophage sample

To investigate markers associated with *L. mexicana* cells infecting macrophages individual cluster analysis was also undertaken for *L. mexicana* transcripts only mapped to the *L. mexicana* custom genome (as Chapter 3), performed by Dr. Pawel Herzyk.

5.3.3.1 Marker analysis of isolated *L. mexicana* sample from macrophage infections

Shown in Figure 5-15 below were the quality control filters considered, as for previous *L. mexicana* samples.

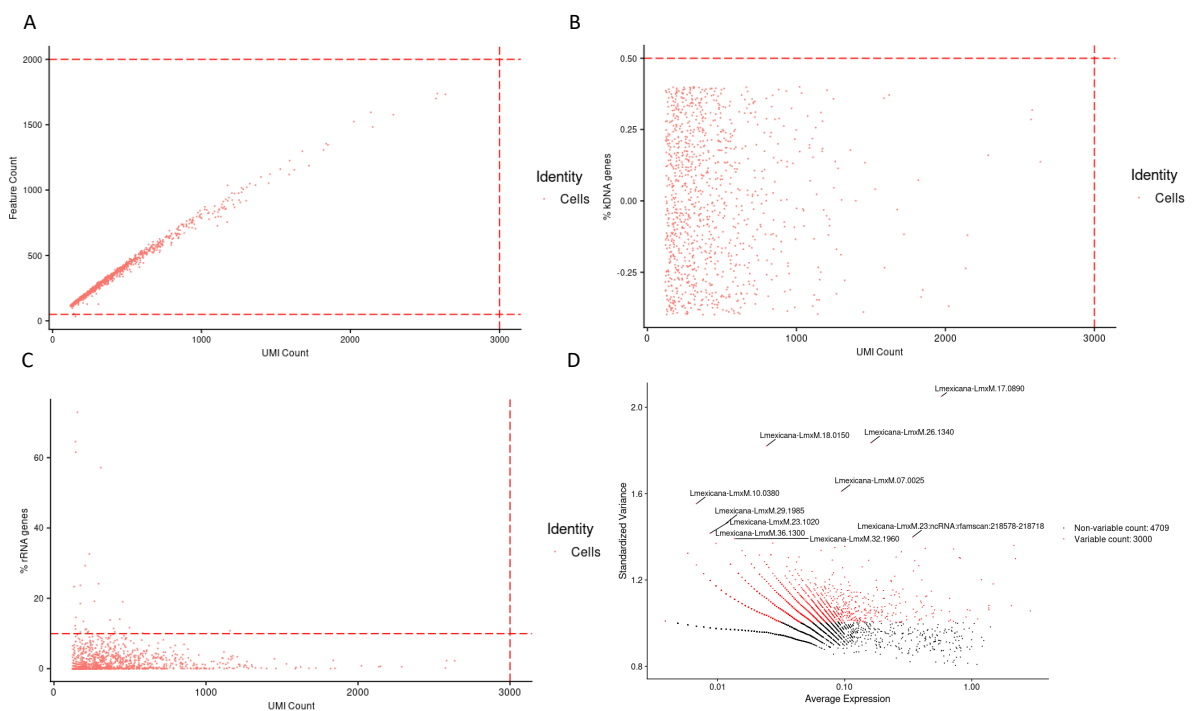


Figure 5-15 Quality control and filtering of transcriptomes in amastigote sample isolated from infected macrophages and top 3,000 variable markers.

Scatter plots of quality control and filtering cut-offs. Each dot representing one captured transcriptome. Unique Molecular Identifiers (UMI) are plotted against **A**) gene counts (features) with UMI counts <3000, feature counts >50 and <2000 to exclude for multiplets, **B**) percentage of kDNA maxi circle genome included with mapping reads < 0.5% to exclude lysed or apoptotic cells and **C**) percentage of ribosomal RNA (rRNA) <10%. Red dashed lines indicate cut-offs used to filter cells per experiment. **D**) Scatter plot of top 3,000 variable features (red) selected for downstream analysis in isolated amastigote cells, plotting average expression of feature against standard variance. Top ten variable features labelled.

Cells which cleared quality control cut-offs were selected, with their top 3,000 variable features used for PC analysis and resolution by Clustree in Figure 5-16.

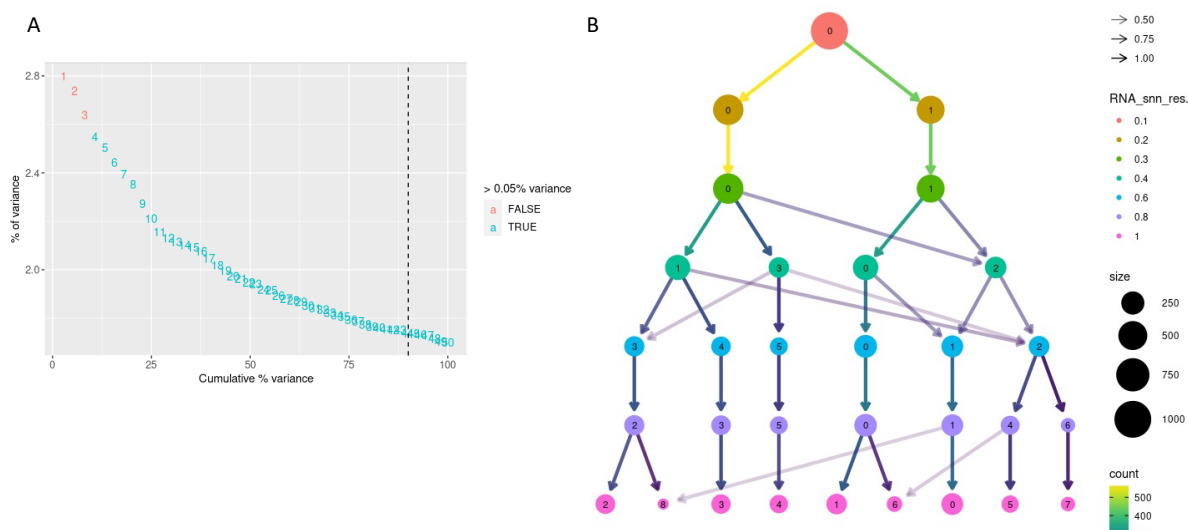


Figure 5-16 Elbow plot of variance found in principle components observed in *L. mexicana* amastigote sample and Clustree plot demonstrating movement with increasing cluster resolution.

A) Elbow plot to rank principle components (PC) found in *L. mexicana* amastigote sample assessing the percentage of variance explained by each PC. In the *L. mexicana* amastigote sample, an “elbow” is seen around PC 3, suggesting that the majority of variance is captured in the first 3 PCs. **B)** Clustree plot for *L. mexicana* amastigote sample showing changes in sample movement as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells portioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.2 (coloured in ochre) is chosen resolving into 2 clusters for further clustering analysis.

3 PCs were selected and a resolution of 0.2 selected to give 2 clusters, as shown in Figure 5-17.

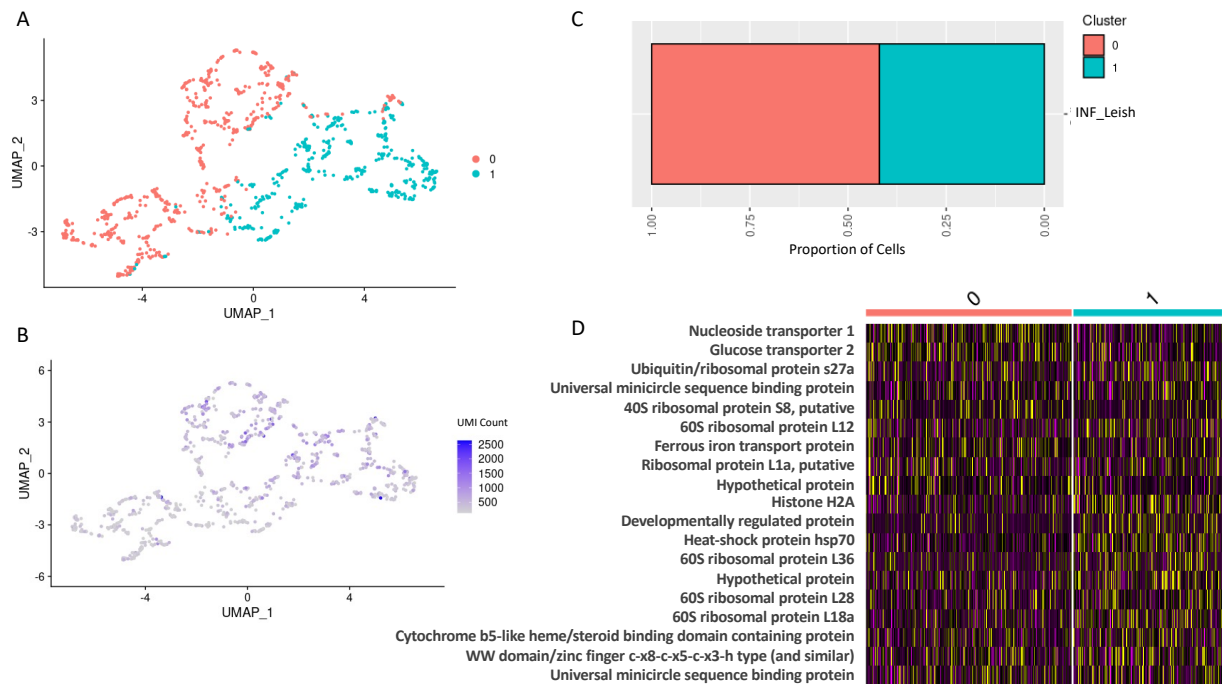


Figure 5-17 Clustering and top differentiated marker analysis by heatmap in *L. mexicana* amastigote sample.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in *L. mexicana* amastigote sample (Ama) to visualise relative relationships between individual transcriptomes. **A)** UMAP of Ama sample at a resolution of 0.2 reveals 2 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** Proportional grouping of cells in each cluster and coloured by cluster. **D)** Heatmap of top ten distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster.

In total, 90 *L. mexicana* gene markers were isolated from the infected macrophage sample. Based on the low number of markers and lack of distinct differences in the Heatmap, the *L. mexicana* cells captured seem mostly to be a uniform population. Of note were the top two markers for cluster 0 (NT1 and GT2) (coloured in red), previously seen in promastigote markers in *in vitro* experiments (See Figure 3-8 and Figure 3-10). This further indicated a possible promastigote-like cluster preceding differentiation into amastigote forms. Top markers for the amastigote-like cluster featured ribosomal proteins and a zinc finger gene previously seen in axenic amastigote clusters (see Figure 3-10 and Figure 3-29). Comparisons to amastigote markers previously described by bulk RNA-Sequencing by Fiebig et al., (2015) are shown in Figure 5-18.

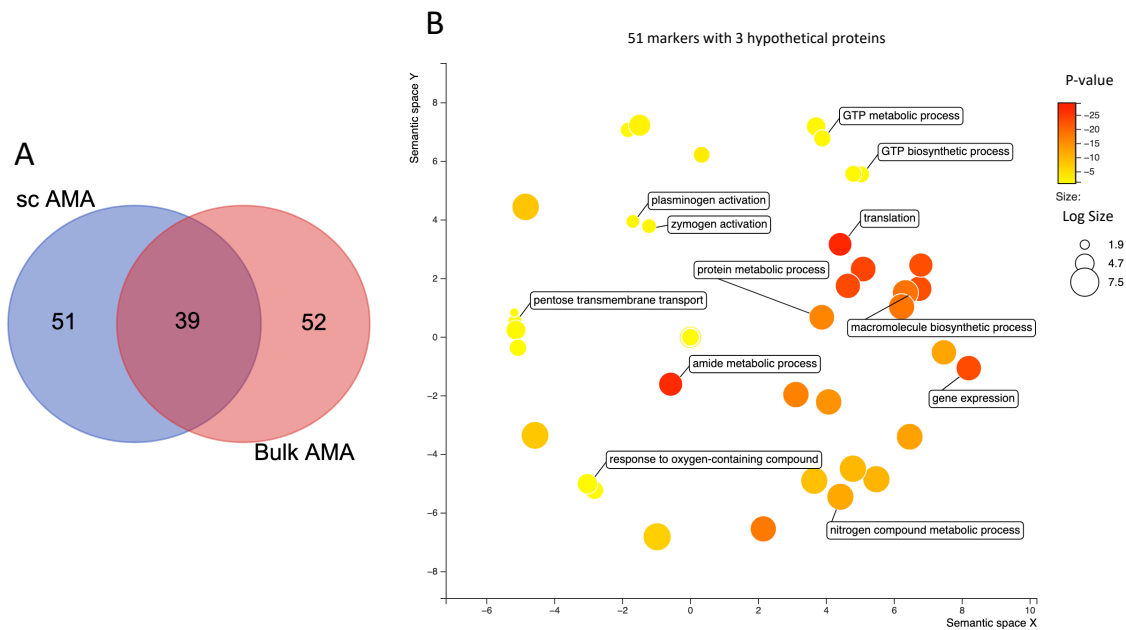


Figure 5-18 Venn diagram and gene ontology enrichment of markers exclusive to infected macrophage sample.

A) Venn diagram for single cell RNA-sequencing markers found uniquely in infecting *L. mexicana* sample (sc AMA) against unique bulk RNA-sequencing markers (Bulk AMA) derived from Fiebig et al., (2015). 51 markers were found to be unique to the sc AMA sample. **B)** Gene ontology (GO) term analysis by REVIGO of 51 unique sc AMA markers, with 3 hypothetical proteins. Non-linear projection of multidimensional scaling by REVIGO for semantically similar GO enrichment terms clustered close together in semantic space. 51 markers for the sc AMA sample were analysed for GO terms with TryTrypDB and ontologies for biological processes plotted by REVIGO in a scatter plot coloured by p-value with log size relative to the size of dots. GO terms of interest highlighted.

In Figure 5-18, **A)** the comparison of scRNA-seq amastigote markers from cluster 1 (sc AMA) to distinct bulk RNA-Seq markers (Bulk AMA) revealed 51 new markers not previously described by bulk RNA-Seq analysis. Present in these 51 markers expressed in these cells were amastigote marker genes, corresponding to amastigotes being identified, as expected in these experiments.

These 51 novel amastigote markers contained 3 hypothetical proteins and were analysed for GO terms as previously described. Notable GO terms included gene expression and translation associated with ribosomal genes, and metabolic processes.

5.3.4 Integrated samples from promastigote, axenic amastigote and amastigotes from macrophage infections

Finally, *L. mexicana* cells detected by scRNA-seq within the infected macrophage cells were isolated and integrated with previous *in vitro L. mexicana* scRNA-seq datasets, as described in Chapter 3. All previous quality

controls were used, as described in Figure 5-15. Principle component and resolution analysis for integrated samples is shown in Figure 5-19.

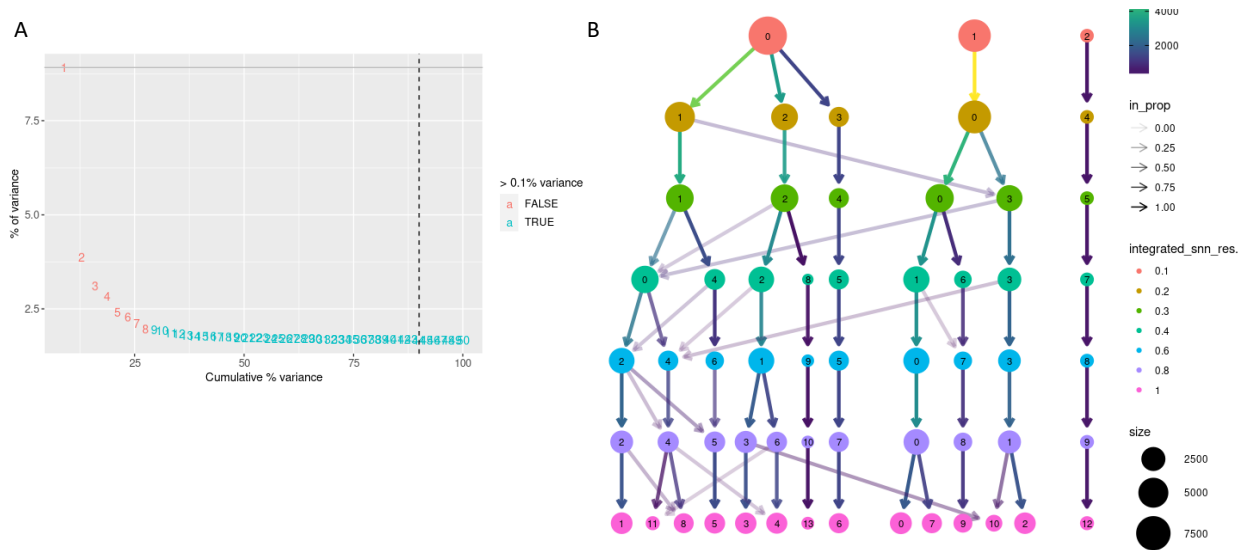


Figure 5-19 Elbow plot of variance found in principle components observed in all integrated *L. mexicana* samples and Clustree plot demonstrating movement with increasing cluster resolution.

A) Elbow plot to rank principle components (PC) found in all integrated *L. mexicana* samples assessing the percentage of variance explained by each PC. In the integrated *L. mexicana* samples, an “elbow” is seen around PC 8, suggesting that the majority of variance is captured in the first 8 PCs. **B)** Clustree plot for all integrated *L. mexicana* samples showing changes in sample movement as resolution increases (RNA_snn_res), from top to bottom. Number of cells moving from one cluster to another is represented by the colour of the arrow, where yellow arrows indicate a large proportion of cells moving from one cluster to another as resolution is increased. The size of clusters determined by the number of cells portioned within as resolution increases (from top to bottom) is indicated by the relative size of depicted circles. A resolution of 0.4 (coloured in sky blue) is chosen resolving into 9 clusters for further clustering analysis.

Plotting these integrated experiments on UMAP for revealed clustering as displayed below in Figure 5-20. Various iterations were considered for clustering analysis, with the below version used for marker selection due to being the lowest resolution providing clusters with differentially expressed cluster markers.

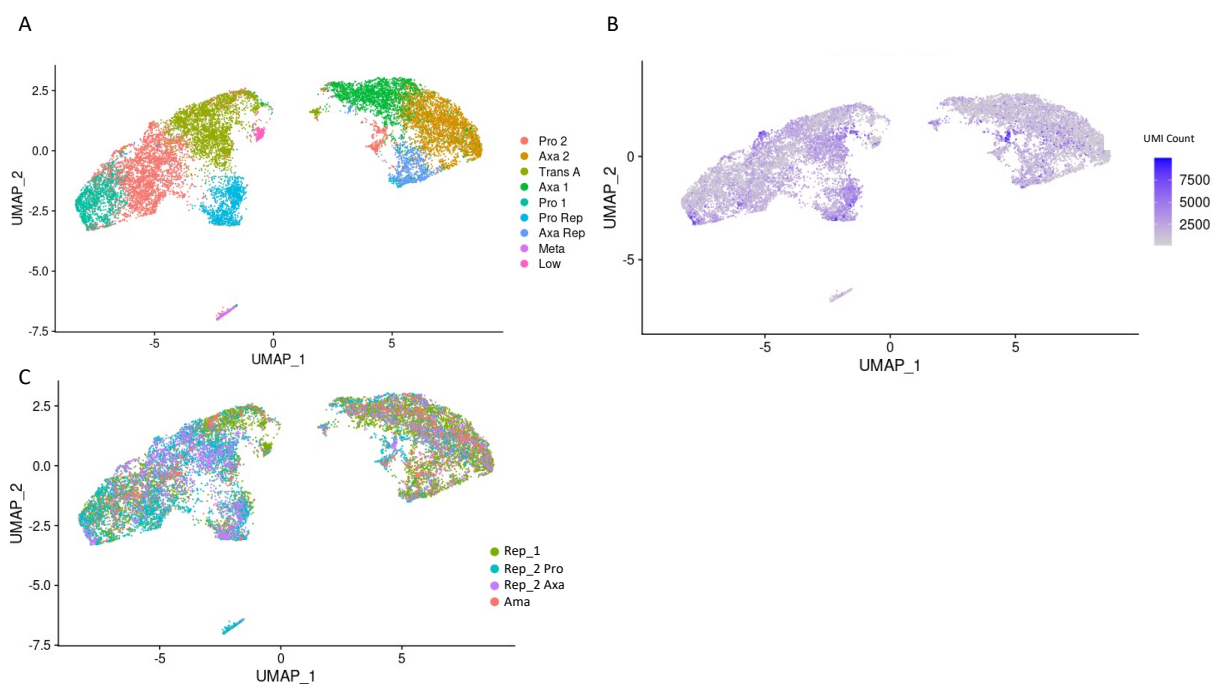


Figure 5-20 Clustering of all integrated *L. mexicana* samples.

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction of clustering in integrated samples to visualise relative relationships between individual transcriptomes. **A)** UMAP of integrated samples at a resolution of 0.6 reveals 10 distinct clusters. **B)** UMAP coloured by total raw transcript counts per cell. Scale shows raw transcript counts per cell. **C)** UMAP of integrated samples coloured by original sample identity; Rep1 (green), Rep2_Pro (turquoise) and Rep2_Axa (purple) and Ama samples from infected macrophage experiments (red).

Preliminary clustering shown in Figure 5-20 **A** revealed 9 distinct clusters.

Applying marker analysis used in previous *in vitro* experiments, clusters were defined by life cycle stages. 4 promastigote clusters were revealed, named Pro1 (coloured in apple green), Pro 2 (coloured in red), Pro Rep for a proliferative cluster, S phase enriched as previous integrated analysis in Chapter 3, (coloured in turquoise) and Meta (coloured in purple). Three amastigote clusters were allocated and named Axa 1 (coloured in lime green) Axa 2 (coloured in ochre) and Axa Rep (coloured in cobalt blue) for a proliferating amastigote cluster, as previously identified. The transitional cluster detected *in vitro*, named Trans A here, is coloured in ochre, and was again positioned between promastigote and amastigote clusters and proximal to the Pro 2 and Pro Rep clusters. Notably, the low expressing cluster was separated in this integrated object (coloured in pink) and again placed between the Trans A cluster and amastigote clusters. In **B** the raw gene expression count is displayed for each cell, coloured in purple for cells with higher expression counts. Consistent with previous analysis of the Rep2_Pro and *in vitro* integrated analysis, very low transcript levels were seen in the Meta cluster (coloured in pink in **A**, consistent with a non-cycling life cycle stage. As previously demonstrated, the Meta cluster also sat noticeably separate from the

rest of the promastigote cells, placed between promastigote and amastigote clusters. Shown in **C** are the cells coloured by the replicate they originated from following integration, where Rep1 is coloured in green, Rep2_Pro coloured in turquoise, Rep2_Axa coloured in purple, and amastigotes isolated from macrophage infections named Ama and coloured in red. Cells from the Ama cluster appear to be integrated in place with in vitro cells, rather than forming a distinct cluster.

5.3.4.1 Marker analysis of all integrated *L. mexicana* experimental samples

Proportions of cells in each cluster, separated by the sample they originated from, were analysed, and plotted in **A**, coloured by cluster. A dot plot with previously used life cycle and cell cycle stage markers clusters was produced. Clusters plotted on the y-axis were placed in order of estimated life cycle progression, and promastigote to amastigote life cycle markers, as described above, ordered left to right on the x-axis with cell cycle markers in the centre, shown in **B**. The top five differentially expressed markers for each cluster were plotted in a heatmap shown in **C** in Figure 5-21.

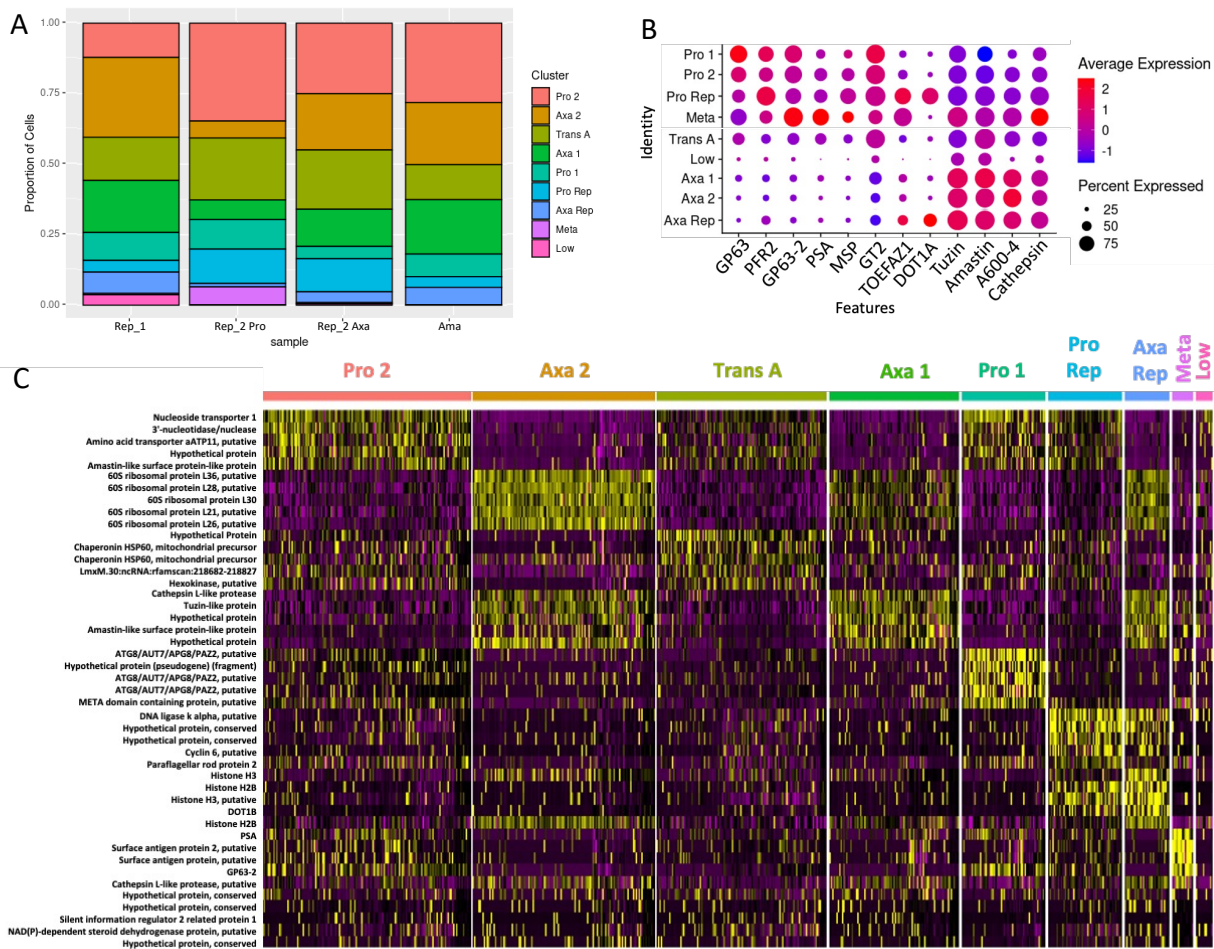


Figure 5-21 Allocation of clusters by life cycle stage and top differentiated marker analysis by heatmap with dot plot life cycle stage marker progression in all integrated *L. mexicana* samples.

A) Proportional groupings of cells in each cluster and coloured by cluster, as in Figure 5-20. **B)** Dot plot of marker genes for life cycle specific stages, as Figure 5-20, ordered from left to right for promastigote to amastigote, with life cycle progression for clusters ordered from top to bottom. Coloured by average expression and percentage expressed in allocated cluster indicated by size of dot. **C)** Heatmap of top five distinguishing markers for each cluster for relative expression levels (log₂ normalised z-score), where each row represents a gene coloured by relative expression and each column is a single cell categorised by assigned cluster.

Cluster proportions (shown in **A**) for the fourth experiment featuring *L. mexicana* transcripts isolated from the infected macrophage sample (named Ama) surprisingly contained cells from early promastigote clusters, such as Pro 1 and Pro 2 (coloured in apple green and red, respectively), again suggesting the presence of promastigote stages phagocytosed by macrophages and most likely present in the 4 h timepoint. Also featured in the Ama experiment was the second largest proportions of cells clustered into Axa 1, Axa 2 and Axa 3 clusters. The Ama experiment featured no cells clustered into the Metacyclic cluster, suggesting life cycle progression from the metacyclic stage within the captured samples and any metacyclic cells progressed into amastigote forms. However, low transcript expression of metacyclic genes may hamper identification of cells with lower relative mRNA abundance compared to other

life cycle stages (Inbar, et al., 2017; Cortazzo da Silva et al., 2022). The promastigote to amastigote life cycle stage markers in **B**, as previously selected in marker analysis in Chapter 3, again matched expression profiles for clusters allocated to life cycle stages, with a consistent trend of proportional expression from the top left to bottom right. The two markers used to identify clusters enriched for cells highly expressing cell cycle progression markers (TOEFA1 and DOT1A) again also showed higher expression in promastigote and amastigote clusters Pro Rep and Axa Rep featuring cells with higher cell cycle gene expression. Top 5 marker proteins for each cluster, shown in heatmap **C**, remained largely consistent with previous cluster markers as shown in Figure 3 - 36. This, combined with cluster data and overlap of replicates as shown in Figure 5-20 **C**, indicated no distinct clustering of cells in the Ama sample and instead a general mix across 7 clusters was seen. Additionally, the Low cluster, previously excluded from analysis in the Rep1 experimental (Figure 3 - 7 **A**) sample due to no significant markers during analysis, was reproduced at this resolution in a similar proximal location near the Trans A cluster and now with significant markers (see below).

Investigation into gene ontology for the markers associated with the Low, Trans A and Meta clusters was performed where markers were entered into TriTrypDB (Amos et al., 2022) and GO term enrichment was examined for biological processes. The resulting GO terms were plotted via the Reduce and Visualise Gene Ontology (REVIGO) application via scatterplot (Supek et al., 2011). GO terms of interest were labelled in the scatterplots produced via REVIGO below in Figure 5-22.

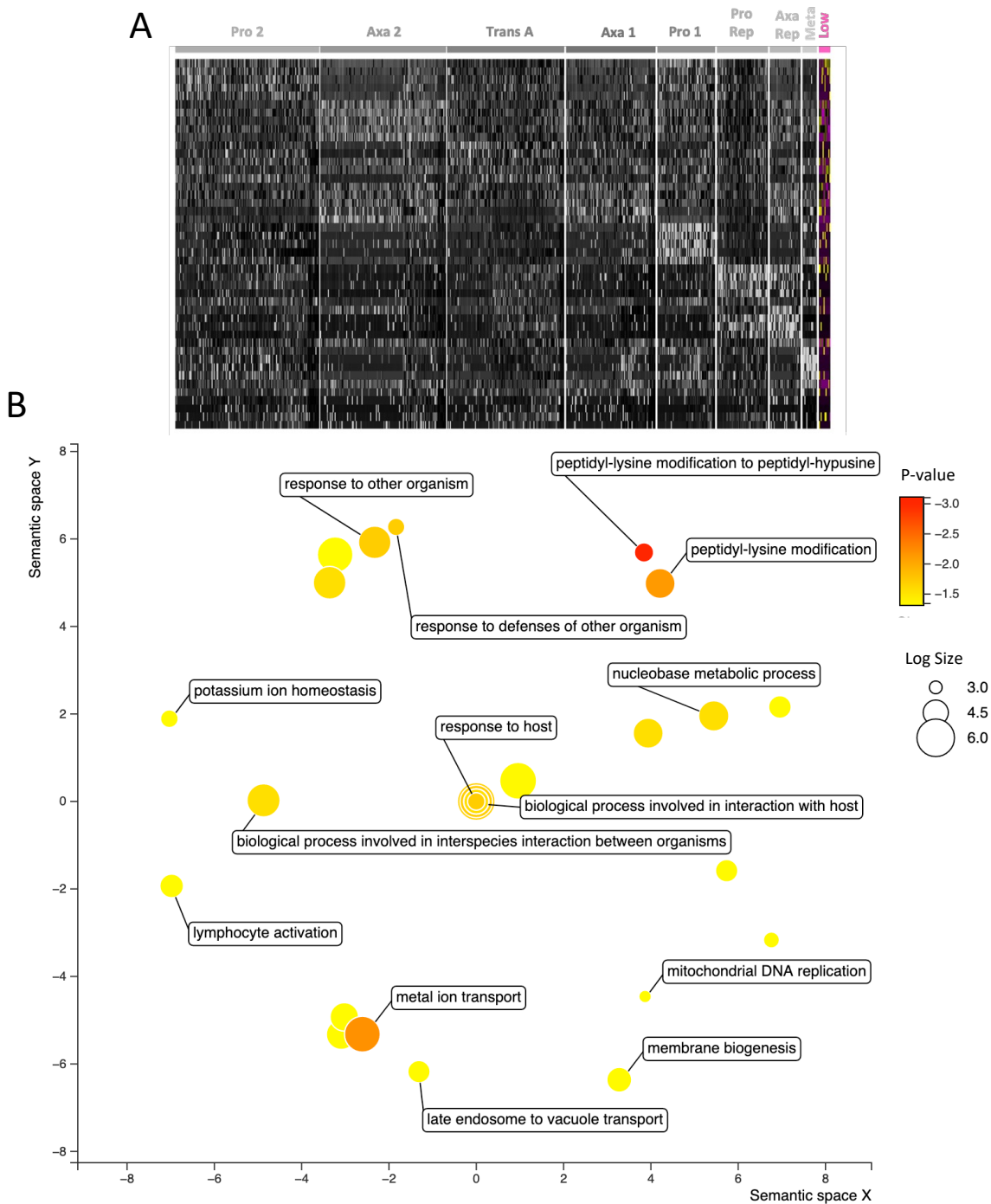


Figure 5-22 Gene ontology term analysis by REVIGO of Low cluster.

A) 342 markers from the Low cluster with 132 hypothetical proteins were analysed for gene ontology (GO) enrichment **B)** Non-linear projection of multidimensional scaling by REVIGO for GO enrichment terms to reduce the dimensionality of the matrix of GO terms pairwise semantic similarities. Semantically similar GO terms clustered close together in semantic space. 342 markers were analysed through TryTrypDB and ontologies for biological processes plotted by REVIGO in a scatter plot coloured by p-value with log size relative to the size of dots. GO terms of interest highlighted.

Scatterplots generated via REVIGO were plotted for the Low cluster, with GO terms of interest highlighted and plotted in semantic space. In these scatterplots p-value of the GO terms is indicated by colour, where red indicates a lower p-value (legend in upper right-hand corner), and the size of the dots indicates the frequency of the GO term in the underlying Gene Ontology Annotation database,

where dots of more general terms are larger in size. For the Low cluster, 342 markers were found, including 132 hypothetical proteins. GO terms included peptidyl-lysine modification to peptidyl-hypusine, peptide-lysine modification, and metal ion transport. Also of note were GO terms associated with interspecies host interactions

To further compare these data, Venn diagrams of each of the marker sets were compared between Meta and Low clusters below in Figure 5-23.

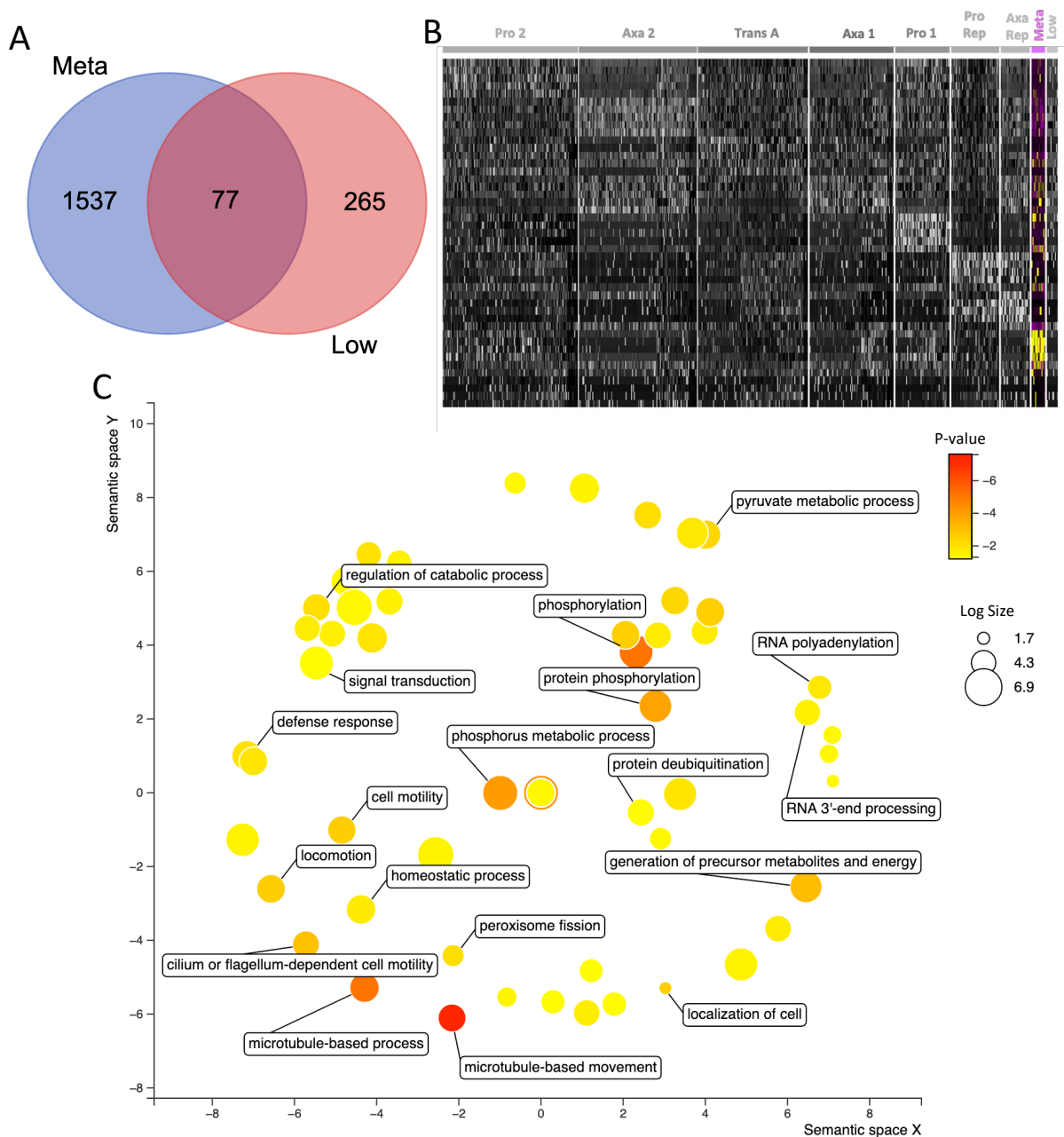


Figure 5-23 Gene ontology term analysis by REVIGO of Meta cluster.

A) Venn diagram for Meta and Low clusters in all integrated *L. mexicana* samples reveals 1614 distinct markers for the Meta cluster and 236 distinct markers for the Low cluster. **B)** 1526 markers from the Meta cluster with 651 hypothetical proteins were analysed for gene ontology (GO) enrichment **C)** Non-linear projection of multidimensional scaling by REVIGO for GO enrichment terms to reduce the dimensionality of the matrix of GO terms pairwise semantic similarities. Semantically similar GO terms clustered close together in semantic space. 1614 markers were analysed through TryTrypDB and ontologies for biological processes plotted by REVIGO in a scatter plot coloured by p-value with log size relative to the size of dots. GO terms of interest highlighted.

The Meta cluster was found to contain 1,614 distinct markers with the Low cluster containing 236 markers. An overlap of 77 genes, constituting 4.5%, was seen between these two clusters. The top three GO terms for the Meta cluster were ‘microtubule-based movement’, ‘phosphorylation’ and ‘microtubule-based process’. Also of note were GO terms associated with ‘RNA polyadenylation’ and

'RNA 3'-end processing', possibly indicating a pre-emptive modification of gene expression associated with life cycle development.

To further compare these data, a Venn diagram of each of the marker sets were compared between Low, Meta and Trans A clusters in Figure 5-24.

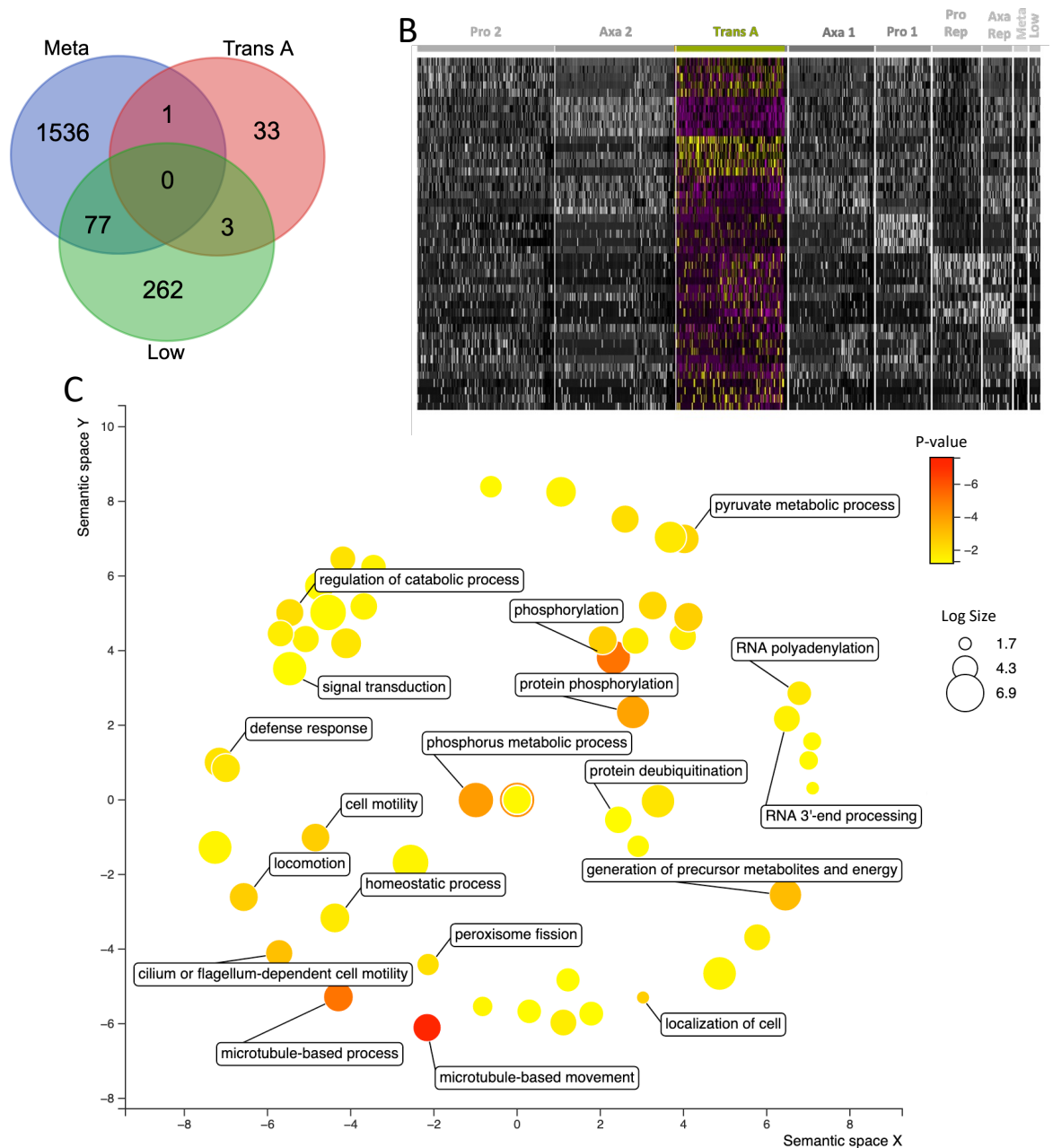


Figure 5-24 Gene ontology term analysis by REVIGO of Trans A cluster.

A) Venn diagram for Meta and Low clusters in all integrated *L. mexicana* samples reveals 37 markers for the Trans A cluster with 33 being unique to the Trans A cluster when compared to Meta and Low clusters. **B)** 1526 markers from the Meta cluster with 651 hypothetical proteins were analysed for gene ontology (GO) enrichment **C)** Non-linear projection of multidimensional scaling by REVIGO for GO enrichment terms to reduce the dimensionality of the matrix of GO terms pairwise semantic similarities. Semantically similar GO terms clustered close together in semantic space. 1614 markers were analysed through TryTrypDB and ontologies for biological processes plotted by REVIGO in a scatter plot coloured by p-value with log size relative to the size of dots. GO terms of interest highlighted.

Interestingly, no overlap between Low, Meta and Trans A clusters were discovered, with most similarity seen in Low and Meta clusters, as demonstrated previously. Trans A shared only 1 marker with Meta, being the gene LmxM.28.0140, pantothenate kinase subunit associated with coenzyme A (CoA) biosynthesis in synthesis and oxidation of fatty acids. The three markers shared between Trans A and Low clusters were LmxM.34.0270, a hypothetical protein; LmxM.07.0990, a nucleolar RNA-binding protein, putative; and LmxM.23.1665, Putative phosphatidic acid phosphatase (PAP2) superfamily protein.

5.4 Discussion

To investigate the host and parasite gene expression when *L. mexicana* infects human macrophages, scRNA-seq was used to compare transcriptional responses from both species by dual transcriptome analysis. Here we presented both the first example of human cells in *L. mexicana* infections analysed by scRNA-seq, as well as the first example of *L. mexicana* cells during their infection of human cells. ScRNA-seq analysis of human macrophages infected with *L. mexicana* reveals new markers associated with *Leishmania* infections not previously described in bulk RNA-sequencing analysis

The analysis presented above produced 11,949 differentially expressed markers for THP-1 derived macrophages infected with *L. mexicana*. When compared to uninfected macrophage controls taken at the same timepoints post-differentiation, an overlap of 7,566 genes (63.3%) was found. Thus, 1,893 genes, or 15.8%, were discovered that were unique to infected samples (Figure 5-13). PANTHER gene list analysis (Mi et al., 2013) for GO enrichment revealed 18 GO terms within the 1,814 distinct genes for infected macrophage markers by scRNA-seq when compared to bulk RNA-sequencing markers generated by Fernandes et al., (2016) (Figure 5-14). Of note in the 18 GO term categories were biological process involved in interspecies interaction between organisms (GO:0044419), containing 4 markers, and immune system process (GO:0002376), containing 9 markers, of which 4 were also found in GO:0044419 shown in Table 1-1 (Figure 5-14).

Of the nine genes found to be involved with interspecies interactions and immune system process GO categories, each has previously been associated with

regulation and immune responses in *Leishmania* infections, as described below. Indeed, four of these genes were found in both interspecies interaction between organisms and immune system process: *RNASE2*, *SRC1*, *MYD88*, and *RELB*.

RNASE2 codes for a non-secretory ribonuclease regulated upstream by the TNF pathway and has previously been identified as a differentially regulated pathway with patients contracting visceral leishmaniasis (VL), where lymph node aspirates were examined following treatment with sodium stibogluconate using Affymetrix microarrays (Salih et al., 2017). Of note, this transcript was also described as being upregulated in macrophages infected after 4 hours by both Dillon et al. (2015) and Fernandes (2016), indicating consistent markers between different methods of transcriptome analysis and between *Leishmania* species. Additionally, *RNASE2* has been described as having immune system roles in antiviral activity, and cleavage of ncRNA in macrophages induced from THP-1 cell lines (Lu et al., 2022).

SRC1 codes for a non-receptor tyrosine protein kinase and has previously been identified as a co-activator protein in VL infections (E. Y. Osorio et al., 2012) and reprogramming of host metabolism in *Leishmania* infections (Reverte et al., 2021). *SRC1* is a co-activator protein, along with RNA polymerase II and several other subunits, where together they form a complex enhancer-binding element in the promotor region of Arginase 1, which initiates expression via transcription with STAT6 (Gray et al., 2005). STAT6 is a well described key regulator of the IL-4 signalling pathway (Chu et al., 2021). Osorio et al., (2012) describe VL progression by parasite induced STAT6 activation, indicating a possible key component of *L. mexicana* infection progression. Additionally, Reverte, et al. (2021) describe SRC family of protein tyrosine kinases (SFKs) activating nuclear factor erythroid 2-related factor 2 (*NRF2*), where upregulation was found in *Leishmania* infections. Interestingly, SRC kinases have also been described as activating Abl family kinases, which were found to be required for efficient phagocytosis for *Leishmania* infection, indicating a strong role in Leishmaniasis disease responses and increased parasitaemia (Wetzel et al., 2012, 2016).

MyD88 is a myeloid differentiation primary response protein, which functions as a scaffold and adaptor protein, previously described in infections with mice lacking the Toll-like receptor (TLR) pathway adaptor protein MyD88 (de Veer et

al., 2003). In these experiments *MyD88* knock-out mice were more prone to infection with *L. major* than wild-type C57BL/6 mice. These results indicate a possible ligand on *L. major* cells for TLRs and infer a novel and a critical role for the infection response from the innate immune cells.

RELB is a Rel homology transcription factor, forming part of the NF- κ B pathway involved in host defences against *Leishmania* infection, where reduced expression of NF- κ B family members in modified mice resulted in increased *Leishmania* infection rates (Bichiou et al., 2021). Other roles inferred from *relB* deficiencies include multi-organ inflammation in mice, demonstrating the inhibitory effect of *relB* protein in inflammation (Weih et al., 1995).

Five of the remaining nine genes categorised into the GO term only immune system processes (Table 1-1) were *CTSH*, *HLA-B*, *FCGRT*, *CD99*, and *CXCL16*. *CTSH* codes for a Pro-cathepsin H, cysteine protease, expression of which has been described as downregulated in *Leishmania amazonensis* infections in comparisons between BALB/c and C57BL/6 mice by dual global transcriptome profiles of bone marrow-derived macrophages (BMDMs) using bulk RNA-sequencing with Illumina NovaSeq by Aoki, et al. (Aoki et al., 2019).

HLA-B and *FCGRT* are MHC genes. *HLA-B* is a Human Leukocyte Antigen and forms part of the major histocompatibility complex, being an HLA class I histocompatibility antigen and B alpha chain protein. HLA has recently been considered a candidate for Leishmaniasis vaccines (Mosaad, 2015; Ranjan Dikhit et al., 2018) and was shown in *Leishmania donovani* (*L. donovani*) infections to evoke an anti-*Leishmania* CD8⁺ T lymphocyte response (Dikhit et al., 2018). *FCGRT* also codes for subunit of the major histocompatibility complex, as part of the IgG receptor FcRn large subunit p51 and has been described as a marker for the pro-inflammatory Th1-type immune response by macrophages in *Leishmania chagasi* when examined by RNA microarray (Rodriguez et al., 2004).

The CD99 antigen is a junctional adhesion molecule (JAM) and plays a role in transendothelial migration of leukocytes, where migration to locations of inflammation is well described as reliant on adhesion molecules (Bazzoni, 2003). Adherent leukocytes have previously been demonstrated as moving past endothelial monolayers with JAM based progression (Ley et al., 2007).

Experiments by Ballet et al (2014) demonstrated that blocking JAM-C junctions increased dendritic cell movement and relocation, thereby increasing the immune system response in *L. major* infections of both C57BL/6 and susceptible BALB/c mice (Ballet et al., 2014), where junctions are inhibited between JAMs and resulting in increasing vascular permeability after *L. major* infection.

And lastly, C-X-C motif chemokine 16 (CXCL16) plays roles in the mTOR pathway and has previously been linked with mTOR pathway mediated parasitaemia with *Leishmania* (Rashidi et al., 2021). mTOR signalling is well described in cell homeostasis and critical roles in autophagy, cell growth and regulation of metabolism. Autophagy is a strategy employed by various immune cells as a response to infections and is strategically important for immunosurveillance and pathogenesis (Rashidi, et al., 2021). CXCL16 has multifarious roles as a chemokine and is largely expressed in immune cells, such as macrophages. Yet, its role in parasitic infections is still to be fully explored (Rashidi et al., 2021; Veinotte et al., 2016). Recent studies in *L. donovani* by Chaparro et al. (2019) have demonstrated with Smart-seq2 transcriptomic analysis that an upregulation of CXCL16 occurs in BMDMs. Further observations were made between leishmanial-lipophosphoglycan (LPG) and initiation of CXCL16 expression (Chaparro et al. 2019). This expression of CXCL16 transcripts following *L. donovani* infections was also seen to be increased due to AKT/mTOR pathway processes. However, the mechanisms and interactions between CXCL16, LPG, and mTOR are yet to be fully understood (Chaparro et al. 2019).

Taken together, the markers described above unique to infected samples in this scRNA-seq analysis represent further evidence of TNF, IL-4, NF- κ B, and mTOR pathways associated with *Leishmania* infections. These pathways are also implicated by marker analysis in bulk RNA-sequencing samples by Dillon et al., (2016) and Fernandes et al. (2016), and would therefore justify further investigation into how these pathways may be modulated in Leishmaniasis.

5.4.1 Markers for *L. mexicana* infection process revealed by scRNA-seq

When analysing individual samples of infected macrophages, a cluster made up of *L. mexicana* cells (cluster 2, coloured in turquoise) was found, making up

12.2% of the cells in this sample. Marker analysis for cluster 2 produced 193 *L. mexicana* genes (Figure 5-12).

Additionally, a comparison to bulk RNA-sequencing markers was made using markers generated by Fernandes et al., (2016), using transcripts from *L. major* infections at the 4 h timepoint in Figure 5-18, where 51 markers unique to amastigote markers derived from scRNA-seq analysis were discovered. These 51 novel amastigote markers contained 3 hypothetical proteins and were analysed for GO terms by TriTrpDB and REVIGO. Notable GO terms revealed included gene expression and translation associated with ribosomal genes. In addition, metabolic GO terms likely associated with metabolic processes involved with survival within the host and nutrient recruitment were found. (Westrop et al., 2015) have shown that amino acid use and metabolism, particularly for tryptophan, aspartate, arginine, and proline vary between various forms of Leishmaniasis. Additionally, variation in amino acid recruitment could be speculated as being specific to pathologies caused by *Leishmania* and may provide further insight into metabolic processes involved in host cell prevalence (Westrop et al., 2015).

5.4.2 Novel markers for *L. mexicana* life cycle stages and prospective new life cycle stage defined by transcriptomic profile

GO term analysis via TriTrypDB and REVIGO was also undertaken for clusters in the integrated *L. mexicana* samples (clustering in Figure 5-20A). The cluster of interest in this analysis were the Low cluster reformed from Rep1 (Figure 3-10), the Trans A cluster and Meta cluster. In the Low cluster, 342 markers were found with 132 hypothetical proteins. GO terms highly differentiated included peptidyl-lysine modification to peptidyl-hypusine, peptide-lysine modification and metal ion transport associated with eukaryotic translation of amino acids and metabolic processes (Park & Wolff, 2018). Also of note were GO terms associated with interspecies host interactions (Figure 5-22, B)

The Meta cluster was found to contain 1,614 distinct markers with the Low cluster containing 236 markers. An overlap of 77 genes, constituting 4.5%, was seen between these two clusters. The top three GO terms for the Meta cluster

were microtubule-based movement, phosphorylation and microtubule-based process, indicating markers associated with flagellum extension also associated with metacyclic differentiation (Wheeler et al., 2011). Also of note were GO terms associated with RNA polyadenylation and RNA 3'-end processing, possibly indicating a pre-emptive modification of gene expression associated with life cycle development (Bates et al., 2000). This finding may also be indicative of higher polyadenylation seen in amastigote forms compared to promastigotes, where polyadenylation of ribosomal RNA has been described as transitory and variable between life cycle stages by Decuypere et al., (2005)

Strikingly, no markers were shared between Low, Meta and Trans A clusters, with most similarity seen in Low and Meta clusters (as stated above, they shared 77 markers). Trans A shared only 1 marker with Meta, the gene LmxM.28.0140, encoding pantothenate kinase subunit associated with coenzyme A (CoA) biosynthesis in synthesis and oxidation of fatty acids. Pantothenate kinase (PanK) catalyses the initial reaction in CoA synthesis and has been extensively characterised in *Plasmodium falciparum* (Tjhin et al., 2018), where also pantothenate mimetic compounds were catabolised into CoA antimetabolites clearing parasites in a humanised rodent infection model (de Villiers et al., 2017; Schalkwijk et al., 2019). The three markers shared between the Trans A and Low clusters were LmxM.34.0270, a hypothetical protein; LmxM.07.0990, a nucleolar RNA-binding protein, putative; and LmxM.23.1665, Putative phosphatidic acid phosphatase (PAP2) superfamily protein. The latter is a membrane-bound protein associated with controlling cell cycle progression, intracellular lipid content and increased virulence in sand flies compared to the wild type cells when knocked-out by CRISPR-Cas9 deletion in *Leishmania guyanensis* (Zakharova et al., 2022). These targets represent markers for further investigation in life cycle progression in *L. mexicana*.

All together, these datasets in this chapter offer an insight into both the unique gene expression associated with infections of THP-1 derived macrophages with *L. mexicana*, and the response to infection as seen in the parasite. This represents a focused analysis of the initial stages of the infection process for both interacting organisms, which is crucial in elucidating parasite persistence and

clearance. The results herein offer a resource for future studies toward possible better understanding of the life cycle, and infection by amastigotes.

5.5 Summary

- First to generate and analyse single cell transcriptomic data to identify markers associated with human macrophage infections with *Leishmania*.
- Identified genes differentially expressed in *L. mexicana* during macrophage infections.
- Combined all experimental single-cell experiments to generate a foundational cell atlas for prospective future analysis of life cycle and cell cycle progression in *L. mexicana*.

Chapter 6 Concluding remarks

Taken together, this thesis presents the first application of scRNA-seq analysis to deconvolve the timing and patterns of gene expression changes during life cycle progression in the *L. mexicana* parasite. Previous application of scRNA-seq are currently limited to analysis of the gamete fusogen *HAP2* to deconvolve *L. tropica* reproduction (Louradour, et al., 2020), and infection of mouse models with *L. major* and *L. donovani* (Venugopal et al., 2022; Karagiannis et al., preprint). In Chapter 3, known life cycle stage markers were used to determine the identities of promastigote and amastigote clusters, before DE analysis was used to compare clusters and identify novel markers of life cycle forms, and sub-populations within each, not previously revealed by bulk RNA-Seq (Fiebig, et al., 2015). Cell cycle phase markers, taken from orthologues of *T. brucei* cell cycle phase from Archer et al., (2011), were used to identify proliferative populations both in individual experiments and in integrated sample analysis. The former revealed transcriptionally distinct proportions of promastigote stages overlapping with cell cycle phases, previously postulated as being overlapping (Wheeler et al., 2011; Sunter and Gull 2017).

Notably, these data provide a resource of cell cycle regulated transcription dynamics currently lacking for *Leishmania* and would certainly represent a formative step in further investigations for comparisons of life cycle stages to cell cycle phases in promastigotes. Markers identified for the infective metacyclic promastigote forms, not previously described using bulk RNA-Seq analysis in *L. mexicana*, were also described in Chapter 3. While bulk analysis of promastigote stages was available for promastigote stage labelling, studies are limited to *L. infantum*. Here, scRNA-seq was able to identify the different promastigote stages without the need to sort cell types first. Thus, these data could be mined further for stage makers to identify *L. mexicana* promastigote stages experimentally. Additionally, haptomonad forms may also be considered, being an alternative life cycle stage originating from leptomonads, which are so far lacking in bulk RNA-Seq experiments (Yasmin et al., 2022). Also for consideration was the finding of metacyclic forms enriched in the G1 cell cycle stage when labelled with *T. brucei* cell cycle stage orthologues (Archer et al., 2011). This results also corresponds with infective metacyclic form of *T. brucei* identified as being arrested in G1/G0 by flow cytometry when taken from the

tsetse fly (Shapiro et al., 1984). Further experiments could investigate if this is represented in *in vitro* and *in vivo* metacyclic forms, by flow cytometry.

A potentially new life cycle stage was also described in Chapter 3, here named Trans A, that was defined by being transcriptionally overlapping with promastigote and amastigote genes, including examples for promastigote and amastigote stages being NT2 and Cathepsin, respectively. Further experimental investigations to characterise this cluster could use tagging strategies to study the 45 Trans A cluster markers identified here. With fluorescent tagging, as demonstrated in Chapter 4, cell lines could be generated and further analysed using fluorescence activated cell sorting (FACS) to isolate potential Trans A populations in both promastigote and amastigote stages. Tagging of one marker (LmxM.08.0810), in Chapter 4, suggested the Trans A cluster may be present in both, but more in promastigotes. If image stream flow cytometry was also adapted with FACS technology, it would be possible to further associate morphological forms with their respective single-cell transcriptomic profiles (Dandugudumula et al., 2022).

Additionally, we were able to order hundreds of genes using life cycle progression using pseudotime analysis, not previously applied to *Leishmania* scRNA-seq analysis. Using this analytical method, we found the Trans A cluster to sit between promastigote and amastigote life cycle stages, with a second separate trajectory ending in a metacyclic cluster, indicating the possibility of promastigote to amastigote progression without requiring progression through the metacyclic stage.

If correct, then metacyclics are not an essential intermediate stage in the formation of amastigotes from promastigotes, at least *in vitro*. Having identified 218 genes specifically expressed in metacyclics, and 45 in the Trans A cluster (with 2 overlapping), the possibility now exists to use CRISPR to knock-out these markers in promastigotes and analyse if differentiation is affected. Promising candidates could also be extended into *in vivo* to investigate if a metacyclic bypass route can be adopted in true host to vector transmission.

Further scRNA-seq analysis of *L. mexicana* life cycle development could also consider an amastigote to promastigote sample for comparison of developmental

genes required for differentiation between both stages, and indeed, identify those specific only to promastigote to amastigote life cycle transition, potentially revealing therapeutic targets that could be targeted to block transmission.

Also described in Chapter 4 were 91 hypothetical proteins noted to be differentially expressed across life cycle pseudotime progression, which were tagged using the high-throughput CRISPR/Cas9 method formulated by Beneke and Gluenz (2019) and remain available for further validations. These 91 tagged cell lines of hypothetical proteins represent potential new biology to explore as the parasite progresses through its life cycle development. Preliminary experiments notably showed a general trend in mRNA and protein expression coinciding together, however further work is needed, such as analysis of expression by flow cytometry across the same five timepoints selected for scRNA-seq samples for further comparisons to pseudotime expression profiles.

Human markers for *L. mexicana* infections in THP-1 derived macrophages were found in Chapter 5, indicating potential new cytokines pathways for interrogation in the infection response of host human cells. While current infection models using scRNA-seq analysis exist using mouse models (Venugopal et al., 2022; Karagiannis et al., preprint) the work presented here represents the first example of human cell models used for *Leishmania* infections analysed by scRNA-seq. Further analysis should consider human isolates for examination using scRNA-seq to further compare and identify new infection response pathways in humans. Additionally, markers identified in these data could be used to generate mutants (if not already available) and used to investigate how the infection dynamic by *Leishmania* is changed. This further work may well form the foundation of new molecular targets for medical interventions.

Another aspect of *Leishmania* biology could be considered is exploring cell cycle arrested, or “persister-like” amastigote forms, which are thought to be resistant to pathogen clearance by typical chemotherapeutic means (Barrett et al., 2019; M. C. Fernandes & Andrews, 2012; Kelly et al., 2020; Mandell & Beverley, 2017; Sánchez-Vald e et al., 2018; Vickerman, 1985; L. Zhang & Tarleton, 1999). Although questions persist as to how effective transcriptomic analysis of potentially transcriptionally sedentary cells may be, the possibility and potential

of examining very small, niche populations with few marker genes to distinguish one sub-population from another, the application of scRNA-seq to investigate these cell types remains enticing. Indeed, the previous requirements and limitations of bulk RNA-Seq are no longer a consideration for scRNA-seq, which opens the door for revealing rare populations in samples extracted straight from the host *in vivo*. For consideration, as described in Chapter 5, was the capture of *L. mexicana* transcripts from macrophages without needing to extract them, supporting the application of scRNA-seq to investigate persisters from tissue samples. Studies by the Lewis group have demonstrated persister cells in chronic *T. cruzi* infections move rapidly from one organ to another in anatomical reservoirs, using bioluminescence-guided tissue sampling mouse models (Ward et al., 2020). Rapid localisation followed by tissue isolation could reveal possible transcriptomic dynamics in the parasite as it moves and adapts from one organ to another. New research from the same group demonstrates chronic tissue parasite persistence in the colon of their mouse models, where *T. cruzi*-host interactions drive Digestive Chagas disease progression, and cell numbers during chronic stages of infection can be less than 200 in localised reservoirs (Khan et al., 2021). Such dynamics between chronic infections, persister cells and parasite-host interactions would be thoroughly described using scRNA-seq methodologies. Other examples of localised reservoirs of *T. cruzi* in mouse models include the heart, skeletal muscle, peripheral nerve, bladder, spleen, liver, adrenal gland, brain, and adipose tissue (Buckner, Wilson, and Van Voorhis 1999; Ward et al., 2020).

6.1 Spatial transcriptomics

In addition to these experimental considerations and prospects, spatial transcriptomic platforms can now be employed to investigate cell-to-cell interactions in both composition and function for health and disease, as reviewed by Briggs et al. (2021a). By combining scRNA-seq methodology developed by Stahl et al. (2016) with cell imaging technologies, it is now possible to measure spatial transcriptomics with positioning information in one experiment (K. H. Chen et al., 2015), as reviewed recently by Williams et al. (2022) and winning Nature's 'Method of the Year' in 2020 (Marx, 2021). Cellular RNAs are allocated barcodes dependant on their relative cellular position in a tissue section before the tissue sample is homogenised and then run through

scRNA-seq (Larsson et al., 2021). Other examples of how understanding could be improved for host-parasite interactions include physically interacting cell analysis with scRNA-seq (PIC-Seq), which compares interacting cells with non-interacting cells in sample to find gene markers associated specifically with cell-cell interactions (Giladi et al., 2020). Such analysis could be applied to tissue sections of both vector and parasite, to deconvolve gut specific interaction throughout the fly, and tissue specific locations in hosts and potential links between life cycle progression. Where spatial transcriptomics would add new knowledge for *Leishmania*; such as the site of genetic exchange in the fly; or the localisation of Trans A in the fly.

Further spatial-omic applications used in tandem with spatial transcriptomics could include proteomic approaches, such as hyperplexed localization of organelle proteins by isotope tagging (hyperLOPIT). This method previously applied to deconvolving the subcellular atlas of the *Toxoplasma* proteome (Barylyuk et al., 2020), which could be applied as a high-resolution proteomic resource for comparison with scRNA-seq analysis to compare mRNA and protein levels at higher resolutions.

6.2 Cell atlases

The vast amounts of high-resolution data now being generated through scRNA-seq can be brought together as a “cell atlas”, which forms an integrated resources for interrogation. Another application and important paradigm for the power of scRNA-seq datasets can be seen in cell atlases, as reviewed by Briggs et al., (2021a). Here, diverse datasets from studies looking at life cycle stages, cell cycle populations, rare cell types and varying developmental processes can be integrated to give a complete transcriptional profile of the combined cell types, providing greater detail than bulk datasets. Examples of the variable life cycle and cell cycle populations present here in *Leishmania* differentiation, once brought together, could provide an invaluable reference for mapping samples with low cell numbers and lower-quality captured transcriptomes (Howick 2019, Briggs 2021a). An example of the application of cell atlases can be found in the *Malaria* cell atlas compiled by Howick et al., (2019), who compiled the entire *Plasmodium berghei* life cycle. Consisting of 10 life cycle stages isolated from rodent models, this atlas was used to enable functional allocations of both

conserved and hypothetical genes and pointed towards further unexplored niches in conventional life cycle stages. Using this malarial cell atlas, transcriptomic samples from patient isolates containing *P. falciparum* and *P. malariae* could be mapped and assigned to developmental stages based on their respective parasitaemia. With these methods and the application of scRNA-seq strategies to tackle the deconvolution of the host-parasite interaction discussed above, further advances to describe cell types, and their developmental progress, will inevitably be forthcoming. The experiments present in this thesis constitute the *in vitro* *L. mexicana* life cycle progression from promastigotes to amastigotes, including development through the promastigote subtypes, and both axenic and macrophage derived amastigotes. Provision of these integrated data could be provided as a cell atlas for the community to interrogate.

6.3 Summary

In summary, the work presented here represent the first application of scRNA-seq pseudotime analysis of life cycle stage development in *Leishmania*. In pursuing this work, we provide thousands of new markers for life cycle stages, and cells labelled in various cell cycle phases for further investigation. 91 tagged cell lines of hypothetical proteins displaying transient expression along pseudotime life cycle stage development were also generated, potentially associating these uninvestigated proteins with life cycle progression and revealing new biology for further investigation. Additionally, human macrophage cells infected with *Leishmania* were also analysed by scRNA-seq for the first time and represents a novel addition to elucidating the parasite-host dynamic. Combined, the author hopes these data provide an additional example of the critical role for scRNA-seq in advancing the field of parasitology.

Appendices

Appendix I

6.4 Rep2_Pro quality control plots

6.4.1.1 Quality control filtering of single cell RNA-sequencing for second promastigote only replicate

Following unique reads being mapped and aligned to gene annotations in the custom reference genome and maxi-circle, counts were assigned to cell barcodes within single-cell GEMs using the Cell Ranger count function. Cell Ranger v3.0.2 (<http://software.10xgenomics.com/single-cell/overview/welcome>) was ran with default settings. Dual-species multipllets were identified by Cell Ranger count, as shown in Figure 0-1 A).

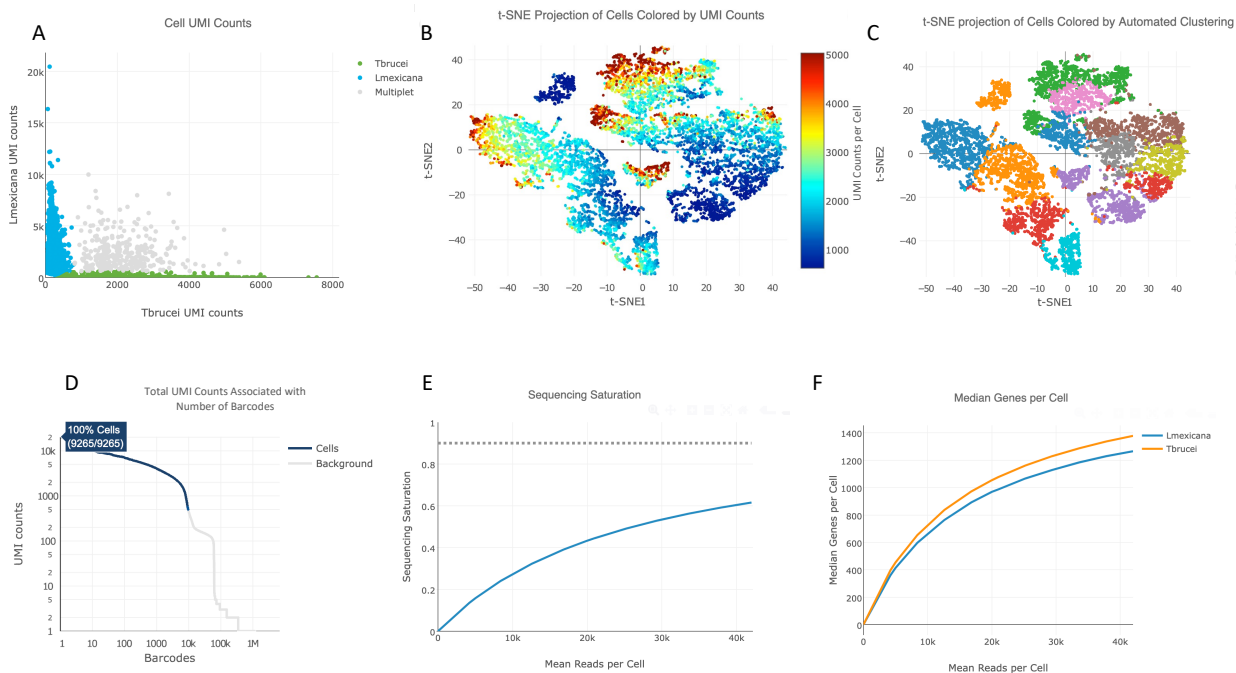


Figure 0-1 Cell Ranger summaries of sequenced Rep2_Pro sample.

Web summary plots of sequenced Rep2_Pro sample containing *Leishmania mexicana* (*L. mexicana*) and *Trypanosoma brucei* (*T. brucei*). Samples ran by Julie Galbraith and plots generated by Pawel Herzyk of Glasgow Polyomics in Cell Ranger (v3.0.2). A) Cell Unique Molecular Identifier (UMI) counts for each sample, *L. mexicana* in green, *T. brucei* in blue and multipllets in grey. B) t-distributed Stochastic Neighbour Embedding (t-SNE) plot of random 10,000 subset of cells in sample coloured by UMI counts. C) t-SNE plot of automated cell clustering in Cell Ranger clustered by similar expression profiles. D) Total UMI counts in mixed sample against barcoded counts of individual transcripts in mixed sample. E) Sequencing saturation of downsampled sequencing depth against mean reads per cell, with reasonably approximated saturation value point indicated by dotted line. F) Plot for median genes per cell against mean reads per cell for each sample mapped to their respective transcriptome for each species. Mapping run by Emma Briggs with Cell Ranger plots generated by Pawel Herzyk of Glasgow Polyomics.

Extensive explanations for each Cell Ranger plot can be found in Section 3.3.2.1. Briefly, Figure 0-1 A), 9,583 GEMs were present in this sample, with 6.5% multiplets having more than one GEM, and a mean UMI count purity of 95.6%. In Figure 0-1 B and C clustering in t-SNE generated by Cell Ranger again formed into two distinct groupings representative of the *L. mexicana* and *T. brucei* samples. In Figure 0-1 D, cells are represented by the blue line on the curve peaking around 9,500 cells, 5,550 of which were *L. mexicana* cells. Fraction of reads for the *L. mexicana* sample was estimated at 90.6%. In Figure 0-1 E) sequencing depth was again chosen as 50,000 mean reads per cell with approximate saturation point indicating suitable saturation reached at 50,000 reads here. In Figure 0-1 F) the median genes per cell for each mapped transcriptome to their relative species are plotted. Median genes per cell for the *L. mexicana* sample were 1,266, with total genes detected being 8,236 and median UMI counts per cell 1,889.

Outputs from Cell Ranger are then read, as explained for Rep1 in Section 3.3.2.1, with all consistent pre-processing quality control removing multiplets and *T. brucei* transcripts applied.

Shown in Figure 0-2 below, the quality control filters considered were number of unique features, UMIs, kDNA and rRNA contaminates, as performed for the Rep1 sample.

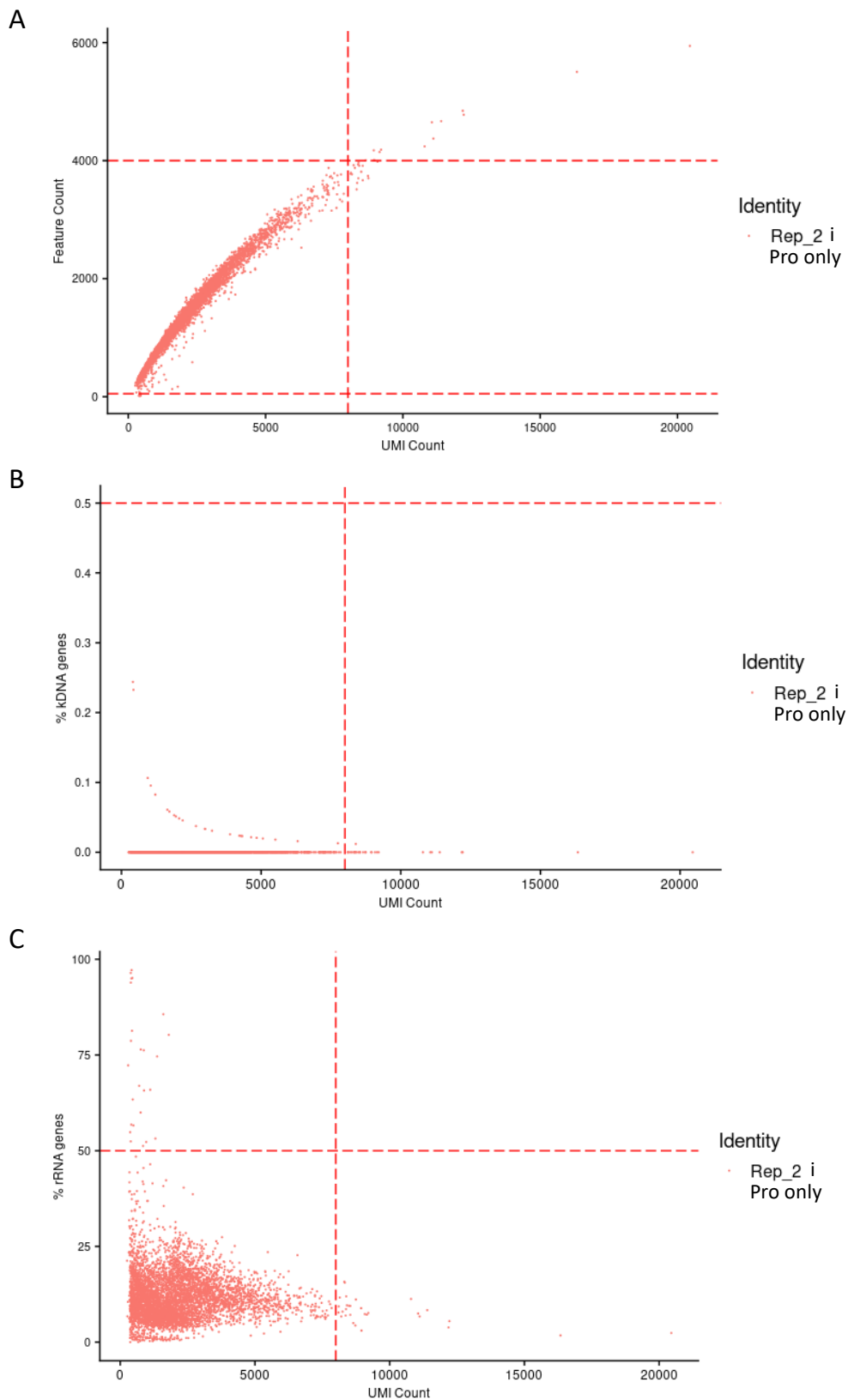


Figure 0-2 Quality control and filtering of transcriptomes in second biological replicate containing only promastigote stages.

Scatter plots of quality control and filtering cut-offs. Each dot representing one captured transcriptome. Unique Molecular Identifiers (UMI) are plotted against **A**) gene counts (features) with UMI counts <8000, feature counts >50 and <4000 to exclude for multiplerts, **B**) percentage of kDNA maxi circle genome included with mapping reads < 0.5% to exclude lysed or apoptotic cells and **C**) percentage of ribosomal RNA (rRNA) <50% and >5% contamination in sample. Red dashed lines indicate cut-offs used to filter cells per experiment.

In Figure 0-2 A) the number of unique features is set at a cut-off of 4,000 at the upper limit and 50 at the lower, to exclude empty droplets or cell multiplets which display unusually high gene counts. This lower limit is lower than typical to account for possible stationary phase growth typical of some life cycle stages in *Leishmania*. Similarly, the number of UMI for the total number of molecules found within a cell is also set at 8,000 as a cut-off. In B) kDNA metrics are calculated with the `PercentageFeatureSet()` function to provide a percentage of counts mapped to the kDNA maxicircle sequence included in mapping, and set as a cut-off for cells with more than 0.5% kDNA in their counts, as cells with low-quality or in the process of dying display higher mitochondrial contamination. Similarly, in C) rRNA contaminants are set at cells with more than 50% rRNA and then all rRNA transcripts removed to avoid clustering or markers associated with these counts.

As explained for Rep1, `scater::logNormCounts` was used to normalise feature counts in Rep2_Pro, and normalised values are then stored in the Seurat object under `Rep2_Pro[["RNA"]]`@data.

Next, a subset of highly variable features from cell to cell were selected as a starting point of genes of interest for further examination in downstream analysis using the `FindVariableFeatures()` function, selecting for the top 2,000 variable features, as in Rep1. The top ten variable features are labelled in the below scatter plot in Figure 0-3

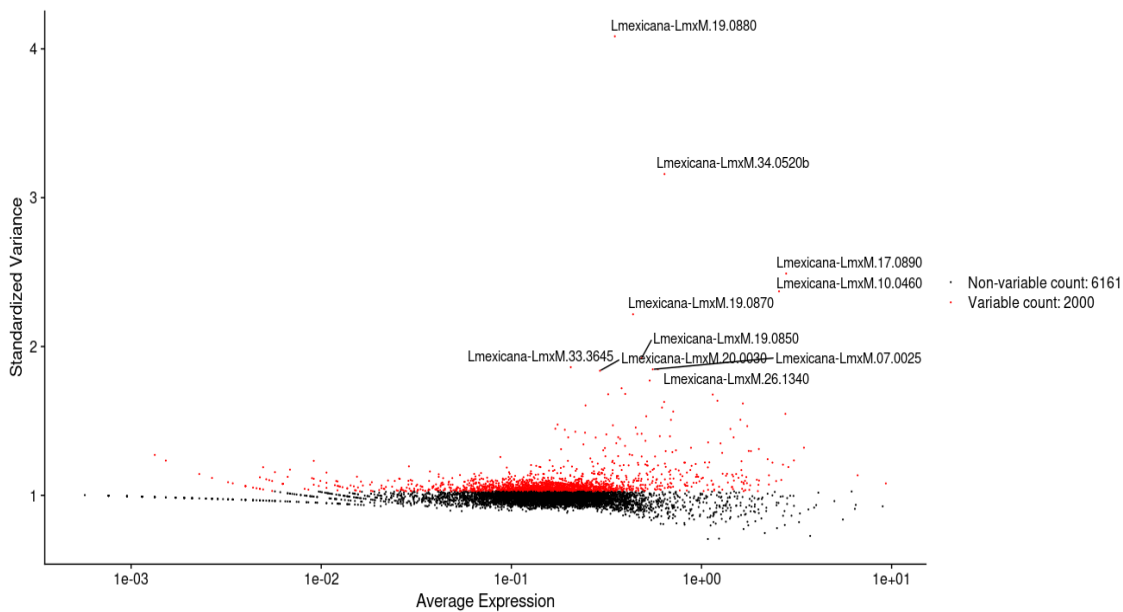


Figure 0-3 Scatter plot of top 2,000 variable features in Rep2_Pro dataset

Scatter plot of top 2,000 variable features (red) selected for downstream analysis in sample Rep2_Pro, plotting average expression of feature against standard variance. Top ten variable features labelled.

Known markers associated with promastigote life cycle stages within *Leishmania*, also previously seen in Rep1, were detected. These include GP63, LmxM.10.0460, which was also a top ten variable feature in Rep1, as well as META domain containing protein, LmxM.17.0890, a protein associated with the infective metacyclic promastigote stage. Also in the top ten variable features in Rep1. 3 of the top ten variable features being forms of ATG8/AUT7/APG8/PAZ2; LmxM.19.0850, LmxM.19.0870 and LmxM.19.0880, the first two forms also seen in the top ten variable features in Rep1.

Also of note was the absence of previously detected amastigote associated markers, such as LmxM.08.0810, encoding Cathepsin L. In addition, the previously used cell cycling markers enriched in the S-phase cells in the Rep1, DOT1A, LmxM.07.0025 and DOT1B, LmxM.20.0030, were recovered as two of the top ten variable features. In addition to these cell cycle markers, DNA ligase k alpha, LmxM.26.1340 was also present. These two markers combined indicate a variability in cycling populations within this sample, consistent with replicating and non-replicating promastigotes stages. Lastly, the hypothetical protein pseudogene fragment LmxM.34.0520b was detected, as also seen in Rep1. Taken together, these data suggest data reproducibility in the approach, and confirm the life cycle predictions in Rep1.

Next, linear dimension reduction was performed by PCA on the REP2_Pro data with scaling on the top 2,000 features (Figure 0-4), as in Rep 1.

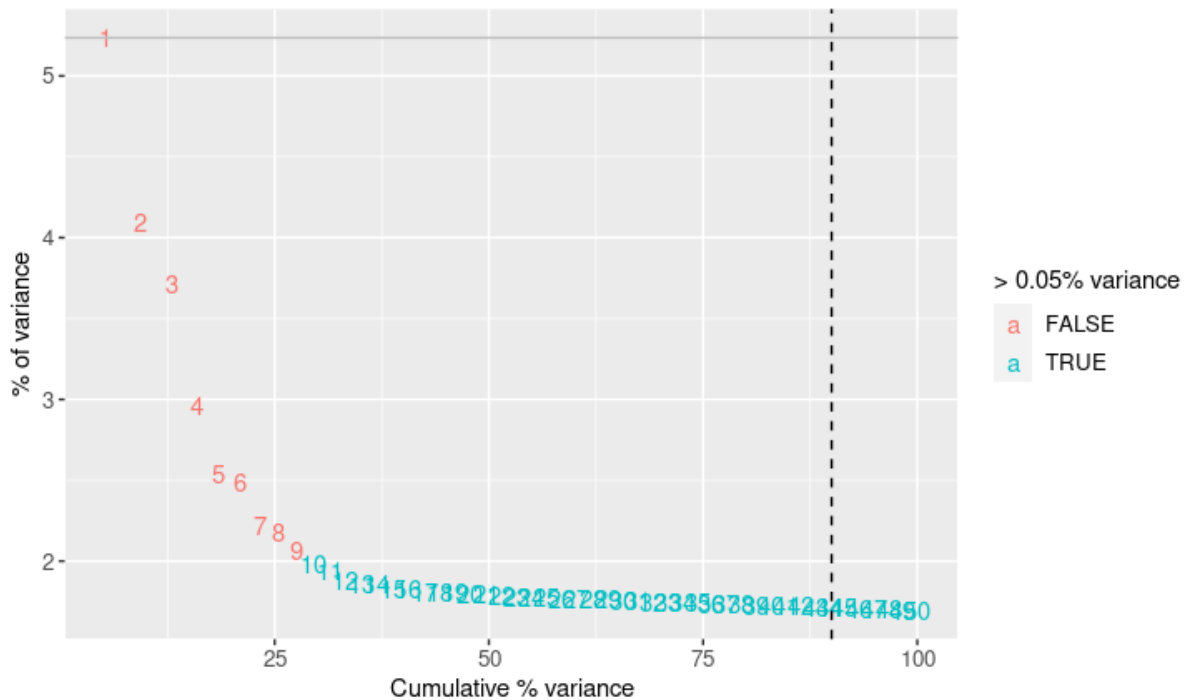


Figure 0-4 Elbow plot of variance found in principle components observed in Rep2_Pro sample.

Elbow plot to rank principle components (PC) found in Rep2_Pro sample assessing the percentage of variance explained by each PC. In the Rep2_Pro sample, an “elbow” is seen around PC 9, suggesting that the majority of variance is captured in the first 9 PCs.

In Figure 0-4 an ‘elbow’ was present around PC 8 -9, which suggests most of the significant variance is captured in the first nine PCs. 9 PCs was therefore chosen for the number of dimensions in the dataset for further downstream analysis.

6.5 Rep2_Axa quality control plots

6.5.1.1 Quality control filtering of single cell RNA-sequencing sample for second axenic amastigote only replicate

After unique reads were mapped and aligned to gene annotations in the custom reference genome and maxi-circle, counts were assigned to cell barcodes within single-cell GEMs using the Cell Ranger count function. Cell Ranger v3.0.2 (<http://software.10xgenomics.com/single-cell/overview/welcome>) was ran with default settings. Dual-species multiplets were identified by Cell Ranger count, as shown in Figure 0-5 A).

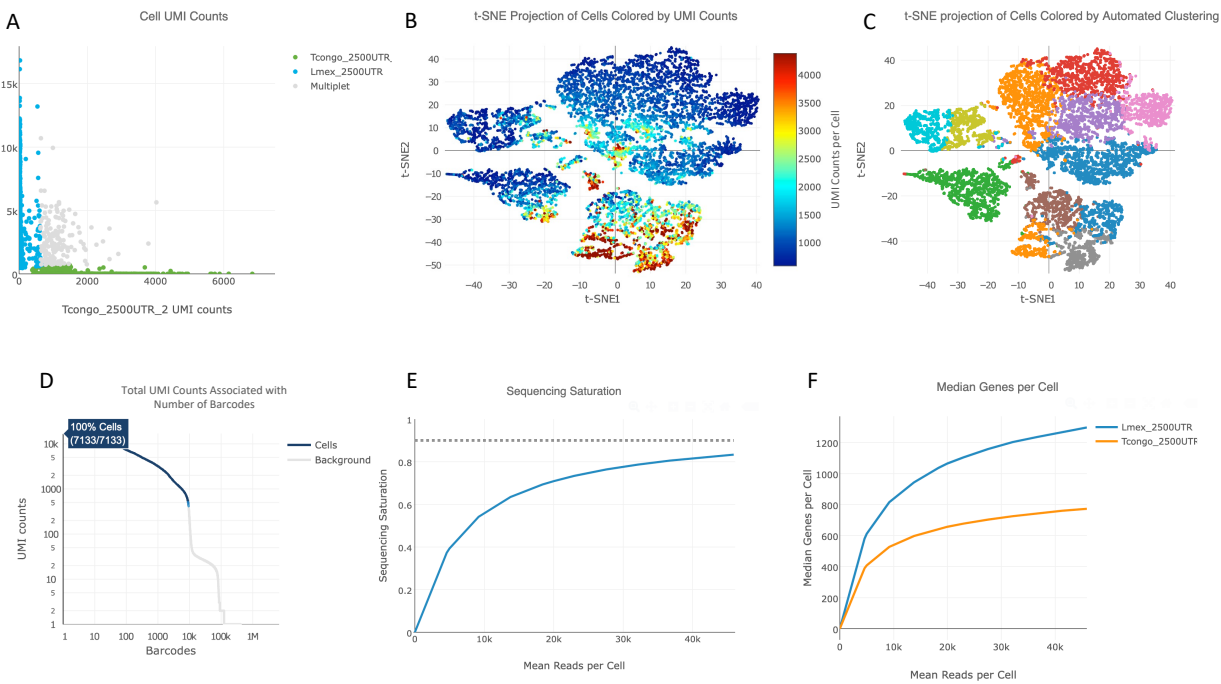


Figure 0-5 Cell Ranger summaries of sequenced Rep2_Axa sample.

Web summary plots of sequenced Rep2_Axa sample containing *Leishmania mexicana* (*L. mexicana*) and *Trypanosoma congo* (*T. congo*). Samples ran by Julie Galbraith and plots generated by Emma Briggs in Cell Ranger (v3.0.2). A) Cell Unique Molecular Identifier (UMI) counts for each sample, *L. mexicana* in green, *T. brucei* in blue and multipllets in grey. B) t-distributed Stochastic Neighbour Embedding (t-SNE) plot of random 10,000 subset of cells in sample coloured by UMI counts. C) t-SNE plot of automated cell clustering in Cell Ranger clustered by similar expression profiles. D) Total UMI counts in mixed sample against barcoded counts of individual transcripts in mixed sample. E) Sequencing saturation of downsampled sequencing depth against mean reads per cell, with reasonably approximated saturation value point indicated by dotted line. F) Plot for median genes per cell against mean reads per cell for each sample mapped to their respective transcriptome for each species. Mapping run by Emma Briggs with Cell Ranger plots generated by Pawel Herzyk of Glasgow Polyomics.

Extensive explanations for each Cell Ranger plot can be found in Section 3.3.2.1.

Briefly, Figure 0-5 A), 9,210 GEMs were present in this sample, with 4% multipllets having more than one GEM, and a mean UMI count purity of 99.0%. In Figure 0-5 B) and C) clustering in t-SNE generated by Cell Ranger formed into a smaller distinct group representative of the *L. mexicana* cells, and three larger groups representative of the *T. congo* sample. In Figure 0-5 D) cells are represented by the blue line on the curve peaking around 9,210 cells, 3,151 of which were *L. mexicana* cells. In Figure 0-5 E) sequencing depth was again chosen as 50,000 mean reads per cell with approximate saturation point indicating suitable saturation reached at 50,000 reads here. In Figure 0-5 F) the median genes per cell for each mapped transcriptome to their relative species are plotted: the *L. mexicana* sample had 1,297, with total genes detected 8,220, and median UMI counts per cell 2,182.

3.3.2.1 In Figure 0-6 below, the quality control filters considered are number of unique features, UMIs, kDNA and rRNA contaminates, as performed for the Rep1 sample.

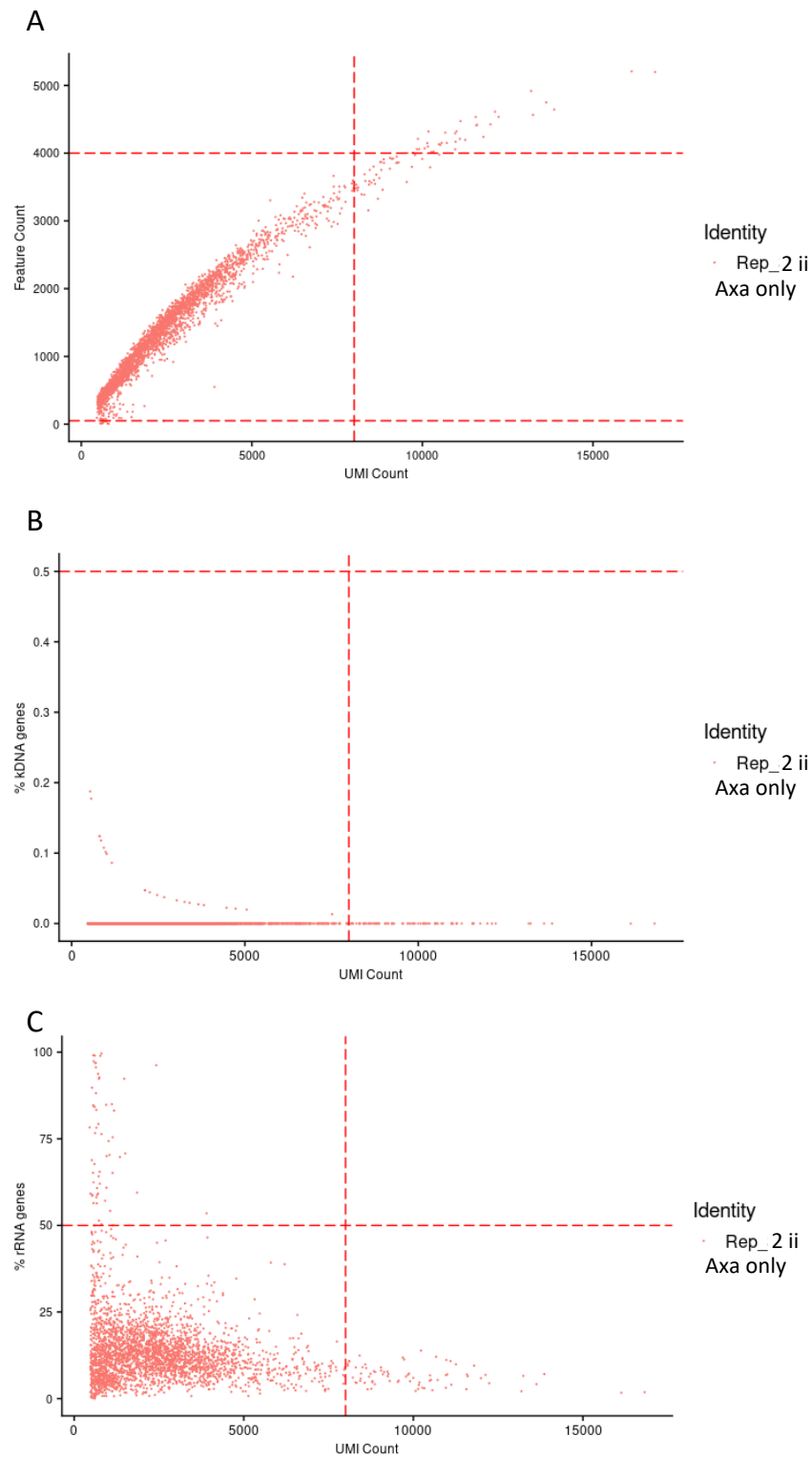


Figure 0-6 Quality control and filtering of transcriptomes in second axenic amastigote replicate.

Scatter plots of quality control and filtering cut-offs. Each dot representing one captured transcriptome. Unique Molecular Identifiers (UMI) are plotted against **A**) gene counts (features) with UMI counts <8000, feature counts >50 and <4000 to exclude for multiplets, **B**) percentage of kDNA maxi circle genome included with mapping reads < 0.5% to exclude lysed or apoptotic cells

and **C**) percentage of ribosomal RNA (rRNA) <50% and >5% contamination in sample. Red dashed lines indicate cut-offs used to filter cells per experiment.

In Figure 0-6 **A**) the number of unique features is set at a cut-off of 4,000 at the upper limit and 50 at the lower, to exclude empty droplets or cell multiplets which display unusually high gene counts. This lower limit is lower than typical to account for possible stationary phase growth typical of some life cycle stages in *Leishmania*. Similarly, the number of UMI for the total number of molecules found within a cell is also set at 8,000 as a cut-off, as for Rep2_Pro. In **B**) kDNA metrics are calculated with the PercentageFeatureSet() function to provide a percentage of counts mapped to the kDNA maxicircle sequence included in mapping and set as a cut-off for cells with more than 0.5% kDNA in their counts, as cells with low-quality or in the process of dying display higher mitochondrial contamination. Similarly, in **C**) rRNA contaminants are set at cells with more than 50% rRNA and then all rRNA transcripts removed to avoid clustering or markers associated with these counts.

As explained for Rep1, scater::logNormCounts was used to normalise feature counts in Rep2_Axa, and normalised values are then stored in the Seurat object under Rep2_Axa[["RNA"]]`@data`.

Next, a subset of highly variable features from cell to cell were selected as a starting point of genes of interest for further examination in downstream analysis using the FindVariableFeatures() function, selecting for the top 2,000 variable features, as in Rep1 and Rep2_Pro. The top ten variable features for the Rep2_Axa sample are labelled below in a scatter plot, Figure 0-7.

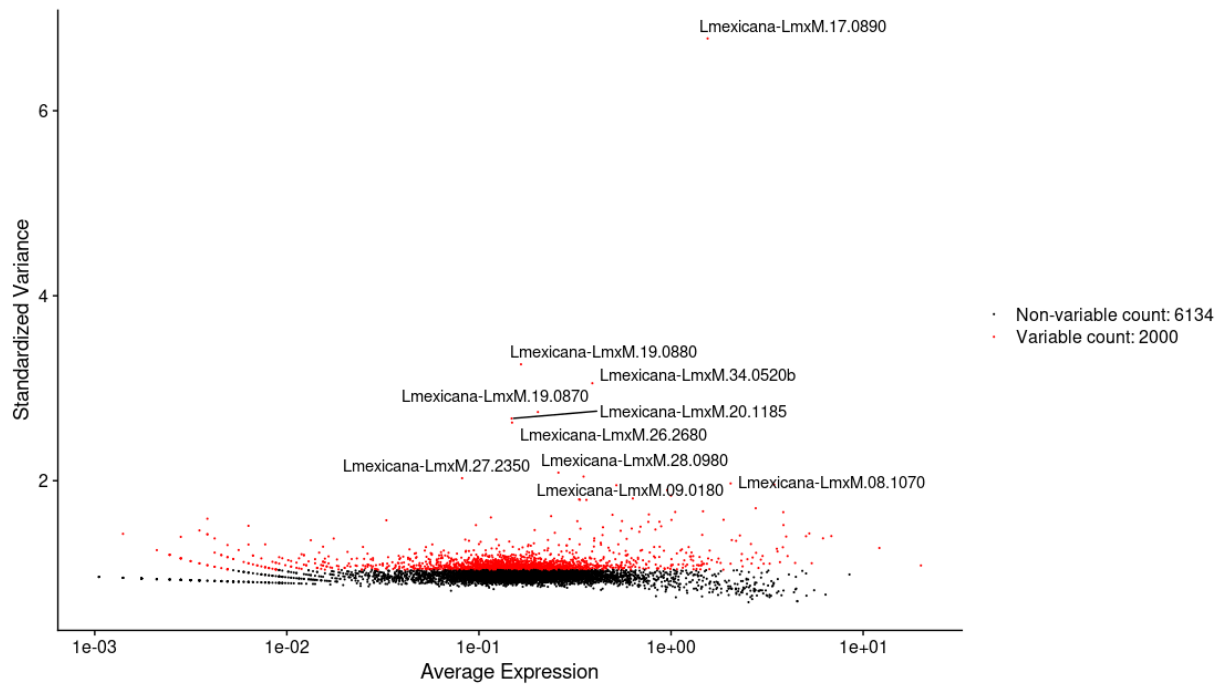


Figure 0-7 Scatter plot of top 2,000 variable features in Rep2_Axa dataset

Scatter plot of top 2,000 variable features (red) selected for downstream analysis in sample Rep2_Axa, plotting average expression of feature against standard variance. Top ten variable features labelled.

Of note are the return of the known marker Cathepsin L protease, LmxM.08.1070, associated with amastigote life cycle stages within *Leishmania*, also previously seen in Rep1. Additionally, the calpain-like cysteine peptidase was also in the top ten variable features. This peptidase has been demonstrated as having roles in amastigote interaction with macrophages, where Marinho et al., (2017) used a calpain inhibitor in mouse models to reduce the number of amastigotes per macrophage in both a time and dose-dependent manner. Interestingly, four promastigote markers were also present, being ATG8/AUT7/APG8/PAZ2, LmxM.09.0180, LmxM.19.0870 and LmxM.19.0880, the latter two also seen in the top ten markers in Rep2_Pro. Additionally the META domain containing protein, LmxM.17.0890. was also present as a marker of the infective metacyclic promastigote stages. This metacyclic markers suggests a population very distinct from the rest of the sample with high variance in these promastigote features. This potentially indicates a promastigote population within this sample, likely from the 24 h timepoint and could indicate incomplete differentiation of promastigotes into axenic amastigotes. Of the remaining four top variable genes, three were hypothetical proteins; LmxM.34.0520b, LmxM.26.2680 and LmxM.28.0980. One of these hypothetical genes (LmxM.34.0520b) is, in fact, a putative pseudogene, and was also present in the

top ten variable genes in both Rep1 and Rep2_Pro. Lastly, the Vesicle-associated membrane protein 7 (VAMP7), LmxM.27.2350, was detected.

Next, linear dimension reduction is performed by PCA on the REP2_Axa data with scaling on the top 2,000 features, as in Rep1 and Rep2_Pro, shown below in Figure 0-8

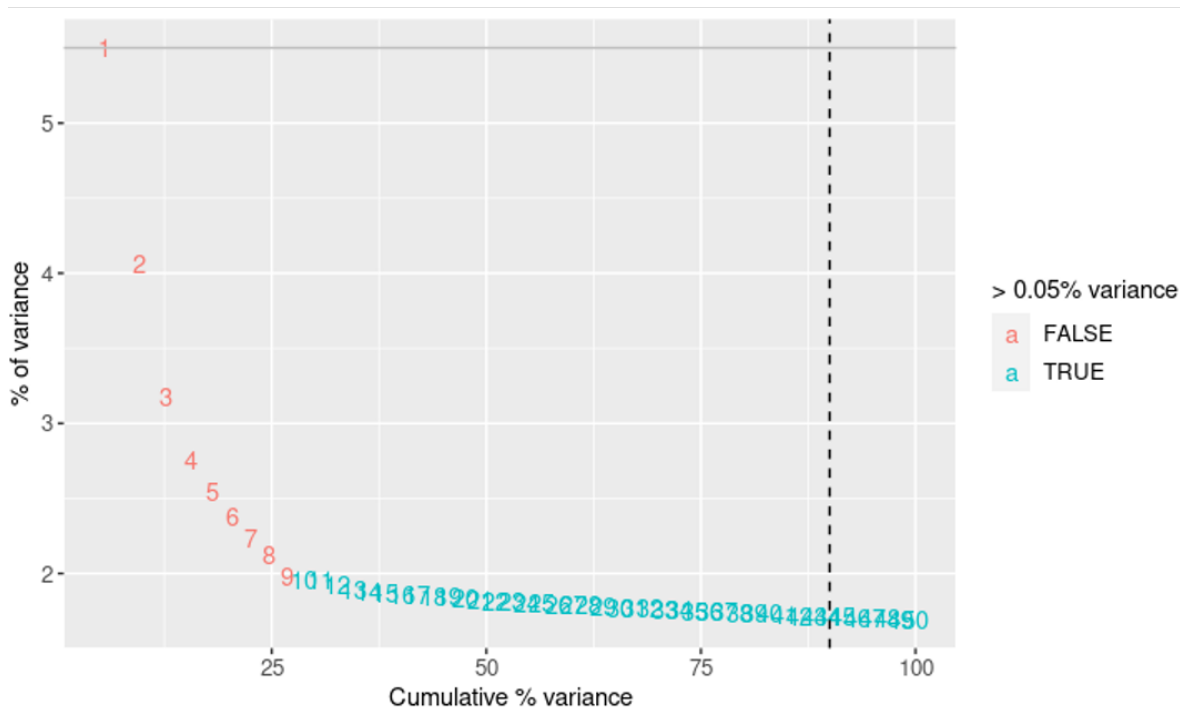


Figure 0-8 Elbow plot of variance found in principle components observed in Rep2_Axa sample.

Elbow plot to rank principle components (PC) found in Rep2_Axa sample assessing the percentage of variance explained by each PC. In the Rep2_Axa sample, an “elbow” is seen around PC 9, suggesting that the majority of variance is captured in the first 9 PCs.

In Figure 0-8 an ‘elbow’ is present around PC 8 -9, which suggests most of the significant variance is captured in the first 9 PCs. 9 PCs is therefore chosen for the number of dimensions in the dataset for further downstream analysis.

6.6 Human macrophage Cell Ranger plots, pre-quality control

Sequencing data was first mapped to the human reference genome GRCh38, combined with the custom *L. mexicana* reference genome as in Chapter 3, by Pawel Herzyk of Glasgow Polyomics, the output of which is shown below in Figure 3-2 for the uninfected control sample and Figure 0-10 for the infected macrophage sample.

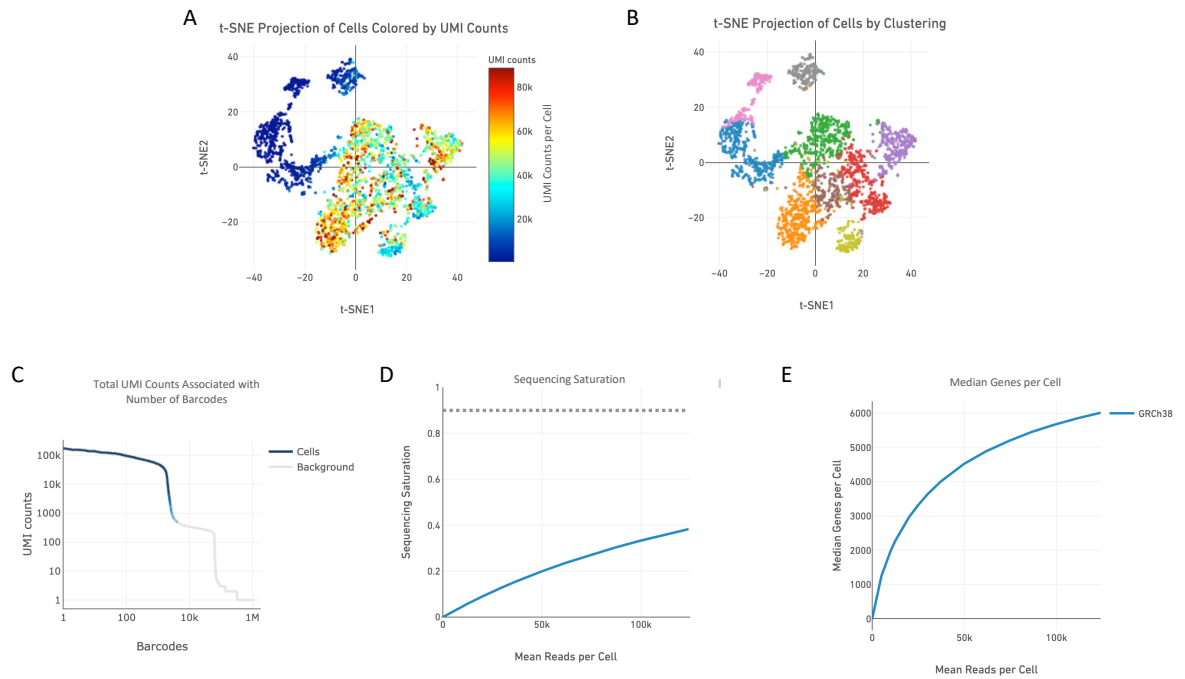


Figure 0-9 Cell Ranger summaries of sequenced uninfected macrophage control sample. Web summary plots of sequenced control sample containing uninfected macrophages. Samples ran by Julie Galbraith and plots generated by Pawel Herzyk of Glasgow Polyomics in Cell Ranger (v6.0.2). **A)** t-distributed Stochastic Neighbour Embedding (t-SNE) plot of cells in sample coloured by Unique Molecular Identifier (UMI) counts. **B)** t-SNE plot of automated cell clustering in Cell Ranger clustered by similar expression profiles. **C)** Total UMI counts in sample against barcoded counts of individual transcripts in sample. **D)** Sequencing saturation of downsampled sequencing depth against mean reads per cell, with reasonably approximated saturation value point indicated by dotted line at 90%. **E)** Plot for median genes per cell against mean reads per cell. Mapping and Cell Ranger plots generated by Pawel Herzyk of Glasgow Polyomics.

Extensive explanations for each Cell Ranger plot can be found in Section 3.3.2.1. Briefly, 2,867 cells were present in this sample, with 96.6% of reads mapped to the genome. In Figure 3-2 **A)** and **B)** clustering in t-SNE generated by Cell Ranger UMI counts per cell and potential clustering within from automated Cell Ranger t-SNE plots, indicative of cells present in the sample for further analysis. In Figure 3-2 **C)** barcodes in the sample were plotted against total UMIs, with the blue line representing cells containing UMI counts, showing 100% of cells having at least 10,000 UMI counts. In Figure 3-2 **D)** the sequencing saturation is shown the number of UMI detected as sequencing depth is increased. Saturation is an indicator of the library complexity and indicates when increased sequencing depth is unlikely to identify further UMIs in the sample. Here, sequencing was run up to 100,000 reads per cells, resulting in 40% of estimated UMIs in the human sample recovered. In Figure 3-2 **E)** the upper boundary for sequence depth in this sample of 100,000 gave median genes per cell of 6,015 with total genes across all cells of 24,180 and median UMI counts per cell of 39,373.

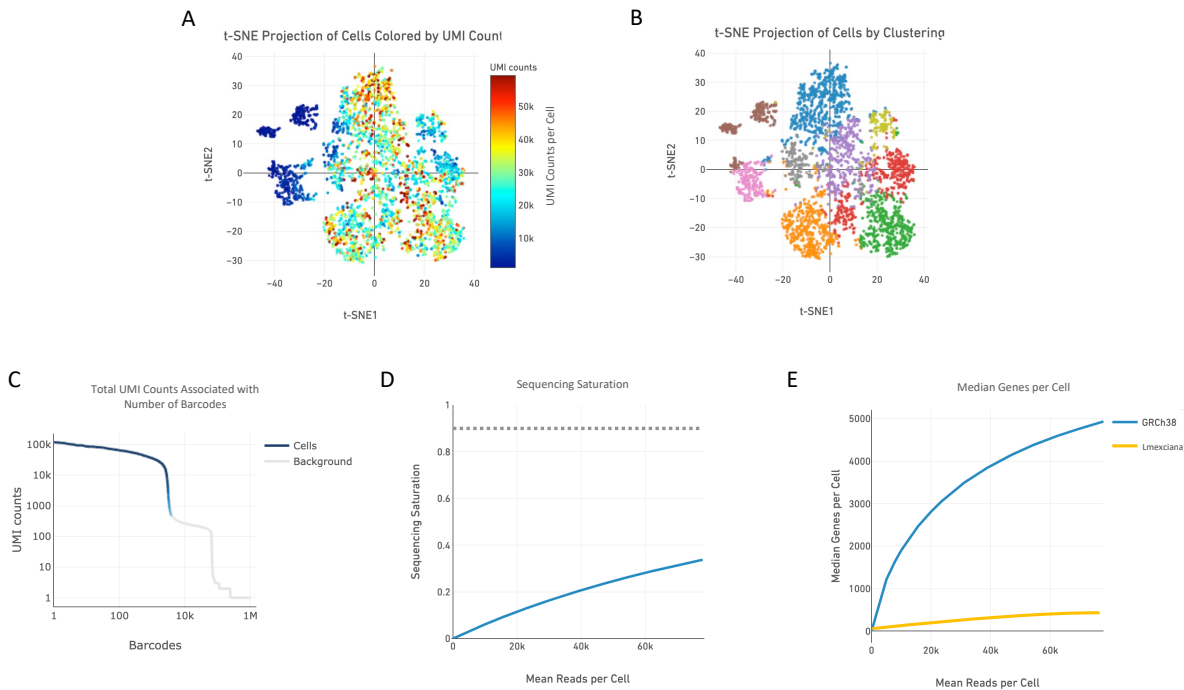


Figure 0-10 Cell Ranger summaries of sequenced infected macrophage sample.

Web summary plots of sequenced control sample containing infected macrophages. Samples ran by Julie Galbraith and plots generated by Pawel Herzyk of Glasgow Polyomics in Cell Ranger (v6.0.2). **A)** t-distributed Stochastic Neighbour Embedding (t-SNE) plot of cells in sample coloured by Unique Molecular Identifier (UMI) counts. **B)** t-SNE plot of automated cell clustering in Cell Ranger clustered by similar expression profiles. **C)** Total UMI counts in sample against barcoded counts of individual transcripts in sample. **D)** Sequencing saturation of downsampled sequencing depth against mean reads per cell, with reasonably approximated saturation value point indicated by dotted line at 90%. **E)** Plot for median genes per cell against mean reads per cell for human reads (blue) and *L. mexicana* (orange). Mapping and Cell Ranger plots generated by Pawel Herzyk of Glasgow Polyomics.

As above, 3,365 cells were present in this sample, with 96.0% of reads mapped to genomes. In Figure 0-10 **A)** and **B)** clustering in t-SNE generated by Cell Ranger UMI counts per cell and potential clustering within from automated Cell Ranger t-SNE plots, indicative of cells present in the sample for further analysis. In Figure 3-2 **C)** barcodes in the sample associated with human reads were plotted against total UMIs, with the blue line representing cells containing UMI counts, again showing 100% of human cells having at least 10,000 UMI counts. In Figure 3-2 **D)** the sequencing saturation for human cells is shown the number of UMI detected as sequencing depth is increased. Saturation is an indicator of the library complexity and indicates when increased sequencing depth is unlikely to identify further UMIs in the sample. Here, sequencing was run up to 100,000 reads per cells and resulted in 40% of estimated UMIs in the human sample recovered. In Figure 3-2 **E)** the upper boundary for sequence depth in the human cells in this sample was 100,000 giving median genes per cell of 4,935 with total genes across human cells of 23,262 and median UMI counts per cell was 26,953.

Appendix II

Hypothetical protein primers.

Sequence Name	Primer Sequence
LmxM.04.0220_DFP	AAAGACGAGGACGGCGAGAATGTGAGCCGCGTTCTGGTAGTGGTTCCGG
LmxM.04.0510_DFP	ACAGCGCGCAAGGCGGTGACGCGATGCTGGGTTCTGGTAGTGGTTCCGG
LmxM.04.0630_DFP	GGCCAGGCCCCGAGTACGCCGAGTTCAAGGGTTCTGGTAGTGGTTCCGG
LmxM.04.0670_DFP	CAGAGAGAGCACCCGATGCAGCCTCCACAGGGTTCTGGTAGTGGTTCCGG
LmxM.04.0720_DFP	TGCAGTGGCGAGTAACCTGATGACGGTTGGTTCTGGTAGTGGTTCCGG
LmxM.04.0740_DFP	GGCAGCTCCTAAGCAGCGACGACAGCGACGGTTCTGGTAGTGGTTCCGG
LmxM.05.1110_DFP	CAAGTGCAGCAGATCCTGCAGCTGCTGAGCGGTTCTGGTAGTGGTTCCGG
LmxM.07.0025_DFP	GAAGGCAGATTCTTCATACACGCGGAAGGGTTCTGGTAGTGGTTCCGG
LmxM.07.0830_DFP	GTGTTCCCCACGGCGGCACCGCGCAGTGGTTCTGGTAGTGGTTCCGG
LmxM.08.29.0940_DFP	GTGATGGAGCTTGAACGTATAGAAAGAAGGGTTCTGGTAGTGGTTCCGG
LmxM.08.0630_DFP	CGTCTGTGGAGGGTGGCAGCGAAACAGTGGTTCTGGTAGTGGTTCCGG
LmxM.08.0810_DFP	CATGAAGACGGTAAATACGTAGAACAGGAGGGTTCTGGTAGTGGTTCCGG
LmxM.08.1222_DFP	AAGCGCCAGGCTGCCGACGACGCCACCCGGGTTCTGGTAGTGGTTCCGG
LmxM.12.0480_DFP	GCGTGATGGCAGCCGTGCGACTACTCCAGGGTTCTGGTAGTGGTTCCGG
LmxM.12.0490_DFP	TGGCGGCGCAGCAGATGGCCGTGTTGACGGTTCTGGTAGTGGTTCCGG
LmxM.13.1070_DFP	ATCCGCGAGCTTCTCCGATGACACACCAGGTTCTGGTAGTGGTTCCGG
LmxM.13.1320_DFP	GGCGAAATTAAGGAGAGCGATGTGAACCTGGGTTCTGGTAGTGGTTCCGG
LmxM.14.0560_DFP	CCGCTATTGGGCGACAGACCTTTGGCGTGGTTCTGGTAGTGGTTCCGG
LmxM.14.0720_DFP	AAGGTGCAGCAAGCGAAGCCCAAGGCGCGGGTTCTGGTAGTGGTTCCGG
LmxM.15.1520_DFP	AGCCACCAGGGCAAGCCGAAGTCTGTGCGGGGTTCTGGTAGTGGTTCCGG
LmxM.16.0500_DFP	AGCAATCTCCAGAATGGCGGCCGTATCCAGGTTCTGGTAGTGGTTCCGG
LmxM.16.1340_DFP	GCATCTGTATCGAAAACCATGGCATTCCAGGTTCTGGTAGTGGTTCCGG
LmxM.16.1370_DFP	GCCGAGAGGGAGACGCCGCGTTGATGGTTGGTTCTGGTAGTGGTTCCGG
LmxM.16.1420_DFP	GACGTTGCCGTGGTGGTGAAGCGACTTCAGGTTCTGGTAGTGGTTCCGG
LmxM.16.1430_DFP	CGCGGAAGACGGTGCAACAGATCACCGAGGGTTCTGGTAGTGGTTCCGG
LmxM.17.0340_DFP	CGCGCTTGCTTACCTTCAACCTGCTTTGGGTTCTGGTAGTGGTTCCGG
LmxM.17.0860_DFP	CATTCAAATCAAGGGTACGATCAGGTCTACGGTTCTGGTAGTGGTTCCGG
LmxM.17.0870_DFP	TTCAGGGTGCTGTACGACGCTCAGCCATTGGTTCTGGTAGTGGTTCCGG
LmxM.18.1350_DFP	GGTGCCGGCAGATGTCGAGTCGTCCTGTTCTGGTAGTGGTTCCGG
LmxM.18.1640_DFP	TACAAGTTCTTCTCGTCCGCGGTGCTCATGGTTCTGGTAGTGGTTCCGG
LmxM.19.1160_DFP	ACTGGCAACAAGTCGCCACGTCGGCGCGGGTTCTGGTAGTGGTTCCGG
LmxM.20.0030_DFP	GGCCCGTTTTCTTGATGTGAAGCGCACAGGTTCTGGTAGTGGTTCCGG
LmxM.20.1220_DFP	CGTCTGAGGCACCCAAAAGCAGCCAATTTGGTTCTGGTAGTGGTTCCGG
LmxM.20.1260_DFP	ACCGGCGCAACCCCTTGACATCTCCTCGGTTCTGGTAGTGGTTCCGG
LmxM.20.1700_DFP	CTGACAGCAGCACCTTCGGCTCAGTACAACGGTTCTGGTAGTGGTTCCGG
LmxM.21.1480_DFP	GTGAGCGGTGGCAACAGCATCTCTAGTCGGTTCTGGTAGTGGTTCCGG
LmxM.21.1555_DFP	GCGACGGTAAGGCACCGCAGGTGATGGCGGGTTCTGGTAGTGGTTCCGG
LmxM.22.0100_DFP	TTTCTCCGCAACTCCTCGTGCCCCCTCAGGTTCTGGTAGTGGTTCCGG
LmxM.22.0240_DFP	CAGTTGTCGATGATGCGGTCTCGCCGCGAGGTTCTGGTAGTGGTTCCGG
LmxM.23.0080_DFP	ACGGTGCTGCGGAATCGACTTGAGTTGTTGGGTTCTGGTAGTGGTTCCGG

LmxM.23.0610_DFP	TCTCCGAGGCATCCCCTCGTCGGTAAAGGGTTCTGGTAGTGGTTCCGG
LmxM.23.1290_DFP	TTCTTTATCTTATCGAGCTCGTCGACGCGTTCTGGTAGTGGTTCCGG
LmxM.24.0350_DFP	CGGTACCGCGATAGTCGCGATGCACGCGGCGTTCTGGTAGTGGTTCCGG
LmxM.24.1410_DFP	ACAGATGCGTGGCCAACGAAGCTCTTGCTAGTTCTGGTAGTGGTTCCGG
LmxM.24.1550_DFP	GGTGAGGGGGCAGGGGCAGCCCCATCGACGTTCTGGTAGTGGTTCCGG
LmxM.24.1560_DFP	TTGGAGCCGTTCAATGACGTGGTAGCGTACGGTTCTGGTAGTGGTTCCGG
LmxM.25.0590_DFP	GTGTACAAGGACTTGGAGAAGGCCATGCACGTTCTGGTAGTGGTTCCGG
LmxM.25.0715_DFP	GAGAAGTCACAGAAGAAAGTGACCTCGAAGGGTTCTGGTAGTGGTTCCGG
LmxM.25.1620_DFP	CACAAGAGCGGTGCGGCTGGTGCCACCACGGTTCTGGTAGTGGTTCCGG
LmxM.25.2375_DFP	CTGTTTACCGTGAAATCGTTCTGAGCGAGGGTTCTGGTAGTGGTTCCGG
LmxM.26.2680_DFP	TTTGGTGACCGCATTTATCTGCCTGTATTCGGTTCTGGTAGTGGTTCCGG
LmxM.27.0660_DFP	CGCCGACCTCGAGCGTGATCTTCATTGACGGTTCTGGTAGTGGTTCCGG
LmxM.27.1080_DFP	TATAGGCGAATCGAAGAGCTGAACCGCAAGGGTTCTGGTAGTGGTTCCGG
LmxM.28.0570_DFP	GCAGTGCTTCTGGTGGTGCTTTTCATGGCCGGTTCTGGTAGTGGTTCCGG
LmxM.28.0980_DFP	CCGGGTCCGCTGATGTCGGCCACGGCGTGGGTTCTGGTAGTGGTTCCGG
LmxM.29.0770_DFP	GGGTTCTAGAGCGAGCCGGCACCGGTAGAGTTCTGGTAGTGGTTCCGG
LmxM.29.2330_DFP	TCCATGCAGGCTGCCGAGGACCCCTACGTGGTTCTGGTAGTGGTTCCGG
LmxM.29.2850_DFP	GAGACGGTAAAGCTCTTCTTTACGCGCGGGTTCTGGTAGTGGTTCCGG
LmxM.29.3140_DFP	AAGTGCAGGCGAGCGGAAGTCTCTTCTGGGTTCTGGTAGTGGTTCCGG
LmxM.30.0080_DFP	GAGGAAATCGAGGGCGCGAAAAGAAAGAGGGTTCTGGTAGTGGTTCCGG
LmxM.30.0710_DFP	GTGCTCACTCCGCCCACTTCGGGTCTGGGTTCTGGTAGTGGTTCCGG
LmxM.30.0760_DFP	GCGGCGAGCGTGCCACCGGCAATGTCAGCGGGTTCTGGTAGTGGTTCCGG
LmxM.30.0800_DFP	GCGGCCAAGGCAATGATCGGTCGATTTTTGTTCTGGTAGTGGTTCCGG
LmxM.30.0900_DFP	GGAAACGGCAAGAGCACCTTGAGAACATCGGTTCTGGTAGTGGTTCCGG
LmxM.30.0970_DFP	CTGCGACTCGCCAGGAGACCGCCGACGCGGGTTCTGGTAGTGGTTCCGG
LmxM.30.1050_DFP	GAGCAGCCCGAGGAGGATGCGACTGAGCTTGGTTCTGGTAGTGGTTCCGG
LmxM.30.1090_DFP	AAGAAGTTGTTTACGTGAAGCCTCTGGAGGGTTCTGGTAGTGGTTCCGG
LmxM.30.1380_DFP	TCACACAGAAACGTGACTCGAGACAAGGGTTCTGGTAGTGGTTCCGG
LmxM.30.1600_DFP	CGCATGGCGAAGGACGTGACCTTGCTGACGGTTCTGGTAGTGGTTCCGG
LmxM.30.2100_DFP	GACGGCGAGAAACCGGAAACCGCGATCAGGGTTCTGGTAGTGGTTCCGG
LmxM.30.2110_DFP	TCCGAAAATGCGGATTGGCGAAGGCGTAACGGTTCTGGTAGTGGTTCCGG
LmxM.30.2310_DFP	CGCTTTGTTTACTACGAGCCGTATCCCTGGGTTCTGGTAGTGGTTCCGG
LmxM.30.2450_DFP	CGCGCACAGTGGCAACAGATACGGCCCAAGGGTTCTGGTAGTGGTTCCGG
LmxM.31.0270_DFP	CGGTCGAGGATGACGAGGTCGAGAGTTGGGTTCTGGTAGTGGTTCCGG
LmxM.31.0350_DFP	GCCAGCCGCTCGGCCAAGGCGCCCTCTGCGGGTTCTGGTAGTGGTTCCGG
LmxM.31.0840_DFP	AAGGCGACTGAGGAGCGGGCGAACGAGCAGGGTTCTGGTAGTGGTTCCGG
LmxM.31.3610_DFP	CCCGCACAGAGAAGTACATCACGGAGATGGGTTCTGGTAGTGGTTCCGG
LmxM.32.1070_DFP	AGCTCAGCGACGAGCGCCATGATAGCGTGGGTTCTGGTAGTGGTTCCGG
LmxM.32.2890_DFP	CCGTTGGCGGCCACCCCACTTGCGGGCAGGTTCTGGTAGTGGTTCCGG
LmxM.33.0250_DFP	CCTCAGTCTACCAGCCTACTTCCCTTTGGGGTTCTGGTAGTGGTTCCGG
LmxM.33.2560_DFP	TCATCGGAGAGAAGTATACCGTATTCATGTGGTTCTGGTAGTGGTTCCGG
LmxM.33.3645_DFP	TCGGAGGATGATTCATCTGCGGACATGGTGGGTTCTGGTAGTGGTTCCGG
LmxM.34.0230_DFP	TCAAGCTGCATAATGGAGAGAGCATTTGGGGTTCTGGTAGTGGTTCCGG
LmxM.34.0270_DFP	CCTAAGGCGTCGTGCAAAGGCACTAAAAGGGTTCTGGTAGTGGTTCCGG
LmxM.34.0400a_DFP	GCTGCACCGGCTGCGGAGAAGCGCCGAGGGTTCTGGTAGTGGTTCCGG
LmxM.34.0520b_DFP	ATGCGCAAGCAGCTGTTGGCGGTGAGGAAGGGTTCTGGTAGTGGTTCCGG

LmxM.34.1460_DFP	GCAGACCCCTCGAGCAGCCGACCGTGGCGGGTTCTGGTAGTGGTTCCGG
LmxM.34.4190_DFP	GCAAGCGAGGCCGAGAGTAAGAAGGAACGCGTTCTGGTAGTGGTTCCGG
LmxM.34.4380_DFP	CAGAGTGTCAAGCTCGGCAGCGACAGTGACGGTTCTGGTAGTGGTTCCGG
LmxM.36.0480_DFP	CACGAGTTCGGCGTCTACCTCAATGCTGAGGGTTCTGGTAGTGGTTCCGG
LmxM.36.1380_DFP	CTGCCAGCACTCATAGCAAGCACACCCGCGTTCTGGTAGTGGTTCCGG
LmxM.36.2770_DFP	GAAGCGGACGGATCTTCGGTGTGGTGCCGGTTCTGGTAGTGGTTCCGG
LmxM.36.4040_DFP	AAGAGGGTGCCTGCAGAATCGGCGTCTGCAGGTTCTGGTAGTGGTTCCGG
LmxM.36.5000_DFP	GGTCGACGACGCGCTGGGTTGTGGTGGTGGTTCTGGTAGTGGTTCCGG
LmxM.36.5100_DFP	ACGAGTGCGCCGAGCTCTCCAACGCTGAGGGTTCTGGTAGTGGTTCCGG
LmxM.36.5480_DFP	GTGCTGCCCTCAAACGCTACTTCTACTCGGGTTCTGGTAGTGGTTCCGG
LmxM.04.0220_DRP	CTTGCCCCATCCCTCCCCATCCCCCAATTTGAGAGACCTGTGC
LmxM.04.0510_DRP	CAGGCTTTCGCGTGTGGGTGCGATACCCTCCAATTTGAGAGACCTGTGC
LmxM.04.0630_DRP	TAGTATCCTAAGTGCTAAGATTTCTCCACTCCAATTTGAGAGACCTGTGC
LmxM.04.0670_DRP	CACCCACCACCATCACACAAGGTTTCCCTCCAATTTGAGAGACCTGTGC
LmxM.04.0720_DRP	AGCGACTGCAAGAGGGATCGGTGCCCCCCCAATTTGAGAGACCTGTGC
LmxM.04.0740_DRP	ACACGCCTTGTCTGGTCCCCCGTCCCCCAATTTGAGAGACCTGTGC
LmxM.05.1110_DRP	GCCCCAACCCGCAACACAGCCACAGCCCAATTTGAGAGACCTGTGC
LmxM.07.0025_DRP	AAAGATTTTACGAGCAGGCAGAGGCATGCCAATTTGAGAGACCTGTGC
LmxM.07.0830_DRP	GTGTGTGTGTGGGTGTGTGTACACCTCCAATTTGAGAGACCTGTGC
LmxM.08.29.0940_DRP	GACGTGTGTCAACGTGGCGACGCTGCTCCGCCAATTTGAGAGACCTGTGC
LmxM.08.0630_DRP	TATGCGTGTGTGTGTGTGGCCCTACCGCCAATTTGAGAGACCTGTGC
LmxM.08.0810_DRP	CCGCACAGCTATCAGCATCGCCCTCCGCTCCAATTTGAGAGACCTGTGC
LmxM.08.1222_DRP	TGCCGACGAGAGGGGCAAGAAATGCGCCCAATTTGAGAGACCTGTGC
LmxM.12.0480_DRP	GAGAGCACACGAGGAGTCCCGTGTCTCCCTCCAATTTGAGAGACCTGTGC
LmxM.12.0490_DRP	TGTTTGTGGATATGGCTATGTGGCACACCCAATTTGAGAGACCTGTGC
LmxM.13.1070_DRP	AACTACGTACGATGCAACTGCCTCAGTCTCCAATTTGAGAGACCTGTGC
LmxM.13.1320_DRP	CATGCGCTCGCTGCGTGTGTGCGTGTGCCTCCAATTTGAGAGACCTGTGC
LmxM.14.0560_DRP	AGGGTGCCCGACACGACCTCTCCACCCCAATTTGAGAGACCTGTGC
LmxM.14.0720_DRP	ACACGTGCTCCACACAGGCCGAGCGTCTCCAATTTGAGAGACCTGTGC
LmxM.15.1520_DRP	CAAGCACTCCTGTTCTGCTATCAGCACCAATTTGAGAGACCTGTGC
LmxM.16.0500_DRP	AAGTGGCAGAAGAAGTTACGCTTATGTTGCCCAATTTGAGAGACCTGTGC
LmxM.16.1340_DRP	TGCCCCATGCGGAGAGGATATAGGGAGGGGCAATTTGAGAGACCTGTGC
LmxM.16.1370_DRP	TTCTCCCCCTTACCCTTCTCCGTTCCGCCAATTTGAGAGACCTGTGC
LmxM.16.1420_DRP	AAAACGCATTGACCGCAAGACAACGAAGGCCAATTTGAGAGACCTGTGC
LmxM.16.1430_DRP	GAACCTCCAAAAGATGATACAGGCGACCAATTTGAGAGACCTGTGC
LmxM.17.0340_DRP	GCTTGTGGGAAAAAGGAGTCTTCTGGTCCAATTTGAGAGACCTGTGC
LmxM.17.0860_DRP	AAAAAAAAGTATGAATGGGTCTTTTTCCGCCAATTTGAGAGACCTGTGC
LmxM.17.0870_DRP	TAATTCAAATCCATATCACACAGTTTTTCCCAATTTGAGAGACCTGTGC
LmxM.18.1350_DRP	CGAGTCACGCGTGCCTCGTGAAGCCGCCAATTTGAGAGACCTGTGC
LmxM.18.1640_DRP	CCGCAGATCCGTGTGAGACATACGCATTACCAATTTGAGAGACCTGTGC
LmxM.19.1160_DRP	TCCTATTCTATCACGCTGCTCCGCGCCACCAATTTGAGAGACCTGTGC
LmxM.20.0030_DRP	GAACAGCCGGAAGGGTAGCTGCACCAGCAACCAATTTGAGAGACCTGTGC
LmxM.20.1220_DRP	CGTGAAGAGACAGTGTGTGGAGAAGGAGGCCAATTTGAGAGACCTGTGC
LmxM.20.1260_DRP	GAAGAGACTGCACAAGCTCACCCGATCCGCCAATTTGAGAGACCTGTGC
LmxM.20.1700_DRP	GATAAGGAGCTGCGTGTGATTCAGGCATCCAATTTGAGAGACCTGTGC
LmxM.21.1480_DRP	CACCGCTCCCGATGGCACGGGCAGCACCCCAATTTGAGAGACCTGTGC

LmxM.21.1555_DRP	CTGAGGCCGACTACACACGCGCATAAACCTCCAATTTGAGAGACCTGTGC
LmxM.22.0100_DRP	CTGCTTGGACTTCTTCCGTAGTGCTAATCACCAATTTGAGAGACCTGTGC
LmxM.22.0240_DRP	AAATGAAAAAGGAATAAGAGATCAAACCACCAATTTGAGAGACCTGTGC
LmxM.23.0080_DRP	CACCACCACGAAAGACAAAACATTACATAGCCAATTTGAGAGACCTGTGC
LmxM.23.0610_DRP	TTTCGCTTCGACAGGTTTCTTCGTCGCCCCACCAATTTGAGAGACCTGTGC
LmxM.23.1290_DRP	CACAGGCCACGACAACCAACATACCCCGCCAATTTGAGAGACCTGTGC
LmxM.24.0350_DRP	GAGGCGGAGTCGCTAAGACACGCCAGCCTCCAATTTGAGAGACCTGTGC
LmxM.24.1410_DRP	GGGGAAGTGATGCGGGCGAGAAGCAAGCTGCCAATTTGAGAGACCTGTGC
LmxM.24.1550_DRP	GCACTGCTCACCAAGTAAAGTGCCTGCACCAATTTGAGAGACCTGTGC
LmxM.24.1560_DRP	CGCGCGTCGTAGATCTTAGTCTGTGCCTCCAATTTGAGAGACCTGTGC
LmxM.25.0590_DRP	CTACTCTCTACACCATCAGCGGCGTGCCACCAATTTGAGAGACCTGTGC
LmxM.25.0715_DRP	ACCGGTACAGTTTTTTTTCTTTTTGGCAAACCAATTTGAGAGACCTGTGC
LmxM.25.1620_DRP	GGCTACACCCACTACCGTGTGTGCTTTCCGCCAATTTGAGAGACCTGTGC
LmxM.25.2375_DRP	CCTTCCCCACCTTCGTCCAATTCCTTTCAACCAATTTGAGAGACCTGTGC
LmxM.26.2680_DRP	GCGTACCTTCCACCCTGTACATACCAACACCCCAATTTGAGAGACCTGTGC
LmxM.27.0660_DRP	GTTAGTTCCTTACCTCACCAACTTTCCCCAATTTGAGAGACCTGTGC
LmxM.27.1080_DRP	AAAAGAAACCGAAGCACGCGCAGGCCACCCCAATTTGAGAGACCTGTGC
LmxM.28.0570_DRP	CACACAGACACAACAGGCGCATGAATGCCCAATTTGAGAGACCTGTGC
LmxM.28.0980_DRP	TGTGGGGGGGGGAAGGGGGACACGTGCCCCAATTTGAGAGACCTGTGC
LmxM.29.0770_DRP	TTTGGCCTGTGTTTCAGAGTATCGTAACGCCAATTTGAGAGACCTGTGC
LmxM.29.2330_DRP	ACACGCACCTCACAATGCTTCACTGTCCCGCCAATTTGAGAGACCTGTGC
LmxM.29.2850_DRP	GGCGTTGACCCTAATCAATGAAAAAGAAAACCAATTTGAGAGACCTGTGC
LmxM.29.3140_DRP	TCATGAAGCATCAAGGGAGGAGGGAAGAGACCAATTTGAGAGACCTGTGC
LmxM.30.0080_DRP	AGAATGAGAGGAAATCACCTCGGCTACCCCAATTTGAGAGACCTGTGC
LmxM.30.0710_DRP	TCCTCCACGTCTGTGATGCCGCGCTGCTTCCAATTTGAGAGACCTGTGC
LmxM.30.0760_DRP	TTTCTCGAACACGCCAAGGAAGCAACCCCAATTTGAGAGACCTGTGC
LmxM.30.0800_DRP	GAAGCGCAAGCACAAAAAAACAATATAGACCAATTTGAGAGACCTGTGC
LmxM.30.0900_DRP	GCGTGCCGCTCGGTGCAATCGCATTCTCCAATTTGAGAGACCTGTGC
LmxM.30.0970_DRP	AGACACGCAGCCCTTCCCAAGACACCCTCCAATTTGAGAGACCTGTGC
LmxM.30.1050_DRP	AACCGGTACTCTTGCTCTTACGTACAGATACCAATTTGAGAGACCTGTGC
LmxM.30.1090_DRP	GAATGTTCACTTCTTTCACGGCGTGCCGCAATTTGAGAGACCTGTGC
LmxM.30.1380_DRP	AGAGCGCCATCGTCATCCTATCTGGCTCCCAATTTGAGAGACCTGTGC
LmxM.30.1600_DRP	AACGGAAGTCAAAAAAAATAAGCACAGGCCAATTTGAGAGACCTGTGC
LmxM.30.2100_DRP	TCCTCTTACCCTAGACCCACCACGGCCCAATTTGAGAGACCTGTGC
LmxM.30.2110_DRP	TACTTGCCATCTGACAGGTTCTGCGGCCACCAATTTGAGAGACCTGTGC
LmxM.30.2310_DRP	AGACTTGGCTTTCGTCTGCCACAGGCACCGCCAATTTGAGAGACCTGTGC
LmxM.30.2450_DRP	GATCAGACGCCACGTCACTTTGTTGCTCCAATTTGAGAGACCTGTGC
LmxM.31.0270_DRP	GCCCGCGTGTGCTGCTGTTGTGGCGCACACCAATTTGAGAGACCTGTGC
LmxM.31.0350_DRP	TGGGAAAAGCAGGTACGCTCAGAGTCTATCCAATTTGAGAGACCTGTGC
LmxM.31.0840_DRP	AGGAGAAAAACCAACCATATATGCACGTACCAATTTGAGAGACCTGTGC
LmxM.31.3610_DRP	GCGAGCACAAGATCAGCCGTGCGCAAGTCCCAATTTGAGAGACCTGTGC
LmxM.32.1070_DRP	CCATCTGTTGCTCCTCTGGTGCTCCTCCCTCCAATTTGAGAGACCTGTGC
LmxM.32.2890_DRP	ATAACGTAGGGTCAAGGCACACACAGGCCCAATTTGAGAGACCTGTGC
LmxM.33.0250_DRP	GCGGCACAGGCGTGACCGTCGCAGCACCGCCAATTTGAGAGACCTGTGC
LmxM.33.2560_DRP	ATCCTGAATAGACTGCCCTGAACGACCTCCAATTTGAGAGACCTGTGC
LmxM.33.3645_DRP	CGAATCCTCAGAGGCGAGAGCGACTTCCCGCCAATTTGAGAGACCTGTGC

LmxM.34.0230_DRP	AGGAAGGGGGGCATCTATACTGAAGCACCGCCAATTTGAGAGACCTGTGC
LmxM.34.0270_DRP	CCGCATTCGACTCGTCCACAGGGCGCGTGCCAATTTGAGAGACCTGTGC
LmxM.34.0400a_DRP	CCCACCCGACAAAAGACCAGCCACGCTGAGCCAATTTGAGAGACCTGTGC
LmxM.34.0520b_DRP	ACAGACACGCACACCCATCCGCACACACCACCAATTTGAGAGACCTGTGC
LmxM.34.1460_DRP	ACACCCGCATATACATACGCACACGTCCTCCAATTTGAGAGACCTGTGC
LmxM.34.4190_DRP	TCCACGACCAAAGCACGGCATGCAGTGGTTCCAATTTGAGAGACCTGTGC
LmxM.34.4380_DRP	CGCCCCAATACTTTCTGTATCTACGCATCCCCAATTTGAGAGACCTGTGC
LmxM.36.0480_DRP	ATACACCGACACAGACACACACATGCAACCAATTTGAGAGACCTGTGC
LmxM.36.1380_DRP	AGGCATCAAAGTGGGTATAGACACCACCACCAATTTGAGAGACCTGTGC
LmxM.36.2770_DRP	TCCACCTGCTCTCTTTCCCTCGAAACCGCCAATTTGAGAGACCTGTGC
LmxM.36.4040_DRP	CCCACGCCCTTCTCGCCACCCACGCACCCCAATTTGAGAGACCTGTGC
LmxM.36.5000_DRP	ACTGCATCCCCTCGCCACCCCTCTTGCCCAATTTGAGAGACCTGTGC
LmxM.36.5100_DRP	CGGGGGGAGGGGACGCAACGCGCTTTAGTTCCAATTTGAGAGACCTGTGC
LmxM.36.5480_DRP	CAAACAACAACAAAAAAGAAGAACCACCACCAATTTGAGAGACCTGTGC
LmxM.04.0220_gRNA	GAAATTAATACGACTCACTATAGGCGACAGCAAAGAAACAGCGAGTTTTAGAGCTAGAAATAGC
LmxM.04.0510_gRNA	GAAATTAATACGACTCACTATAGGTATCTGTTGCATCGTCTCCGTTTTAGAGCTAGAAATAGC
LmxM.04.0630_gRNA	GAAATTAATACGACTCACTATAGGTTGACAAACACACAAAAAGTTTTAGAGCTAGAAATAGC
LmxM.04.0670_gRNA	GAAATTAATACGACTCACTATAGGGCATCATGTAGGCAACTGTGGTTTTAGAGCTAGAAATAGC
LmxM.04.0720_gRNA	GAAATTAATACGACTCACTATAGGTGAGTGACGCGGGTGGAGGGTTTTAGAGCTAGAAATAGC
LmxM.04.0740_gRNA	GAAATTAATACGACTCACTATAGGGCCACGGCGAGCTCAGATGGTTTTAGAGCTAGAAATAGC
LmxM.05.1110_gRNA	GAAATTAATACGACTCACTATAGGGCGATGGCAAGCGCGAGATGGTTTTAGAGCTAGAAATAGC
LmxM.07.0025_gRNA	GAAATTAATACGACTCACTATAGGACGTGCAGTACGTTAACAGCGTTTTAGAGCTAGAAATAGC
LmxM.07.0830_gRNA	GAAATTAATACGACTCACTATAGGGAAGAGACACAGAATTGGTGGTTTTAGAGCTAGAAATAGC
LmxM.08.29.0940_gRNA	GAAATTAATACGACTCACTATAGGAGAGTCTCTGTGATGCCGAGTTTTAGAGCTAGAAATAGC
LmxM.08.0630_gRNA	GAAATTAATACGACTCACTATAGGGGGCAACGACAACGGGAGCGGTTTTAGAGCTAGAAATAGC
LmxM.08.0810_gRNA	GAAATTAATACGACTCACTATAGGAGGCAAGTAAAAGCGGAAGGGTTTTAGAGCTAGAAATAGC
LmxM.08.1222_gRNA	GAAATTAATACGACTCACTATAGGGACTGTGTGGGTGGGAGATGGTTTTAGAGCTAGAAATAGC
LmxM.12.0480_gRNA	GAAATTAATACGACTCACTATAGGAATGTGGGTGCGTGAAGTGGTTTTAGAGCTAGAAATAGC
LmxM.12.0490_gRNA	GAAATTAATACGACTCACTATAGGACACACACAAACGTGCACCAGTTTTAGAGCTAGAAATAGC
LmxM.13.1070_gRNA	GAAATTAATACGACTCACTATAGGAGCTGTAAAAGCACCTTGCCGTTTTAGAGCTAGAAATAGC
LmxM.13.1320_gRNA	GAAATTAATACGACTCACTATAGGGGGACGAACAACAAGGTAAGTTTTAGAGCTAGAAATAGC
LmxM.14.0560_gRNA	GAAATTAATACGACTCACTATAGGCCCAAGTGGATGACAGAAGTTTTAGAGCTAGAAATAGC
LmxM.14.0720_gRNA	GAAATTAATACGACTCACTATAGGAGAGCACAGCCCCAAGCGCGTTTTAGAGCTAGAAATAGC
LmxM.15.1520_gRNA	GAAATTAATACGACTCACTATAGGCAAGAAAAAAGAAACGTGTAGTTTTAGAGCTAGAAATAGC
LmxM.16.0500_gRNA	GAAATTAATACGACTCACTATAGGTGCTCTTCATGAGCTGTGAGTTTTAGAGCTAGAAATAGC
LmxM.16.1340_gRNA	GAAATTAATACGACTCACTATAGGGAACGCCAACCCAGCATGCAGTTTTAGAGCTAGAAATAGC
LmxM.16.1370_gRNA	GAAATTAATACGACTCACTATAGGGGTACATGGATCAAGCAGGGTTTTAGAGCTAGAAATAGC
LmxM.16.1420_gRNA	GAAATTAATACGACTCACTATAGGGCGGATGCGTCACCCCTCTCGTTTTAGAGCTAGAAATAGC
LmxM.16.1430_gRNA	GAAATTAATACGACTCACTATAGGTAGGGCGGCGCCGCTGGAAAGTTTTAGAGCTAGAAATAGC
LmxM.17.0340_gRNA	GAAATTAATACGACTCACTATAGGTGATTGCGCGAAGAAGAAAGTTTTAGAGCTAGAAATAGC
LmxM.17.0860_gRNA	GAAATTAATACGACTCACTATAGGTAGCGGACGGAGTGCCTCAGTTTTAGAGCTAGAAATAGC
LmxM.17.0870_gRNA	GAAATTAATACGACTCACTATAGGTCTGAAAAATAGGAAGGAATGTTTTAGAGCTAGAAATAGC
LmxM.18.1350_gRNA	GAAATTAATACGACTCACTATAGGGGTACCCGACGCGCAACAGTTTTAGAGCTAGAAATAGC
LmxM.18.1640_gRNA	GAAATTAATACGACTCACTATAGGACACAAGCGGCTTTAGCGGTTTTAGAGCTAGAAATAGC
LmxM.19.1160_gRNA	GAAATTAATACGACTCACTATAGGCTCGTCGCTGCCACCCAGTTTTAGAGCTAGAAATAGC
LmxM.20.0030_gRNA	GAAATTAATACGACTCACTATAGGCAACACTCGAAAATGATGTCGTTTTAGAGCTAGAAATAGC

LmxM.20.1220_gRNA	GAAATTAATACGACTCACTATAGGTGACTCCACCACATCTCCCGTTTTAGAGCTAGAAATAGC
LmxM.20.1260_gRNA	GAAATTAATACGACTCACTATAGGGAGGGTGTCCGGAGACATGTGTTTTAGAGCTAGAAATAGC
LmxM.20.1700_gRNA	GAAATTAATACGACTCACTATAGGGCATCATCATCGTTGCTTATGTTTTAGAGCTAGAAATAGC
LmxM.21.1480_gRNA	GAAATTAATACGACTCACTATAGGGGAATGCTGGCGCGAGTGTGTTTTAGAGCTAGAAATAGC
LmxM.21.1555_gRNA	GAAATTAATACGACTCACTATAGGAACGAACAGGCAAAGGATTAGTTTTAGAGCTAGAAATAGC
LmxM.22.0100_gRNA	GAAATTAATACGACTCACTATAGGGAGAGGCGGGAGTAGTCGGCGTTTTAGAGCTAGAAATAGC
LmxM.22.0240_gRNA	GAAATTAATACGACTCACTATAGGACGCCCTCTCTCATCTTGTGTTTTAGAGCTAGAAATAGC
LmxM.23.0080_gRNA	GAAATTAATACGACTCACTATAGGTAAGAAACTCAGAGAGCAAGTTTTAGAGCTAGAAATAGC
LmxM.23.0610_gRNA	GAAATTAATACGACTCACTATAGGCCTTGCTTACGAAATCCTGTGTTTTAGAGCTAGAAATAGC
LmxM.23.1290_gRNA	GAAATTAATACGACTCACTATAGGAGGGTAGAGTATGGTGGGAGTTTTAGAGCTAGAAATAGC
LmxM.24.0350_gRNA	GAAATTAATACGACTCACTATAGGGCCTGCGTATCCATGCACGCGTTTTAGAGCTAGAAATAGC
LmxM.24.1410_gRNA	GAAATTAATACGACTCACTATAGGCGAAAACCGTAGCACATCGCGTTTTAGAGCTAGAAATAGC
LmxM.24.1550_gRNA	GAAATTAATACGACTCACTATAGGACTACCACGAGCCAAACACGTTTTAGAGCTAGAAATAGC
LmxM.24.1560_gRNA	GAAATTAATACGACTCACTATAGGAGAGGTACTIONCATGCGCGCGTTTTAGAGCTAGAAATAGC
LmxM.25.0590_gRNA	GAAATTAATACGACTCACTATAGGCTCTGATGCCGCTGTGCAGAGTTTTAGAGCTAGAAATAGC
LmxM.25.0715_gRNA	GAAATTAATACGACTCACTATAGGTTCTGTCCGACCCAGGAAATGTTTTAGAGCTAGAAATAGC
LmxM.25.1620_gRNA	GAAATTAATACGACTCACTATAGGCAATACGTACACAGGCGACAGTTTTAGAGCTAGAAATAGC
LmxM.25.2375_gRNA	GAAATTAATACGACTCACTATAGGACAGAATCTCTCTCGTCTATGTTTTAGAGCTAGAAATAGC
LmxM.26.2680_gRNA	GAAATTAATACGACTCACTATAGGACTAGCCACACCGTGGGATGTTTTAGAGCTAGAAATAGC
LmxM.27.0660_gRNA	GAAATTAATACGACTCACTATAGGATGTGAAATGTCCCGTTGCGTTTTAGAGCTAGAAATAGC
LmxM.27.1080_gRNA	GAAATTAATACGACTCACTATAGGATACCAAGGGGACGACAAACGTTTTAGAGCTAGAAATAGC
LmxM.28.0570_gRNA	GAAATTAATACGACTCACTATAGGGTCCCTTGTCACAAGAACCCTTTTTAGAGCTAGAAATAGC
LmxM.28.0980_gRNA	GAAATTAATACGACTCACTATAGGCTTTTAGCGAGAGGTCCTGTTTTAGAGCTAGAAATAGC
LmxM.29.0770_gRNA	GAAATTAATACGACTCACTATAGGAGTGTGTTCCGGAACCATGTGTTTTAGAGCTAGAAATAGC
LmxM.29.2330_gRNA	GAAATTAATACGACTCACTATAGGGAACGGTAGGGTTGAGAGGGTTTTAGAGCTAGAAATAGC
LmxM.29.2850_gRNA	GAAATTAATACGACTCACTATAGGGTGCGGGTGATTGTACATGGTTTTAGAGCTAGAAATAGC
LmxM.29.3140_gRNA	GAAATTAATACGACTCACTATAGGACAGAAGCTGGGCCGAGCGTTTTAGAGCTAGAAATAGC
LmxM.30.0080_gRNA	GAAATTAATACGACTCACTATAGGGCTGAGATCAACCCGTGTAGTTTTAGAGCTAGAAATAGC
LmxM.30.0710_gRNA	GAAATTAATACGACTCACTATAGGCTGTGTGACTTTCACCTCGTTTTAGAGCTAGAAATAGC
LmxM.30.0760_gRNA	GAAATTAATACGACTCACTATAGGGATACGTGGGTGTTCTGTGCGTTTTAGAGCTAGAAATAGC
LmxM.30.0800_gRNA	GAAATTAATACGACTCACTATAGGTAATAACGAGCAACAGAGAGTTTTAGAGCTAGAAATAGC
LmxM.30.0900_gRNA	GAAATTAATACGACTCACTATAGGAGGAGTGCCCTTACCAGGATGTTTTAGAGCTAGAAATAGC
LmxM.30.0970_gRNA	GAAATTAATACGACTCACTATAGGGGATGCAGCTCTGTGGTAAAGTTTTAGAGCTAGAAATAGC
LmxM.30.1050_gRNA	GAAATTAATACGACTCACTATAGGTGCACAGGAGAGCACAGCGGGTTTTAGAGCTAGAAATAGC
LmxM.30.1090_gRNA	GAAATTAATACGACTCACTATAGGAGCCAGGCAGATCGATGATTGTTTTAGAGCTAGAAATAGC
LmxM.30.1380_gRNA	GAAATTAATACGACTCACTATAGGGCTCTGTTGAGTGTGTGCGTTTTAGAGCTAGAAATAGC
LmxM.30.1600_gRNA	GAAATTAATACGACTCACTATAGGTGGTCGATGTGATGCGTGCAGTTTTAGAGCTAGAAATAGC
LmxM.30.2100_gRNA	GAAATTAATACGACTCACTATAGGAGAAGTGACGAATGTGTGCGTTTTAGAGCTAGAAATAGC
LmxM.30.2110_gRNA	GAAATTAATACGACTCACTATAGGCGTGGCGCGCAGCGCTGGTTTTAGAGCTAGAAATAGC
LmxM.30.2310_gRNA	GAAATTAATACGACTCACTATAGGCTGCCTTCGCGTATGCCACGTTTTAGAGCTAGAAATAGC
LmxM.30.2450_gRNA	GAAATTAATACGACTCACTATAGGGAAGCGCTTGCATCGACGAGTTTTAGAGCTAGAAATAGC
LmxM.31.0270_gRNA	GAAATTAATACGACTCACTATAGGGATGCCACTGTCAAGTTGAGTTTTAGAGCTAGAAATAGC
LmxM.31.0350_gRNA	GAAATTAATACGACTCACTATAGGCAGGAGTGACAGCGTCAGTGGTTTTAGAGCTAGAAATAGC
LmxM.31.0840_gRNA	GAAATTAATACGACTCACTATAGGTTGCGTGGTAATCGGCTAAAGTTTTAGAGCTAGAAATAGC
LmxM.31.3610_gRNA	GAAATTAATACGACTCACTATAGGTTCTCGTCTCGACAAGGTGTTTTAGAGCTAGAAATAGC
LmxM.32.1070_gRNA	GAAATTAATACGACTCACTATAGGTTAAATAAGCAACAAGGTGTTTTAGAGCTAGAAATAGC

LmxM.32.2890_gRNA	GAAATTAATACGACTCACTATAGGGTTCAGAAATGTGTGTATGCGTTTTAGAGCTAGAAATAGC
LmxM.33.0250_gRNA	GAAATTAATACGACTCACTATAGGTGTAACACGCGCGTTACATGTTTTAGAGCTAGAAATAGC
LmxM.33.2560_gRNA	GAAATTAATACGACTCACTATAGGTTTTTGAATTGTACCTGTCAGTTTTAGAGCTAGAAATAGC
LmxM.33.3645_gRNA	GAAATTAATACGACTCACTATAGGGACTGACGTCCTTCGTGAGTTTTAGAGCTAGAAATAGC
LmxM.34.0230_gRNA	GAAATTAATACGACTCACTATAGGGCGTTCTCTTCCAGGTAACGTTTTAGAGCTAGAAATAGC
LmxM.34.0270_gRNA	GAAATTAATACGACTCACTATAGGTAGAATCGACACGAACGAATGTTTTAGAGCTAGAAATAGC
LmxM.34.0400a_gRNA	GAAATTAATACGACTCACTATAGGAACACGTCATGCTTGAAGGGTTTTAGAGCTAGAAATAGC
LmxM.34.0520b_gRNA	GAAATTAATACGACTCACTATAGGTTCTTTGTGTGTCATTTGGGTTTTAGAGCTAGAAATAGC
LmxM.34.1460_gRNA	GAAATTAATACGACTCACTATAGGTGCGAGGACGTGCGAGTGAAGTTTTAGAGCTAGAAATAGC
LmxM.34.4190_gRNA	GAAATTAATACGACTCACTATAGGGTCACGAGGAGAGAACAGTGGTTTTAGAGCTAGAAATAGC
LmxM.34.4380_gRNA	GAAATTAATACGACTCACTATAGGTAGGCAGCTTATGTGATTTAGTTTTAGAGCTAGAAATAGC
LmxM.36.0480_gRNA	GAAATTAATACGACTCACTATAGGCATAAGGGACAAGACGGGAAGTTTTAGAGCTAGAAATAGC
LmxM.36.1380_gRNA	GAAATTAATACGACTCACTATAGGAGCTGCGTTGTCTTTCTTGGGTTTTAGAGCTAGAAATAGC
LmxM.36.2770_gRNA	GAAATTAATACGACTCACTATAGGTGTGAGTAACGGTGATGACTGTTTTAGAGCTAGAAATAGC
LmxM.36.4040_gRNA	GAAATTAATACGACTCACTATAGGGCGTCGAGAGTGGCAGTAAAGTTTTAGAGCTAGAAATAGC
LmxM.36.5000_gRNA	GAAATTAATACGACTCACTATAGGGGGATATGACCTCTCCTTCGTTTTAGAGCTAGAAATAGC
LmxM.36.5100_gRNA	GAAATTAATACGACTCACTATAGGAAGATAAAGGAAGCCAGGTAGTTTTAGAGCTAGAAATAGC
LmxM.36.5480_gRNA	GAAATTAATACGACTCACTATAGGTTTCTCTGTCAGTGTGGTGGGTTTTAGAGCTAGAAATAGC

Appendix III

Cell cycle marker orthologues

S.phase	G2.M.phase	Early.G1	Late.G1
LmxM.02.0610	LmxM.01.0620	LmxM.02.0710	LmxM.01.0830
LmxM.04.0280	LmxM.02.0050	LmxM.05.0480	LmxM.02.0740
LmxM.05.0010	LmxM.03.0760	LmxM.06.0410	LmxM.03.0090
LmxM.05.1110	LmxM.03.0970	LmxM.06.0415	LmxM.03.0110
LmxM.06.0010	LmxM.04.0460	LmxM.08_29.1270	LmxM.03.0870
LmxM.06.1040	LmxM.05.0040	LmxM.08_29.1360	LmxM.03.0960
LmxM.08_29.0270	LmxM.05.0070	LmxM.08_29.1370	LmxM.04.1180
LmxM.08_29.0850	LmxM.05.0420	LmxM.08_29.1960	LmxM.05.0050
LmxM.08_29.1720	LmxM.05.1200	LmxM.08_29.2460	LmxM.05.0580
LmxM.08_29.1730	LmxM.06.0010	LmxM.08_29.2461	LmxM.05.0970
LmxM.08_29.1740	LmxM.06.0220	LmxM.10.0290	LmxM.05.1090
LmxM.08.0150	LmxM.06.0270	LmxM.11.0140	LmxM.05.1170
LmxM.08.1120	LmxM.06.0430	LmxM.11.0420	LmxM.05.1190
LmxM.08.1171	LmxM.06.1150	LmxM.11.1130	LmxM.06.0160
		LmxM.11.1130partia	
LmxM.08.1230	LmxM.07.0025		LmxM.06.0210
LmxM.09.0210	LmxM.07.0310	LmxM.11.1340	LmxM.06.0860
LmxM.09.1340	LmxM.07.0930	LmxM.14.0350	LmxM.06.1030
LmxM.10.0870	LmxM.07.0930	LmxM.14.1395	LmxM.07.0040
LmxM.10.0970	LmxM.08_29.1170	LmxM.15.0270	LmxM.07.0830
LmxM.10.0990	LmxM.08_29.1330	LmxM.16.0760	LmxM.07.0831
LmxM.11.0710	LmxM.08_29.1750	LmxM.18.0170	LmxM.07.0840
LmxM.11.0720	LmxM.08_29.1760	LmxM.18.0180	LmxM.07.1080
LmxM.12.1120	LmxM.08_29.2150	LmxM.18.0510	LmxM.07.1140
LmxM.13.0980	LmxM.08_29.2560	LmxM.18.0670	LmxM.08_29.0020
LmxM.13.1650	LmxM.09.0005	LmxM.18.0680	LmxM.08_29.0210
			LmxM.08_29.0480part
LmxM.14.1060	LmxM.09.0210	LmxM.19.0060	ial
LmxM.15.0010	LmxM.09.0910	LmxM.20.1320	LmxM.08_29.0885
LmxM.15.0440	LmxM.09.0920	LmxM.21.0240	LmxM.08_29.0970
LmxM.15.0940	LmxM.09.0930	LmxM.21.0250	LmxM.08.0010
LmxM.16.0570	LmxM.09.1320	LmxM.21.1010	LmxM.08.0410
LmxM.16.0575	LmxM.09.1490	LmxM.23.1580	LmxM.08.0890
LmxM.16.0600	LmxM.10.0195	LmxM.24.0310	LmxM.08.0900
LmxM.16.0610	LmxM.10.0590	LmxM.24.0320	LmxM.08.1140
LmxM.17.0190	LmxM.11.0490	LmxM.24.1630	LmxM.09.0100
LmxM.17.0190partia			
	LmxM.11.0580	LmxM.24.1920	LmxM.09.1540
LmxM.17.0191	LmxM.11.0740	LmxM.26.0170	LmxM.10.0090
LmxM.17.0200	LmxM.12.0410	LmxM.26.0180	LmxM.10.0120
LmxM.17.0235	LmxM.12.0700	LmxM.26.0880	LmxM.10.0500

LmxM.17.0236	LmxM.13.0860	LmxM.26.0890	LmxM.10.0630
LmxM.17.0237	LmxM.13.1180	LmxM.27.1380	LmxM.11.0140
LmxM.17.1220	LmxM.15.0010	LmxM.27.1380a	LmxM.11.1380
LmxM.18.0810	LmxM.15.0540	LmxM.27.1390	LmxM.12.0120
LmxM.19.0030	LmxM.15.0590	LmxM.27.1580	LmxM.12.0510
LmxM.19.0050	LmxM.15.0930	LmxM.27.1580a	LmxM.12.1220
LmxM.20.0260	LmxM.16.0800	LmxM.28.1930	LmxM.12.1230
LmxM.20.0640	LmxM.16.1425partia l	LmxM.28.1930partia l	LmxM.12.1340
LmxM.20.1365	LmxM.16.1430	LmxM.29.2980	LmxM.13.0050
LmxM.20.1370	LmxM.16.1460	LmxM.29.3720	LmxM.13.0080
LmxM.20.1370a	LmxM.16.1460a	LmxM.29.3730	LmxM.13.0090
LmxM.21.0015	LmxM.16.1470	LmxM.31.0400	LmxM.13.0180
LmxM.21.0920	LmxM.17.0070	LmxM.31.0450	LmxM.13.0400
LmxM.21.1080	LmxM.19.0520	LmxM.32.0290	LmxM.13.0660
LmxM.21.1860	LmxM.20.0030	LmxM.32.1955	LmxM.13.1630
LmxM.22.0120	LmxM.20.1400	LmxM.32.3200	LmxM.13.1650
LmxM.22.1110	LmxM.21.0015	LmxM.33.1250	LmxM.14.0120
LmxM.23.1310	LmxM.21.1250	LmxM.34.0950	LmxM.14.0920
LmxM.23.1680	LmxM.22.0730	LmxM.34.3080	LmxM.14.1060
LmxM.25.0980	LmxM.24.1560	LmxM.34.5100	LmxM.15.0060
LmxM.25.1950	LmxM.24.1940	LmxM.36.0980	LmxM.15.0120
LmxM.25.1980	LmxM.24.1980	LmxM.36.0990	LmxM.15.1290
LmxM.25.2200	LmxM.25.0290	LmxM.36.1260	LmxM.15.1420
LmxM.25.2450	LmxM.25.1460	LmxM.36.1940	LmxM.15.1450
LmxM.26.1020	LmxM.25.2450	LmxM.36.2350	LmxM.16.0660
LmxM.26.1340	LmxM.26.0060	LmxM.36.6280	LmxM.16.0980
LmxM.26.2380	LmxM.26.1320	LmxM.36.6290	LmxM.16.1520
LmxM.26.2660	LmxM.27.1945	LmxM.36.6300	LmxM.16.1540
LmxM.27.0490	LmxM.27.2305	LmxM.36.7000	LmxM.17.0190
LmxM.27.0510	LmxM.27.2390		LmxM.17.0190partial
LmxM.27.1080	LmxM.28.0520		LmxM.17.0191
LmxM.27.1750	LmxM.28.1860		LmxM.17.0200
LmxM.27.2170	LmxM.28.2890		LmxM.17.0235
LmxM.27.2590	LmxM.29.3360		LmxM.17.0236
LmxM.28.0210	LmxM.29.3480		LmxM.17.0237
LmxM.28.0610	LmxM.30.0090		LmxM.17.0790
LmxM.28.0710	LmxM.30.1670		LmxM.17.1070
LmxM.28.1600	LmxM.30.1680		LmxM.18.0480
LmxM.28.1860	LmxM.30.3180		LmxM.18.0530
LmxM.28.2880	LmxM.31.0350		LmxM.18.0720
LmxM.29.1810	LmxM.31.0360		LmxM.18.1140
LmxM.29.3330	LmxM.31.0690		LmxM.18.1240
LmxM.30.0230	LmxM.31.1060		LmxM.19.0180
LmxM.30.2750	LmxM.31.1680		LmxM.19.1480
LmxM.30.3180	LmxM.31.1910		LmxM.20.1430
LmxM.31.2550	LmxM.31.2450		LmxM.20.1660
LmxM.31.3780	LmxM.32.1170		LmxM.21.0125
LmxM.32.0792	LmxM.32.1350		LmxM.21.0150

LmxM.32.0794	LmxM.32.2500	LmxM.21.0390
LmxM.32.2320	LmxM.33.0710	LmxM.21.1210
LmxM.32.3070	LmxM.33.3480	LmxM.21.1560
LmxM.32.3070partia l	LmxM.34.0430	LmxM.21.1650
LmxM.33.2080	LmxM.34.1310	LmxM.22.1110
LmxM.33.3880	LmxM.34.1810	LmxM.22.1410
LmxM.33.4160	LmxM.34.5080	LmxM.23.0680
LmxM.34.1310	LmxM.34.5340	LmxM.23.0950
LmxM.34.2400	LmxM.36.0020	LmxM.23.0960
LmxM.34.4690	LmxM.36.0070	LmxM.23.1310
LmxM.34.4760	LmxM.36.0590	LmxM.23.1470
LmxM.34.4800	LmxM.36.1020	LmxM.24.0060
LmxM.34.5390	LmxM.36.1470	LmxM.24.0290
LmxM.36.0020	LmxM.36.1910	LmxM.24.0670
LmxM.36.0550	LmxM.36.2550	LmxM.24.0910
LmxM.36.0950	LmxM.36.4230	LmxM.24.1010
LmxM.36.1900	LmxM.36.4910	LmxM.24.1810
LmxM.36.2035	LmxM.36.5110	LmxM.24.1820
LmxM.36.2790	LmxM.36.5300	LmxM.24.2180
LmxM.36.4980	LmxM.36.5340	LmxM.25.0050
LmxM.36.5350	LmxM.36.5800	LmxM.25.0210
	LmxM.36.5870	LmxM.25.0980
		LmxM.25.1010
		LmxM.25.1630
		LmxM.25.1780
		LmxM.26.0330
		LmxM.26.0660
		LmxM.26.1020
		LmxM.26.1230
		LmxM.26.1380
		LmxM.26.1570
		LmxM.26.2420
		LmxM.26.2440
		LmxM.26.2560
		LmxM.27.0100
		LmxM.27.0110
		LmxM.27.0350
		LmxM.27.0600
		LmxM.27.0770
		LmxM.27.1750
		LmxM.27.2270
		LmxM.27.2590
		LmxM.28.0510
		LmxM.28.0530
		LmxM.28.0610
		LmxM.28.0810
		LmxM.28.0850
		LmxM.28.0890

LmxM.28.1250
LmxM.28.1515
LmxM.28.1990
LmxM.28.2020
LmxM.28.2100
LmxM.28.2280
LmxM.28.2300
LmxM.28.2640
LmxM.28.2880
LmxM.29.0300
LmxM.29.1710
LmxM.29.1900
LmxM.29.1975
LmxM.29.2065
LmxM.29.2130
LmxM.29.2290
LmxM.29.2830
LmxM.29.3310
LmxM.29.3440
LmxM.29.3700
LmxM.30.0590
LmxM.30.0740
LmxM.30.0870
LmxM.30.0880
LmxM.30.0980
LmxM.30.0980
LmxM.30.1000
LmxM.31.0210
LmxM.31.0520
LmxM.31.1160
LmxM.31.1430
LmxM.31.1470
LmxM.31.1635
LmxM.31.1660
LmxM.31.1720
LmxM.31.1890
LmxM.31.2540
LmxM.31.2940
LmxM.31.2960
LmxM.31.3430
LmxM.31.3431
LmxM.31.3790
LmxM.32.0410
LmxM.32.0650
LmxM.32.1230
LmxM.32.1240
LmxM.32.1500
LmxM.32.2540
LmxM.32.3180

LmxM.33.0460
LmxM.33.1040
LmxM.33.2160
LmxM.33.2200
LmxM.33.2250
LmxM.33.3440
LmxM.33.3880
LmxM.33.4160
LmxM.34.0160
LmxM.34.0380
LmxM.34.0950
LmxM.34.1080
LmxM.34.1240
LmxM.34.1650
LmxM.34.1690
LmxM.34.1945
LmxM.34.2630
LmxM.34.3240
LmxM.34.3510
LmxM.34.3890
LmxM.34.4360
LmxM.34.4630
LmxM.34.4810
LmxM.36.0210
LmxM.36.0330
LmxM.36.0340
LmxM.36.0640
LmxM.36.0950
LmxM.36.1180
LmxM.36.1310
LmxM.36.1590
LmxM.36.2450
LmxM.36.2820
LmxM.36.3220
LmxM.36.4040
LmxM.36.4470
LmxM.36.4500
LmxM.36.4790
LmxM.36.4990
LmxM.36.5990
LmxM.36.6310
LmxM.36.6410
LmxM.36.6600
LmxM.36.6980
LmxM.36.7000

Appendix IV

Promastigote life cycle phase orthologues.

Pro2d	Nec4d	Lep8d	Meta14d
LmxM.05.0500	LmxM.04.0310	LmxM.04.0180	LmxM.02.0300
LmxM.05.0510	LmxM.04.0320	LmxM.04.0190	LmxM.08_29.0620
LmxM.09.0610	LmxM.23.0870	LmxM.04.0180	LmxM.11.1220
LmxM.09.1350	LmxM.23.0880	LmxM.04.0190	LmxM.11.1220a
LmxM.21.0710	LmxM.23.0880a	LmxM.04.0210	LmxM.11.1240
LmxM.25.1470	LmxM.27.2340	LmxM.09.0150	LmxM.11.1250
LmxM.31.3320	LmxM.34.0640	LmxM.09.0180	LmxM.11.1270
LmxM.26.1015	LmxM.04.1030	LmxM.16.0490	LmxM.11.1290
LmxM.28.1930	LmxM.13.0090	LmxM.20.1180	LmxM.15.0760
LmxM.28.1930partial	LmxM.32.2540	LmxM.20.1185	LmxM.27.0970
LmxM.32.3200	LmxM.20.1550	LmxM.04.0450	LmxM.27.0980
LmxM.08_29.1500	LmxM.20.1560	LmxM.21.0885	LmxM.08.1225
LmxM.08_29.1500a	LmxM.20.1570	LmxM.26.2680	LmxM.10.0390
LmxM.30.2580	LmxM.29.1380	LmxM.29.0290	LmxM.10.0405
LmxM.32.3240	LmxM.30.1130	LmxM.32.0520	LmxM.10.0470
LmxM.10.0460	LmxM.27.0420		LmxM.28.0570
LmxM.10.0465			LmxM.10.0390
LmxM.10.0465a			LmxM.10.0465a
			LmxM.28.0570
			LmxM.12.0850
			LmxM.12.0860
			LmxM.12.0870partial
			LmxM.12.0890
			LmxM.12.0891
			LmxM.12.0910
			LmxM.12.0980
			LmxM.12.0990
			LmxM.12.1090
			LmxM.19.0540
			LmxM.19.0570
			LmxM.25.1480
			LmxM.28.0570
			LmxM.30.1800
			LmxM.30.1820
			LmxM.30.1800a
			LmxM.30.2460
			LmxM.33.0500
			LmxM.33.0960
			LmxM.33.0961
			LmxM.33.0500

LmxM.33.0960
LmxM.33.0961
LmxM.33.1560
LmxM.33.1980
LmxM.08.0720
LmxM.08.0730
LmxM.08.0740
LmxM.08.0750
LmxM.08.0760
LmxM.08.0770
LmxM.08.0810
LmxM.30.0450
LmxM.30.0450a
LmxM.30.0450b
LmxM.30.0450c
LmxM.30.0450d
LmxM.30.0450e
LmxM.30.0450g
LmxM.30.0450h
LmxM.30.0450i
LmxM.30.0451
LmxM.30.0452
LmxM.30.0452c
LmxM.30.0453
LmxM.30.0454
LmxM.30.1450partial
LmxM.33.1560a
LmxM.33.1580
LmxM.33.1600
LmxM.33.1720
LmxM.33.1720a
LmxM.33.1720b
LmxM.33.1720c
LmxM.33.1721
LmxM.33.1725
LmxM.33.1740
LmxM.33.1820
LmxM.33.1840
LmxM.33.1900
LmxM.33.1900a
LmxM.33.1920
LmxM.33.1920a
LmxM.33.1920b
LmxM.33.1920c
LmxM.33.1920d
LmxM.33.1920e

LmxM.36.1270

LmxM.33.1560

LmxM.33.1980

LmxM.36.3340

Appendix V

Pseudotime and fluorescent expression inferred matching expression.

Matching expression	Hypothetical Protein
y	LmxM.04.0220
y	LmxM.04.0510
y	LmxM.04.0630
y	LmxM.04.0670
y	LmxM.04.0720
n	LmxM.04.0740
y	LmxM.05.1110
y	LmxM.07.0025
y	LmxM.07.0830
y	LmxM.08_29.0940
y	LmxM.08.0630
n	LmxM.08.0810
y	LmxM.08.1222
y	LmxM.12.0480
y	LmxM.12.0490
y	LmxM.13.1070
y	LmxM.13.1320
n	LmxM.14.0560
y	LmxM.14.0720
n	LmxM.15.1520
y	LmxM.16.0500
y	LmxM.16.1340
y	LmxM.16.1370
y	LmxM.16.1420
y	LmxM.16.1430
y	LmxM.17.0340
y	LmxM.17.0860
y	LmxM.17.0870
y	LmxM.18.1350
y	LmxM.18.1640
n	LmxM.19.1160
y	LmxM.20.0030
y	LmxM.20.1220
y	LmxM.20.1260
y	LmxM.20.1700
y	LmxM.21.1480
y	LmxM.21.1555
y	LmxM.22.0100
y	LmxM.22.0240

y	LmxM.23.0080
y	LmxM.23.0610
y	LmxM.23.1290
y	LmxM.24.0350
y	LmxM.24.1410
y	LmxM.24.1550
y	LmxM.24.1560
y	LmxM.25.0590
y	LmxM.25.0715
y	LmxM.25.1620
y	LmxM.25.2375
y	LmxM.26.2680
y	LmxM.27.0660
y	LmxM.27.1080
y	LmxM.28.0570
y	LmxM.28.0980
y	LmxM.29.0770
n	LmxM.29.2330
y	LmxM.29.2850
y	LmxM.29.3140
y	LmxM.30.0080
y	LmxM.30.0710
n	LmxM.30.0760
n	LmxM.30.0800
y	LmxM.30.0900
y	LmxM.30.0970
y	LmxM.30.1050
y	LmxM.30.1090
n	LmxM.30.1380
n	LmxM.30.1600
y	LmxM.30.2100
y	LmxM.30.2110
y	LmxM.30.2310
y	LmxM.30.2450
y	LmxM.31.0270
n	LmxM.31.0350
y	LmxM.31.0840
y	LmxM.31.3610
y	LmxM.32.1070
y	LmxM.32.2890
y	LmxM.33.0250
y	LmxM.33.2560
n	LmxM.33.3645
n	LmxM.34.0230
n	LmxM.34.0270
y	LmxM.34.0400a

y	LmxM.34.0520b
y	LmxM.34.1460
y	LmxM.34.4190
y	LmxM.34.4380
y	LmxM.36.0480
y	LmxM.36.1380
n	LmxM.36.2770
y	LmxM.36.4040
n	LmxM.36.5000
y	LmxM.36.5100
y	LmxM.36.5480

List of References

- Abu-Dayyeh, I., Hassani, K., Westra, E. R., Mottram, J. C., & Olivier, M. (2010). Comparative study of the ability of *Leishmania mexicana* promastigotes and amastigotes to alter macrophage signaling and functions. *Infection and Immunity*, *78*(6), 2438-2445. <https://doi.org/10.1128/IAI.00812-09>
- Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. *Cell*, *61*(7), 1157-1160. [https://doi.org/10.1016/0092-8674\(90\)90674-4](https://doi.org/10.1016/0092-8674(90)90674-4)
- Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P., & Sereno, D. (2016). *A Historical Overview of the Classification, Evolution, and Dispersion of Leishmania Parasites and Sandflies*. <https://doi.org/10.1371/journal.pntd.0004349>
- Alcolea, P. J., Alonso, A., Domínguez, M., Parro, V., Jiménez, M., Molina, R., & Larraga, V. (2016). *Influence of the Microenvironment in the Transcriptome of Leishmania infantum Promastigotes: Sand Fly versus Culture*. <https://doi.org/10.1371/journal.pntd.0004693>
- Alcolea, P. J., Alonso, A., Gómez, M. J., Moreno, I., Domínguez, M., Parro, V., & Larraga, V. (2010). Transcriptomics throughout the life cycle of *Leishmania infantum*: High down-regulation rate in the amastigote stage. *International Journal for Parasitology*, *40*(13), 1497-1516. <https://doi.org/10.1016/J.IJPARA.2010.05.013>
- Alcolea, P. J., Alonso, A., Gómez, M. J., Postigo, M., Molina, R., Jiménez, M., & Larraga, V. (2014a). *Stage-specific differential gene expression in Leishmania infantum: from the foregut of Phlebotomus perniciosus to the human phagocyte*. <https://doi.org/10.1186/1471-2164-15-849>
- Alcolea, P. J., Alonso, A., Gómez, M. J., Postigo, M., Molina, R., Jiménez, M., & Larraga, V. (2014b). *Stage-specific differential gene expression in Leishmania infantum: from the foregut of Phlebotomus perniciosus to the human phagocyte*. <https://doi.org/10.1186/1471-2164-15-849>
- Alcolea, P. J., Alonso, A., Sánchez-Gorostiaga, A., Moreno-Paz, M., Gómez, M. J., Ramos, I., Parro, V., & Larraga, V. (2009). Genome-wide analysis reveals increased levels of transcripts related with infectivity in peanut lectin non-agglutinated promastigotes of *Leishmania infantum*. *Genomics*, *93*(6), 551-564. <https://doi.org/10.1016/J.YGENO.2009.01.007>
- Alcoleaid, P. J., Alonso, A., Molina, R., Jiménez, M., Myler, P. J., & Larraga, V. (2019). *Functional genomics in sand fly-derived Leishmania promastigotes*. <https://doi.org/10.1371/journal.pntd.0007288>
- Alexander, J., Satoskar, A. R., & Russell, D. G. (1999). *Leishmania species: models of intracellular parasitism*. <https://doi.org/10.1046/j.1365-3113.1999.00000.x> *Journal of Cell Science* *112*, 2993-3002 (1999)
- Almeida, R., Gilmartin, B. J., McCann, S. H., Norrish, A., Ivens, A. C., Lawson, D., Levick, M. P., Smith, D. F., Dyal, S. D., Vetrie, D., Freeman, T. C., Coulson, R. M., Sampaio, I., Schneider, H., & Blackwell, J. M. (2004). Expression profiling of the *Leishmania* life cycle: cDNA arrays identify developmentally regulated genes present but not annotated in the genome. *Molecular and Biochemical Parasitology*, *136*(1), 87-100. <https://doi.org/10.1016/J.MOLBIOPARA.2004.03.004>
- Alvar, J., Vélez, I. D., Bern, C., Herrero, M., & Desjeux, P. (2012). Leishmaniasis Worldwide and Global Estimates of Its Incidence. *PLoS ONE*, *7*(5), 35671. <https://doi.org/10.1371/journal.pone.0035671>
- Alvar, J., Yactayo, S., & Bern, C. (2006). Leishmaniasis and poverty. *Trends in Parasitology*, *22*(12), 552-557. <https://doi.org/10.1016/J.PT.2006.09.004>

- Ambit, A., Woods, K. L., Cull, B., Coombs, G. H., & Mottram, J. C. (2011). Morphological events during the cell cycle of *Leishmania major*. *Eukaryotic Cell*, *10*(11), 1429-1438. <https://doi.org/10.1128/EC.05118-11>
- Amiri-Dashatan, N., Rezaei-Tavirani, M., & Ahmadi, N. (2020). A quantitative proteomic and bioinformatics analysis of proteins in metacyclogenesis of *Leishmania tropica*. *Acta Tropica*, *202*, 105227. <https://doi.org/10.1016/J.ACTATROPICA.2019.105227>
- Amiri-Dashatan, N., Rezaei-Tavirani, M., Zali, H., Koushki, M., & Ahmadi, N. (2020). Quantitative proteomic analysis reveals differentially expressed proteins in *Leishmania major* metacyclogenesis. *Microbial Pathogenesis*, *149*, 104557. <https://doi.org/10.1016/J.MICPATH.2020.104557>
- Amos, B., Aurrecochea, C., Barba, M., Barreto, A., Basenko, E. Y., Ba` Zant, W., Belnap, R., Blevins, A. S., Ohme, U. B. ` , Brestelli, J., Brunk, B. P., Caddick, M., Callan, D., Campbell, L., Christensen, M. B., Christophides, G. K., Crouch, K., Davis, K., Debarry, J., ... Zheng, J. (2022). VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*, *50*, 899. <https://doi.org/10.1093/nar/gkab929>
- Amri, M. el, Fitzgerald, U., & Schlosser, G. (2018). *MARCKS and MARCKS-like proteins in development and regeneration*. <https://doi.org/10.1186/s12929-018-0445-1>
- Antonie, J., Lang, T., Prina, E., Courret, N., & Hellio, R. (1999). *H-2M molecules, like MHC class II molecules, are targeted to parasitophorous vacuoles of Leishmania-infected macrophages and internalized by amastigotes of L. amazonensis and L. mexicana*.
- Aoki, J. I., Muxel, S. M., Zampieri, R. A., Laranjeira-Silva, M. F., Müller, K. E., Nerland, A. H., & Floeter-Winter, L. M. (2017). RNA-seq transcriptional profiling of *Leishmania amazonensis* reveals an arginase-dependent gene expression regulation. *PLoS Neglected Tropical Diseases*, *11*(10). <https://doi.org/10.1371/journal.pntd.0006026>
- Aoki, J. I., Muxel, S. M., Zampieri, R. A., Müller, K. E., Nerland, H., Lucile, & Floeter-Winter, M. (2019). *Differential immune response modulation in early Leishmania amazonensis infection of BALB/c and C57BL/6 macrophages based on transcriptome profiles*. <https://doi.org/10.1038/s41598-019-56305-1>
- Aoun, K., & Al`da Bouratbine, A. (2014). *Cutaneous Leishmaniasis in North Africa: a review*. <https://doi.org/10.1051/parasite/2014014>
- Arango Duque, G., & Descoteaux, A. (2015). *Leishmania survival in the macrophage: where the ends justify the means*. *Current Opinion in Microbiology*, *26*, 32-40. <https://doi.org/10.1016/J.MIB.2015.04.007>
- Archer, S. K., Inchaustegui, D., Queiroz, R., & Clayton, C. (2011). The Cell Cycle Regulated Transcriptome of *Trypanosoma brucei*. *PLoS ONE*, *6*(3), 18425. <https://doi.org/10.1371/journal.pone.0018425>
- Archer, S., Queiroz, R., Stewart, M., & Clayton, C. (2008). Chapter 18 *Trypanosomes as a Model to Investigate mRNA Decay Pathways*. *Methods in Enzymology*, *448*, 359-377. [https://doi.org/10.1016/S0076-6879\(08\)02618-9](https://doi.org/10.1016/S0076-6879(08)02618-9)
- Artigas-Jerónimo, S., Villar, M., Estrada-Pena, A., Velázquez-Campoy, A., Alberdi, P., & de La Fuente, J. (2021). Function of cofactor Akirin2 in the regulation of gene expression in model human Caucasian neutrophil-like HL60 cells. *Bioscience Reports*, *41*(7). <https://doi.org/10.1042/BSR20211120>
- Baitsch, D., Bock, H. H., Engel, T., Telgmann, R., Müller-Tidow, C., Varga, G., Bot, M., Herz, J., Robenek, H., von Eckardstein, A., & Nofer, J.-R. (2011).

- Apolipoprotein E Induces Antiinflammatory Phenotype in Macrophages.*
<https://doi.org/10.1161/ATVBAHA.111.222745>
- Baker, N., Catta-Preta, C. M. C., Neish, R., Sadlova, J., Powell, B., Alves-Ferreira, E. V. C., Geoghegan, V., Carnielli, J. B. T., Newling, K., Hughes, C., Vojtkova, B., Anand, J., Mihut, A., Walrad, P. B., Wilson, L. G., Pitchford, J. W., Volf, P., & Mottram, J. C. (2021). *Systematic functional analysis of Leishmania protein kinases identifies regulators of differentiation or survival.* <https://doi.org/10.1038/s41467-021-21360-8>
- Ballet, R., Emre, R., Jemelin, Y., Charmoy, S., & Tacchini-Cottier, M. (2014). Blocking Junctional Adhesion Molecule C Enhances Dendritic Cell Migration and Boosts the Immune Responses against *Leishmania major*. *PLoS Pathog*, *10*(12), 1004550. <https://doi.org/10.1371/journal.ppat.1004550>
- Baron, N., Tupperwar, N., Dahan, I., Hadad, U., Davidov, G. I., ZarivachID, R., & ShapiraID, M. (2021). *Distinct features of the Leishmania cap-binding protein LeishIF4E2 revealed by CRISPR-Cas9 mediated hemizygous deletion.* <https://doi.org/10.1371/journal.pntd.0008352>
- Barrangou, R., et al., & Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*.
<https://doi.org/10.1029/2004GL019460>
- Barrett, M. P., Kyle, D. E., David Sibley, L., Radke, J. B., & Tarleton, R. L. (2019). *Protozoan persister-like cells and drug treatment failure.* <https://doi.org/10.1038/s41579-019-0238-x>
- Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O. M., Lassadi, I., Gupta, V., Tromer, E., Mourier, T., Stevens, T. J., Breckels, L. M., Pain, A., Lilley, K. S., & Waller, R. F. (2020). A Comprehensive Subcellular Atlas of the *Toxoplasma* Proteome via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell Host & Microbe*, *28*(5), 752-766.e9.
<https://doi.org/10.1016/J.CHOM.2020.09.011>
- Bates, E. J., Knuepfer, E., & Smith, D. F. (2000). Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites. In *Nucleic Acids Research* (Vol. 28, Issue 5).
- Bates, P. A. (1994). The Developmental Biology of *Leishmania* Promastigotes. *Experimental Parasitology*, *79*(2), 215-218.
<https://doi.org/10.1006/EXPR.1994.1084>
- Bates, P. A. (2007). Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. In *International Journal for Parasitology* (Vol. 37, Issue 10, pp. 1097-1106). <https://doi.org/10.1016/j.ijpara.2007.04.003>
- Bates, P. A. (2008). *Leishmania* sand fly interaction: progress and challenges. *Current Opinion in Microbiology*, *11*(4), 340-344.
<https://doi.org/10.1016/J.MIB.2008.06.003>
- Bates, P. A. (2018). *Revising Leishmania's life cycle.* <https://doi.org/10.1038/s41564-018-0154-2>
- Bates, P. A., Robertson, C. D., Coombs, G. H., & Tetley, L. (1992). Axenic cultivation and characterization of *Leishmania mexicana* amastigote-like forms. *Parasitology*, *105*(2), 193-202.
<https://doi.org/10.1017/S0031182000074102>
- Bates, P. A., & Tetley, L. (1993a). *Leishmania mexicana*: Induction of Metacyclogenesis by Cultivation of Promastigotes at Acidic pH. *Experimental Parasitology*, *76*(4), 412-423. <https://doi.org/10.1006/EXPR.1993.1050>
- Bates, P. A., & Tetley, L. (1993b). *Leishmania mexicana*: Induction of Metacyclogenesis by Cultivation of Promastigotes at Acidic pH. *Experimental Parasitology*, *76*(4), 412-423. <https://doi.org/10.1006/EXPR.1993.1050>

- Bauxbaum, L., Denise, H., Coombs, G., Alexander, J., Mottram, J., & Scott, P. (2003). *Cysteine Protease B of Leishmania mexicana Inhibits Host Th1 Responses and Protective Immunity*.
<https://doi.org/10.4049/jimmunol.171.7.3711>
- Bazzoni, G. (2003). The JAM family of junctional adhesion molecules. *Current Opinion in Cell Biology*, 15(5), 525-530. [https://doi.org/10.1016/S0955-0674\(03\)00104-2](https://doi.org/10.1016/S0955-0674(03)00104-2)
- Bellatin, J. A., Murray, A. S., Zhao, M., & McMaster, W. R. (2002). Leishmania mexicana: Identification of Genes That Are Preferentially Expressed in Amastigotes. *Experimental Parasitology*, 100(1), 44-53.
<https://doi.org/10.1006/EXPR.2001.4677>
- Beneke, T., & Gluenz, E. (2019). *LeishGEdit: A Method for Rapid Gene Knockout and Tagging Using CRISPR-Cas9*. <http://www.springer.com/series/7651>
- Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., & Gluenz Sir William, E. (2017). *A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids*. <https://doi.org/10.1098/rsos.170095>
- Bennis, I., de Brouwere, V., Belrhiti, Z., Sahibi, H., & Boelaert, M. (2018). Psychosocial burden of localised cutaneous Leishmaniasis: A scoping review. In *BMC Public Health* (Vol. 18, Issue 1). BioMed Central Ltd.
<https://doi.org/10.1186/s12889-018-5260-9>
- Beverley, S. (2002). *PROTOZOMICS: TRYPANOSOMATIDPARASITE GENETICS COMES OF AGE*. <https://doi.org/10.1038/nrg980>
- Bevkal Id, S., Naguleswaran, A., Rehmann, R., Id, M. K., Hellerid, M., & Roditiid, I. (2021). *An Alba-domain protein required for proteome remodelling during trypanosome differentiation and host transition*.
<https://doi.org/10.1371/journal.ppat.1009239>
- Bichiou, H., Bouabid, C., Rabhi, I., & Guizani-Tabbane, L. (2021). *Transcription Factors Interplay Orchestrates the Immune-Metabolic Response of Leishmania Infected Macrophages*.
<https://doi.org/10.3389/fcimb.2021.660415>
- Black, C. (2020). *WHO_EB147 22MAY2020_761.jpg*.
<http://apps.who.int/bookorders>.
- Briggs, E., Rojas, F., McCulloch, R., Matthews, K. R., & Otto, T. D. (2021). Single-cell transcriptomic analysis of bloodstream Trypanosoma brucei reconstructs cell cycle progression and developmental quorum sensing. *Nat. Comms*. <https://doi.org/10.1038/s41467-021-25607-2>
- Briggs, E., Warren, F. S., Matthews, K., McCulloch, R., & Otto, T. (2021). Application of single-cell transcriptomics to kinetoplastid research. *Parasitology*. <https://doi.org/10.1017/S003118202100041X>
- Brotherton, M. C., Racine, G., Foucher, A. L., Drummelsmith, J., Papadopoulou, B., & Ouellette, M. (2010). Analysis of stage-specific expression of basic proteins in leishmania infantum. *Journal of Proteome Research*, 9(8), 3842-3853. <https://doi.org/10.1021/pr100048m>
- Buccitelli, C., & Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics*.
<https://doi.org/10.1038/s41576-020-0258-4>
- Buchholz, V. R., & Flossdorf, M. (2018). Single-Cell Resolution of T Cell Immune Responses. *Advances in Immunology*, 137, 1-41.
<https://doi.org/10.1016/BS.AI.2017.12.001>
- Buchholz, V. R., Schumacher, T. N. M., & Busch, D. H. (2016). T Cell Fate at the Single-Cell Level. *Annual Review of Immunology*, 34, 65-92.
<https://doi.org/10.1146/annurev-immunol-032414-112014>

- Buckner, F. S., Wilson, A. J., & van Voorhis, W. C. (1999). Detection of Live *Trypanosoma cruzi* in Tissues of Infected Mice by Using Histochemical Stain for-Galactosidase. In *INFECTION AND IMMUNITY* (Vol. 67, Issue 1). <https://journals.asm.org/journal/iai>
- Burchmore, R. J. S., & Barrett, M. P. (2001). Life in vacuoles - nutrient acquisition by *Leishmania* amastigotes. *International Journal for Parasitology*, 31(12), 1311-1320. [https://doi.org/10.1016/S0020-7519\(01\)00259-4](https://doi.org/10.1016/S0020-7519(01)00259-4)
- Burchmore, R. J. S., & Hart, D. T. (1995). Glucose transport in amastigotes and promastigotes of *Leishmania mexicana mexicana*. *Molecular and Biochemical Parasitology*, 74(1), 77-86. [https://doi.org/10.1016/0166-6851\(95\)02485-9](https://doi.org/10.1016/0166-6851(95)02485-9)
- Burchmore, R. J. S., & Landfear, S. M. (1998). Differential regulation of multiple glucose transporter genes in *Leishmania mexicana*. *Journal of Biological Chemistry*, 273(44), 29118-29126. <https://doi.org/10.1074/jbc.273.44.29118>
- Burza MBChB, S., Croft, S. L., & Boelaert, M. (2018). Leishmaniasis. *The Lancet*, 392, 951-970. [https://doi.org/10.1016/S0140-6736\(18\)31204-2](https://doi.org/10.1016/S0140-6736(18)31204-2)
- Buxbaum, L. U., Denise, H., Coombs, G. H., Alexander, J., Mottram, J. C., & Scott, P. (2003). Cysteine Protease B of *Leishmania mexicana* Inhibits Host Th1 Responses and Protective Immunity . *The Journal of Immunology*, 171(7), 3711-3717. <https://doi.org/10.4049/jimmunol.171.7.3711>
- Camacho, E., Rastrojo, A., Sanchiz, Á., González-De La Fuente, S., Aguado, B., & Requena, J. M. (2019). *Leishmania Mitochondrial Genomes: Maxicircle Structure and Heterogeneity of Minicircles*. <https://doi.org/10.3390/genes10100758>
- Castro, D. P., Seabra, S. H., Garcia, E. S., Souza, W. de, & Azambuja, P. (2007). *Trypanosoma cruzi*: Ultrastructural studies of adhesion, lysis and biofilm formation by *Serratia marcescens*. *Experimental Parasitology*, 117(2), 201-207. <https://doi.org/10.1016/J.EXPPARA.2007.04.014>
- Chandra, P., Grigsby, S. J., & Philips, J. A. (2022). *Immune evasion and provocation by Mycobacterium tuberculosis*. <https://doi.org/10.1038/s41579-022-00763-4>
- Chandra, U., Yadav, A., Kumar, D., & Saha, S. (2017). *Cell cycle stage-specific transcriptional activation of cyclins mediated by HAT2-dependent H4K10 acetylation of promoters in Leishmania donovani*. <https://doi.org/10.1371/journal.ppat.1006615>
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Publishing Group*, 18. <https://doi.org/10.1038/nrm.2017.48>
- Chang, K.-P., & Dwyer, D. M. (1978). *Leishmania donovani*. Hamster macrophage interactions in vitro: cell entry, intracellular survival, and multiplication of amastigotes. <https://doi.org/10.1084/jem.147.2.515>
- Chaparro, V., Leroux, L.-P., Zimmermann, A., Jardim, A., Johnston, B., Descoteaux, A., & Jaramillo, M. (2019). *Leishmania donovani Lipophosphoglycan Increases Macrophage-Dependent Chemotaxis of CXCR6-Expressing Cells via CXCL16 Induction*. <https://doi.org/10.1128/IAI>
- Charmoy, M., Auderset, F., Allenbach, C., & Tacchini-Cottier, F. (2010). The Prominent Role of Neutrophils during the Initial Phase of Infection by *Leishmania* Parasites. *Journal of Biomedicine and Biotechnology*, 2010. <https://doi.org/10.1155/2010/719361>

- Chattopadhyay, P. K., Gierahn, T. M., Roederer, M., & Love, J. C. (2014). 1 2 8 VOLUME 15 NUMBER 2 FEBRUARY 2014 *nature immunology*.
<https://doi.org/10.1038/ni.2796>
- Chattopadhyay, P. K., Roederer, M., & Chattopadhyay, P. K. (2015). *A Mine Is a Terrible Thing to Waste: High Content, Single Cell Technologies for Comprehensive Immune Analysis*. <https://doi.org/10.1111/ajt.13193>
- Chen, H., Lareau, C., Andreani, T., Vinyard, M. E., Garcia, S. P., Clement, K., Andrade-Navarro, M. A., Buenrostro, J. D., & Pinello, L. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biology*, 20(1). <https://doi.org/10.1186/s13059-019-1854-5>
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). *Spatially resolved, highly multiplexed RNA profiling in single cells*. <https://doi.org/10.1126/science.aaa6090>
- Christiano, R., Nikolay, ‡ §, Kolev, G., Shi, H., Ullu, E., Walther, T. C., & Tschudi, C. (2017). *The proteome and transcriptome of the infectious metacyclic form of Trypanosoma brucei define quiescent cells primed for mammalian invasion*. <https://doi.org/10.1111/mmi.13754>
- Chu, K. H., Lin, S. Y., & Chiang, B. L. (2021). STAT6 Pathway Is Critical for the Induction and Function of Regulatory T Cells Induced by Mucosal B Cells. *Frontiers in Immunology*, 11. <https://doi.org/10.3389/fimmu.2020.615868>
- Clayton, C. (2019). *Regulation of gene expression in trypanosomatids: living with polycistronic transcription history appears in chronological order*. <https://doi.org/10.1098/rsob.190072>
- Clayton, C., & Shapira, M. (2007). Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology*, 156(2), 93-101.
<https://doi.org/10.1016/J.MOLBIOPARA.2007.07.007>
- Coelho, V. T. S., Oliveira, J. S., Valadares, D. G., Chá Vez-Fumagalli, M. A., Duarte, M. C., Lage, P. S., Soto, M., Santoro, M. M., Tavares, C. A. P., Fernandes, A. P., & Coelho, E. A. F. (2012). *Identification of Proteins in Promastigote and Amastigote-like Leishmania Using an Immunoproteomic Approach*. <https://doi.org/10.1371/journal.pntd.0001430>
- Cohen-Freue, G., Holzer, T. R., Forney, J. D., & McMaster, W. R. (2007). Global gene expression in Leishmania. *International Journal for Parasitology*, 37(10), 1077-1086. <https://doi.org/10.1016/J.IJPARA.2007.04.011>
- Corradin, S., Ransijn, A., Corradin, G., Roggero, M. A., Schmitz, A. A. P., Schneider, P., Mauël, J., & Vergères, G. (1999). MARCKS-related protein (MRP) is a substrate for the Leishmania major surface protease leishmanolysin (gp63). *Journal of Biological Chemistry*, 274(36), 25411-25418. <https://doi.org/10.1074/jbc.274.36.25411>
- Cortazzo Da Silva, L., Aoki, J. I., & Floeter-Winter, L. M. (2022). *Finding Correlations Between mRNA and Protein Levels in Leishmania Development: Is There a Discrepancy?* <https://doi.org/10.3389/fcimb.2022.852902>
- Courret, N., Fréhel, C., Gouhier, N., Pouchelet, M., Prina, E., Roux, P., & Antoine, J.-C. (2002). Biogenesis of Leishmania-harbouring parasitophorous vacuoles following phagocytosis of the metacyclic promastigote or amastigote stages of the parasites. In *Journal of Cell Science* (Vol. 115).
- Coutinho-Abreu, I. I. v, Serafim, T. D., Meneses, C., Kamhawi, S., Oliveira, F., & Valenzuelaid, J. G. (2020). Distinct gene expression patterns in vector-residing Leishmania infantum identify parasite stage-enriched markers. *PLOS Neglected Tropical Diseases*. <https://doi.org/10.1371/journal.pntd.0008014>

- Cramer, P. (2019). Eukaryotic Transcription Turns 50. In *Cell* (Vol. 179, Issue 4, pp. 808-812). Cell Press. <https://doi.org/10.1016/j.cell.2019.09.018>
- Cristina Terrão, M., José Rosas de Vasconcelos, E., nia Aquino Defina, T., Myler, P. J., & Kaysel Cruz, A. (2017). *Disclosing 3' UTR cis-elements and putative partners involved in gene expression regulation in Leishmania spp.* <https://doi.org/10.1371/journal.pone.0183401>
- Cruz, A., & Beverley, A. (1990). *Gene replacement in parasitic protozoa.*
- Cuervo, P., de Jesus, J. B., Saboia-Vahia, L., Mendonça-Lima, L., Domont, G. B., & Cupolillo, E. (2009). Proteomic characterization of the released/secreted proteins of *Leishmania (Viannia) braziliensis* promastigotes. *Journal of Proteomics*, 73(1), 79-92. <https://doi.org/10.1016/J.JPROT.2009.08.006>
- Curotto De Lafaille, M. A., Laban, A., & Wirth, D. F. (1992). *Gene expression in Leishmania: Analysis of essential 5' DNA sequences (protozoa/Kinetoplastida/trans-splicing/transcription)* (Vol. 89).
- Cuypers, B., Meysman, P., Erb, I., Bittremieux, W., Valkenburg, D., Baggerman, G., Mertens, I., Sundar, S., Khanal, B., Notredame, C., Dujardin, J. C., Domagalska, M. A., & Laukens, K. (2022). Four layer multi-omics reveals molecular responses to aneuploidy in *Leishmania*. *PLoS Pathogens*, 18(9). <https://doi.org/10.1371/journal.ppat.1010848>
- da Silva, R., Hall, B., & Sacks, D. (1989). CR1, the C3b receptor, mediates binding of infective *Leishmania* major metacyclic promastigotes to human macrophages. *J Immunol*, Jul 15;143(2):, 617-22.
- da Silva Santos, C., Ida Brodskyn, C., Wael Daboul, M., Bogdan, C., & Martins Rodrigues, M. (2014). *The role of CD4 and CD8 T cells in human cutaneous leishmaniasis.* <https://doi.org/10.3389/fpubh.2014.00165>
- Damianoud, A., Burgeid, R. J., Catta-Pretaid, C. M. C., Geogheganid, V., Romina Nievasid, Y., Newlingid, K., Brown, E., Burchmoreid, R., Rodenkoid, B., & Mottramid, J. C. (2020). *Essential roles for deubiquitination in Leishmania life cycle progression.* <https://doi.org/10.1371/journal.ppat.1008455>
- Dandugudumula, R., Fischer-Weinberger, R., & Zilberstein, D. (2022). *Morphogenesis Dynamics in Leishmania Differentiation.* <https://doi.org/10.3390/pathogens11090952>
- de Cássia Ruy, P., Melquie Monteiro-Teles, N., Daniel Miserani Magalhães, R., Freitas-Castro, F., Dias, L., Paula Aquino Defina, T., José Rosas De Vasconcelos, E., Myler, P. J., Kaysel Cruz, A., & Magalhães, M. (2019). *Comparative transcriptomics in Leishmania braziliensis: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages.* <https://doi.org/10.1080/15476286.2019.1574161>
- de Menezes, J. P., Saraiva, E. M., & da Rocha-Azevedo, B. (2016). The site of the bite: *Leishmania* interaction with macrophages, neutrophils and the extracellular matrix in the dermis. In *Parasites and Vectors* (Vol. 9, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13071-016-1540-3>
- de Pablos, L. M., Ferreira, T. R., Dowle, A. A., Forrester, S., Parry, E., Newling, K., & Walrad, P. B. (2019). The mRNA-bound Proteome of *Leishmania mexicana*: Novel genetic insight into an ancient parasite*. *Molecular and Cellular Proteomics*, 18(7), 1271-1284. <https://doi.org/10.1074/mcp.RA118.001307>
- de Sousa, L. R. F., Wu, H., Nebo, L., Fernandes, J. B., da Silva, M. F. das G. F., Kiefer, W., Schirmeister, T., & Vieira, P. C. (2015). Natural products as inhibitors of recombinant cathepsin L of *Leishmania mexicana*. *Experimental Parasitology*, 156, 42-48. <https://doi.org/10.1016/J.EXPPARA.2015.05.016>

- de Souza Leao, S., Lang, T., Prina, E., Hellio, R., & Antoine, J. (1995). *Intracellular Leishmania amazonensis amastigotes internalize and degrade MHC class II molecules of their host cells.*
- de Veer, M. J., Curtis, J. M., Baldwin, T. M., Didonato, J. A., Sexton, A., Mcconville, M. J., Handman, E., & Schofield, L. (2003). *MyD88 is essential for clearance of Leishmania major: possible role for lipophosphoglycan and Toll-like receptor 2 signaling.* <https://doi.org/10.1002/eji.200324128>
- de Villiers, M., Spry, C., Macuamule, C. J., Barnard, L., Wells, G., Saliba, K. J., & Strauss, E. (2017). Antiplasmodial Mode of Action of Pantothenamides: Pantothenate Kinase Serves as a Metabolic Activator Not as a Target. *ACS Infectious Diseases*, 3(7), 527-541. <https://doi.org/10.1021/acsinfecdis.7b00024>
- Dean, S., Sunter, J., Wheeler, R. J., Hodgkinson, I., Gluenz, E., & Gull, K. (2015). *A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids.* <https://doi.org/10.1098/rsob.140197>
- Decuypere, S., Vandesompele, J., Yardley, V., de Doncker, S., Laurent, T., Rijal, S., Llanos-Cuentas, A., Chappuis, F., Arevalo, J., & Dujardin, J. C. (2005). Differential polyadenylation of ribosomal RNA during post-transcriptional processing in Leishmania. *Parasitology*, 131(3), 321-329. <https://doi.org/10.1017/S0031182005007808>
- Devault, A., & Băuls, A. L. (2008). The promastigote surface antigen gene family of the Leishmania parasite: Differential evolution by positive selection and recombination. *BMC Evolutionary Biology*, 8(1). <https://doi.org/10.1186/1471-2148-8-292>
- Dillon, L. A. L., Okrah, K., Hughitt, V. K., Suresh, R., Li, Y., Fernandes, M. C., Belew, A. T., Corrada Bravo, H., Mosser, D. M., & El-Sayed, N. M. (2015). Transcriptomic profiling of gene expression and RNA processing during Leishmania major differentiation. *Nucleic Acids Research*, 43(14), 6799-6813. <https://doi.org/10.1093/nar/gkv656>
- Doehl, J. S. P., Bright, Z., Dey, S., Davies, H., Magson, J., Brown, N., Romano, A., Dalton, J. E., Pinto, A. I., Pitchford, J. W., & Kaye, P. M. (2017). *Skin parasite landscape determines host infectiousness in visceral leishmaniasis.* <https://doi.org/10.1038/s41467-017-00103-8>
- Donovan, M. J., Maciuba, B. Z., Mahan, C. E., & Mcdowell, M. A. (2009). *Leishmania Infection Inhibits Cycloheximide-Induced Macrophage Apoptosis in a Strain Dependent Manner.* <https://doi.org/10.1016/j.exppara.2009.05.012>
- Dos, C., Meira, S., & Gedamu, L. (2019). *microorganisms Protective or Detrimental? Understanding the Role of Host Immunity in Leishmaniasis.* <https://doi.org/10.3390/microorganisms7120695>
- Dostálová, A., & Volf, P. (2012). *Leishmania development in sand flies: parasite-vector interactions overview.* <https://doi.org/10.1186/1756-3305-5-276>
- Doudna, J. A., & Charpentier, E. (2014). *The new frontier of genome engineering with CRISPR-Cas9.* <https://www.science.org>
- Downey, N., Hines, J. C., Sinha, K. M., & Ray, D. S. (2005). Mitochondrial DNA Ligases of Trypanosoma brucei. *Eukaryotic Cell*, 4(4), 765-774. <https://doi.org/10.1128/EC.4.4.765-774.2005>
- Dunkelberger, J. R., & Song, W.-C. (2010). Complement in host immunity 34 Complement and its role in innate and adaptive immune responses. *Cell Research*, 20(1), 34-50. <https://doi.org/10.1038/cr.2009.139>
- Emiliano Da Silva, R., Sampaio, B. M., Tonhosolo, R., Perei Ra Da Costa, A., da Silva Costa, L. E., Nieri-Bastos, F. A., Sperança, M. A., & Marcili, A. (2019).

- Exploring Leishmania infantum cathepsin as a new molecular marker for phylogenetic relationships and visceral leishmaniasis diagnosis.*
<https://doi.org/10.1186/s12879-019-4463-8>
- Esch, K. J., & Petersen, C. A. (2013). Transmission and epidemiology of zoonotic protozoal diseases of companion animals. *Clinical Microbiology Reviews*, 26(1), 58-85. <https://doi.org/10.1128/CMR.00067-12>
- Fadda, A., Ryten, † Mark, Droll, D., Rojas, F., Färber, V., Haanstra, J. R., Merce, C., Bakker, B. M., 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.* <https://doi.org/10.1111/mmi.12764>
- Fernandes, J. C. R., Acuña, S. M., Aoki, J. I., Floeter-Winter, L. M., & Muxel, S. M. (2019). *non-coding RNA Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease.* <https://doi.org/10.3390/ncrna5010017>
- Fernandes, M. C., & Andrews, N. W. (2012). Host cell invasion by *Trypanosoma cruzi*: A unique strategy that promotes persistence. In *FEMS Microbiology Reviews* (Vol. 36, Issue 3, pp. 734-747). <https://doi.org/10.1111/j.1574-6976.2012.00333.x>
- Fernandes, M. C., Dillon, L. A. L., Belew, A. T., Bravo, H. C., Mosser, D. M., & El-Sayed, N. M. (2016). Dual transcriptome profiling of *Leishmania*-infected human macrophages reveals distinct reprogramming signatures. *MBio*, 7(3). <https://doi.org/10.1128/mBio.00027-16>
- Ferreira, T. R., Dowle, A. A., Parry, E., Alves-Ferreira, E. V. C., Hogg, K., Kolokousi, F., Larson, T. R., Plevin, M. J., Cruz, A. K., & Walrad, P. B. (2020). PRMT7 regulates RNA-binding capacity and protein stability in *Leishmania* parasites. *Nucleic Acids Research*, 48(10), 5511-5526. <https://doi.org/10.1093/NAR/GKAA211>
- Fiebig, M., Kelly, S., & Gluenz, E. (2015). *Comparative Life Cycle Transcriptomics Revises Leishmania mexicana Genome Annotation and Links a Chromosome Duplication with Parasitism of Vertebrates.* <https://doi.org/10.1371/journal.ppat.1005186>
- Figueiredo, L. M., Janzen, C. J., & Cross, G. A. M. (2008). *A Histone Methyltransferase Modulates Antigenic Variation in African Trypanosomes.* <https://doi.org/10.1371/journal.pbio.0060161>
- Figueiredo, L. M., M Cross, G. A., & Janzen, C. J. (2009). *Epigenetic regulation in African trypanosomes: a new kid on the block.* <https://doi.org/10.1038/nrmicro2149>
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., & Gottardo, R. (2015). MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology*, 16(1). <https://doi.org/10.1186/s13059-015-0844-5>
- Flannagan, R., Jaumouille, V., & Grinstein, S. (2012). *The Cell Biology of Phagocytosis - annurev-pathol-011811-132445.*
- Fleming, B. D., Chandrasekaran, P., Dillon, L. A. L., Dalby, E., Suresh, R., Sarkar, A., El-Sayed, N. M., & Mosser, D. M. (2015). The generation of macrophages with anti-inflammatory activity in the absence of STAT6 signaling. *J. Leukoc. Biol*, 98, 395-407. <https://doi.org/10.1189/jlb.2A1114-560R>
- Forestier, C. L., MacHu, C., Loussert, C., Pescher, P., & Späth, G. F. (2011). Imaging host cell-leishmania interaction dynamics implicates parasite motility, lysosome recruitment, and host cell wounding in the infection

- process. *Cell Host and Microbe*, 9(4), 319-330.
<https://doi.org/10.1016/j.chom.2011.03.011>
- Forrester, S., Goundry, A., Torres Dias, B., Leal-Calvo, T., Moraes, M. O., Kaye, P. M., Mottram, J. C., Paula, A., & Lima, C. A. (2022). *Tissue specific dual RNA-seq defines host-parasite interplay in murine visceral leishmaniasis caused by Leishmania donovani and Leishmania infantum*.
<https://doi.org/10.1101/2022.02.04.479211>
- França-Costa, J., Nacagami Sotto, M., Bordignon, J., Rogério Pavanelli, W., Tomiotto-Pellissier, F., Taciane da Silva Bortoleti, B., Paulo Assolini, J., Daiele Gonçalves, M., Cristina Machado Carloto, A., Menegazzo Miranda-Sapla, M., & Conchon-Costa, I. (2018). Macrophage Polarization in Leishmaniasis: Broadening Horizons. *Frontiers in Immunology* | *Www.Frontiersin.Org*, 9, 2529. <https://doi.org/10.3389/fimmu.2018.02529>
- Gao, J., Zhu, X., Wu, M., Jiang, L., Wang, F., & He, S. (2021). *IFI27 may predict and evaluate the severity of respiratory syncytial virus infection in preterm infants*. <https://doi.org/10.1186/s41065-020-00167-5>
- Gassen, A., Brechtefeld, D., Schandry, N., Arteaga-Salas, J. M., Israel, L., Imhof, A., & Janzen, C. J. (2012). *DOT1A-dependent H3K76 methylation is required for replication regulation in Trypanosoma brucei*.
<https://doi.org/10.1093/nar/gks801>
- Geissmann, F., Manz, M. G., Jung, S., Sieweke, M. H., Merad, M., & Ley, K. (2010). *Development of Monocytes, Macrophages, and Dendritic Cells*.
<https://www.science.org>
- Gilabert Carbajo, C., Cornell Id, L. J., Madbouly Id, Y., Lai Id, Z., Yates, P. A., Tinti, M., & Tiengweid, C. (2021). *Novel aspects of iron homeostasis in pathogenic bloodstream form Trypanosoma brucei*.
<https://doi.org/10.1371/journal.ppat.1009696>
- Giladi, A., Cohen, M., Medaglia, C., Baran, Y., Li, B., Zada, M., Bost, P., Blecher-Gonen, R., Salame, T.-M., Mayer, J. U., David, E., Ronchese, F., Tanay, A., & Amit, I. (2020). *Dissecting cellular crosstalk by sequencing physically interacting cells*. <https://doi.org/10.1038/s41587-020-0442-2>
- Girard, J. R., Goins, L. M., Vuu, D. M., Sharpley, M. S., Spratford, C. M., Mantri, S. R., & Banerjee, U. (2021). *Paths and pathways that generate cell-type heterogeneity and developmental progression in hematopoiesis*. 10.
<https://doi.org/10.7554/eLife>
- Giraud, E., Martin, O., Yakob, L., & Rogers, M. (2019). *Quantifying Leishmania Metacyclic Promastigotes from Individual Sandfly Bites Reveals the Efficiency of Vector Transmission*. <https://doi.org/10.1038/s42003-019-0323-8>
- Giri, S., & Shaha, C. (2019). *Leishmania donovani parasite requires Atg8 protein for infectivity and survival under stress*. <https://doi.org/10.1038/s41419-019-2038-7>
- Gonzaga dos Santos, M., Fernanda Laranjeira da Silva, M., Andrade Zampieri, R., Marino Lafraia, R., & Maria Floeter-Winter, L. (2011). *Correlation of meta 1 expression with culture stage, cell morphology and infectivity in Leishmania (Leishmania) amazonensis promastigotes* (Vol. 106, Issue 2).
- Gossage, S. M., Rogers, M. E., & Bates, P. A. (2003). *Two separate growth phases during the development of Leishmania in sand flies: implications for understanding the life cycle*.
- Gramiccia, M., & Gradoni, L. (2005). The current status of zoonotic leishmaniasis and approaches to disease control. *International Journal for*

- Parasitology*, 35(11-12), 1169-1180.
<https://doi.org/10.1016/J.IJPARA.2005.07.001>
- Gray, M. J., Poljakovic, M., Kepka-Lenhart, D., & Morris, S. M. (2005). Induction of arginase I transcription by IL-4 requires a composite DNA response element for STAT6 and C/EBP β . *Gene*, 353(1), 98-106.
<https://doi.org/10.1016/J.GENE.2005.04.004>
- Grünebast, J., & Clos, J. (2020). Leishmania: Responding to environmental signals and challenges without regulated transcription. In *Computational and Structural Biotechnology Journal* (Vol. 18, pp. 4016-4023). Elsevier B.V.
<https://doi.org/10.1016/j.csbj.2020.11.058>
- Gschwandtner, M., Derler, R., & Midwood, K. S. (2019). More Than Just Attractive: How CCL2 Influences Myeloid Cell Behavior Beyond Chemotaxis. *Frontiers in Immunology | Www.Frontiersin.Org*, 10, 2759.
<https://doi.org/10.3389/fimmu.2019.02759>
- Guimarães-Costa, A. B., DeSouza-Vieira, T. S., Paletta-Silva, R., Freitas-Mesquita, A. L., Meyer-Fernandes, J. R., & Saraiva, E. M. (2014). 3'-nucleotidase/nuclease activity allows Leishmania parasites to escape killing by neutrophil extracellular traps. *Infection and Immunity*, 82(4), 1732-1740.
<https://doi.org/10.1128/IAI.01232-13>
- Gupta, G., Oghumu, S., & Satoskar, A. R. (2013). *Mechanisms of Immune Evasion in Leishmaniasis*. <https://doi.org/10.1016/B978-0-12-407679-2.00005-3>
- Gurung, P., & Kanneganti, T.-D. (2015). *Microreview Innate immunity against Leishmania infections*. <https://doi.org/10.1111/cmi.12484>
- Hafemeister, C., & Satija, R. (2019). *Method Open Access Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression*. <https://doi.org/10.1186/s13059-019-1874-1>
- Hallé, M., Gomez, M. A., Stuble, M., Shimizu, H., McMaster, W. R., Olivier, M., & Tremblay, M. L. (2009). The Leishmania surface protease GP63 cleaves multiple intracellular proteins and actively participates in p38mitogen-activated protein kinase inactivation. *Journal of Biological Chemistry*, 284(11), 6893-6908. <https://doi.org/10.1074/jbc.M805861200>
- Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. In *Genome Medicine* (Vol. 9, Issue 1). BioMed Central Ltd.
<https://doi.org/10.1186/s13073-017-0467-4>
- Hilton, N. A., Thomas, †, Sladewski, E., Perry, J. A., Pataki, Z., Sinclair-Davis, A. N., Muniz, R. S., Tran, H. L., Wurster, J. I., Seo, J., & de Graffenried, C. L. (2018). *Identification of TOEFAZ1-interacting proteins reveals key regulators of Trypanosoma brucei cytokinesis*.
<https://doi.org/10.1111/mmi.13986>
- Holzer, T. R., McMaster, W. R., & Forney, J. D. (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. *Molecular and Biochemical Parasitology*, 146(2), 198-218.
<https://doi.org/10.1016/J.MOLBIOPARA.2005.12.009>
- Holzer, T. R., Mishra, K. K., LeBowitz, J. H., & Forney, J. D. (2008). Coordinate regulation of a family of promastigote-enriched mRNAs by the 3'UTR PRE element in Leishmania mexicana. *Molecular and Biochemical Parasitology*, 157(1), 54-64. <https://doi.org/10.1016/J.MOLBIOPARA.2007.10.001>

- Hombach, S., & Kretz, M. (2016). Non-coding RNAs: Classification, biology and functioning. *Advances in Experimental Medicine and Biology*, 937, 3-17. https://doi.org/10.1007/978-3-319-42059-2_1
- Howick, V. M., Peacock, L., Kay, C., Collett, C., Gibson, W., & Lawniczak, M. K. N. (2022). *Single-cell transcriptomics reveals expression profiles of Trypanosoma brucei sexual stages*. <https://doi.org/10.1371/journal.ppat.1010346>
- Hutchinson, S., Foulon, S., Crouzols, A., Menafra, R., Rotureau, B., Griffiths, A. D., & Bastin, P. (2021). The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of *Trypanosoma brucei* in the tsetse fly salivary glands. *PLoS Pathogens*, 17(9). <https://doi.org/10.1371/journal.ppat.1009904>
- Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50, 96. <https://doi.org/10.1038/s12276-018-0071-8>
- Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology*, 17(1). <https://doi.org/10.1186/s13059-016-0888-1>
- Inbar, E., Hughitt, V. K., Dillon, L. A. L., Ghosh, K., El-Sayed, N. M., & Sacks, D. L. (2017). *The Transcriptome of Leishmania major Developmental Stages in Their Natural Sand Fly Vector*. <http://www.who.int/mediacentre/>
- Isnard, A., Shio, M. T., Olivier, M., Sreevastan, S., & Bengoechea, J. A. (2012). *Impact of Leishmania metalloprotease GP63 on macrophage signaling*. <https://doi.org/10.3389/fcimb.2012.00072>
- Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.-A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., ... Myler, P. J. (2005). *The Genome of the Kinetoplastid Parasite, Leishmania major*. <https://www.science.org>
- Jacquel, A., Obba, S., Boyer, L., Dufies, M., Robert, G., Gounon, P., Lemichez, E., Luciano, F., Solary, E., & Auberger, P. (2012). *Autophagy is required for CSF-1-induced macrophagic differentiation and acquisition of phagocytic functions*. <https://doi.org/10.1182/blood-2011-11-392167>
- Jansen, R., van Embden, J., Gaastra, W., & Schouls, L. (2002). *Identification of genes that are associated with DNA repeats in prokaryotes*. <https://doi.org/10.1046/j.1365-2958.2002.02839.x>
- Jaramillo, M., Gomez, M. A., Larsson, O., Shio, M. T., Topisirovic, I., Contreras, I., Luxenburg, R., Rosenfeld, A., Colina, R., McMaster, R. W., Olivier, M., Costa-Mattioli, M., & Sonenberg, N. (2011). *Leishmania repression of host translation through mTOR cleavage is required for parasite survival and infection*. *Cell Host and Microbe*, 9(4), 331-341. <https://doi.org/10.1016/j.chom.2011.03.008>
- Jiang, F., & Doudna, J. A. (2017). *CRISPR-Cas9 Structures and Mechanisms*. <https://doi.org/10.1146/annurev-biophys>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). *A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity*. <https://www.science.org>
- Jones, N. G., Catta-Preta, C. M. C., Paula, A., Lima, C. A., & Mottram, J. C. (2018). *Genetically Validated Drug Targets in Leishmania: Current Knowledge and Future Prospects*. <https://doi.org/10.1021/acsinfecdis.7b00244>

- Jones, N. G., Geoghegan, V., Moore, G., Carnielli, J. B. T., Newling, K., Calderón, F., Gabarró, R., Martín, J., Prinjha, R. K., Rioja, I., Wilkinson, A. J., & Mottram, J. C. (2022). *Bromodomain factor 5 is an essential regulator of transcription in Leishmania*. <https://doi.org/10.1038/s41467-022-31742-1>
- Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., Luo, Y., Correspondence, Y., Luo, D., & Jovic, L. (2021). *Single-cell RNA sequencing technologies and applications: A brief overview*. <https://doi.org/10.1002/ctm2.694>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583. <https://doi.org/10.1038/s41586-021-03819-2>
- Kalesh, K., Wei, W., Mantilla, B. S., Roumeliotis, T. I., Choudhary, J., & Denny, P. W. (2022). *Transcriptome-Wide Identification of Coding and Noncoding RNA-Binding Proteins Defines the Comprehensive RNA Interactome of Leishmania mexicana*. <https://journals.asm.org/journal/spectrum>
- Kamhawi, S. (2006). Phlebotomine sand flies and Leishmania parasites: friends or foes? *Trends in Parasitology*, 22(9), 439-445. <https://doi.org/10.1016/J.PT.2006.06.012>
- Kamir, D., Zierow, S., et al., & Bucala, R. (2008). *A Leishmania Ortholog of Macrophage Migration Inhibitory Factor Modulates Host Macrophage Responses*. <https://doi.org/10.4049/jimmunol.180.12.8250>
- Karagiannis, K., Gannavaram, S., Verma, C., Bhattacharya, P., Nakhasi, H. L., Satoskar, A., Spring, S., & Satoskar Professor, A. R. (preprint). *Dual scRNA-Seq analysis reveals rare and uncommon parasitized cell populations in chronic L. donovani infection*. <https://doi.org/10.1101/2022.07.26.501600>
- Kaur, P., Anand, A., Bhat, A., Maras, J. S., & Goyal, N. (2021). Comparative phosphoproteomic analysis unravels MAPK1 regulated phosphoproteins in Leishmania donovani. *Journal of Proteomics*, 240, 104189. <https://doi.org/10.1016/J.JPROT.2021.104189>
- Kautz-Neu, K., Schwonberg, K., Fischer, M. R., & Schermann, A. I. (2012). *Dendritic cells in Leishmania major infections: mechanisms of parasite uptake, cell activation and evidence for physiological relevance*. <https://doi.org/10.1007/s00430-012-0261-2>
- Kaye, P., & Scott, P. (2011). Leishmaniasis: complexity at the host-pathogen interface. *Nature Publishing Group*. <https://doi.org/10.1038/nrmicro2608>
- Kelly, J. M., Ward, A. I., Olmo, F., Atherton, R. L., & Taylor, M. C. (2020). *Trypanosoma cruzi amastigotes that persist in the colon during chronic stage murine infections have a reduced replication rate*. <https://doi.org/10.1098/rsob.200261>
- Khan, A. A., Langston, H. C., Id, F. C. C., Olmo Id, F., Taylorid, M. C., Mccannid, C. J., Kellyid, J. M., & Lewisid, M. D. (2021). *Local association of Trypanosoma cruzi chronic infection foci and enteric neuropathic lesions at the tissue micro-domain scale*. <https://doi.org/10.1371/journal.ppat.1009864>
- Kima, P. E. (2007). The amastigote forms of Leishmania are experts at exploiting host cell processes to establish infection and persist. *International Journal for Parasitology*, 37(10), 1087-1096. <https://doi.org/10.1016/J.IJPARA.2007.04.007>
- Kloehn, J., Saunders, E. C., O'callaghan, S., Dagley, M. J., & Mcconville, M. J. (2015). Characterization of Metabolically Quiescent Leishmania Parasites in

- Murine Lesions Using Heavy Water Labeling. *PLoS Pathog*, 11(2), 1004683. <https://doi.org/10.1371/journal.ppat.1004683>
- Kolli, B. K., Kostal, J., Zaborina, O., Chakrabarty, A. M., & Chang, K.-P. (2008). *Leishmania-released nucleoside diphosphate kinase prevents ATP-mediated cytolysis of macrophages*.
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4), 610-620. <https://doi.org/10.1016/J.MOLCEL.2015.04.005>
- Kramer, S., & Carrington, M. (2011). Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids. *Trends in Parasitology*, 27(1), 23-30. <https://doi.org/10.1016/J.PT.2010.06.011>
- Kramer, S., Kimblin, N. C., & Carrington, M. (2010). Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*. In *BMC Genomics* (Vol. 11). <https://doi.org/10.1186/1471-2164-11-283>
- Krishnan, K., & Moens, P. D. J. (2009). *Structure and functions of profilins*. <https://doi.org/10.1007/s12551-009-0010-y>
- Kröber-Boncardo, C., Clos, J., & Grünebast, J. (2020). *Heat Shock Proteins in Leishmania Parasites*. https://doi.org/https://doi.org/10.1007/7515_2020_27
- Kulkarni, A., Anderson, A. G., Merullo, D. P., & Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology*, 58, 129-136. <https://doi.org/10.1016/J.COPBIO.2019.03.001>
- Kulp, T. R., Hoeft, S. E., Asao, M., Madigan, M. T., Hollibaugh, J. T., Fisher, J. C., Stolz, J. F., Culbertson, C. W., Miller, L. G., & Oremland, R. S. (2008). In Vivo Imaging Reveals an Essential Role for Neutrophils in Leishmaniasis Transmitted by Sand Flies. *Science*, 321(5891), 970-974. <https://doi.org/10.1126/science.1160799>
- Kupfer, B., Vehreschild, J., Cornely, O., Kaiser, R., Plum, G., Viazov, S., Franzen, C., Tillmann, R.-L., Simon, A., Müller, A., & Schildgen, O. (2006). Leishmaniasis in Ancient Egypt and Upper Nubia. In *Proc Natl Acad Sci* (Vol. 12, Issue 10). www.cdc.gov/eid
- Kuru, T., Jirata, D., Genetu, A., Barr, S., Mengistu, Y., Aseffa, A., & Gedamu, L. (2007). *Leishmania aethiopica*: Identification and characterization of cathepsin L-like cysteine protease genes. *Experimental Parasitology*, 115(3), 283-290. <https://doi.org/10.1016/J.EXPPARA.2006.09.011>
- Lahav, T., Sivam, D., Volpin, H., Ronen, M., Tsigankov, P., Green, A., Holland, N., Kuzyk, M., Borchers, ¶ C, Zilberstein, ¶ D, & Myler, P. J. (2011). Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *The FASEB Journal • Research Communication*. <https://doi.org/10.1096/fj.10-157529>
- Lainson, R., & Strangways-Dixon, J. (1964). The epidemiology of dermal leishmaniasis in British Honduras: Part II. Reservoir-hosts of *Leishmania mexicana* among the forest rodents. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 58(2), 136-153. [https://doi.org/10.1016/0035-9203\(64\)90003-3](https://doi.org/10.1016/0035-9203(64)90003-3)
- Lakshmi, B. S., Wang, R., & Madhubala, R. (2014). *Leishmania* genome analysis and high-throughput immunological screening identifies tuzin as a novel vaccine candidate against visceral leishmaniasis. *Vaccine*, 32(30), 3816-3822. <https://doi.org/10.1016/J.VACCINE.2014.04.088>

- Lamontagne, J., & Papadopoulou, B. (1999). Developmental regulation of spliced leader RNA gene in *Leishmania donovani* amastigotes is mediated by specific polyadenylation. *Journal of Biological Chemistry*, 274(10), 6602-6609. <https://doi.org/10.1074/jbc.274.10.6602>
- 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(4). <https://doi.org/10.1128/mBio.01012-15>
- Lantz, C., Radmanesh, B., Liu, E., Thorp, E. B., & Lin, J. (2020). *Single-cell RnA sequencing uncovers heterogenous transcriptional signatures in macrophages during efferocytosis*. 10, 14333. <https://doi.org/10.1038/s41598-020-70353-y>
- Larsson, L., Frisén, J., & Lundeberg, J. (2021). *Spatially resolved transcriptomics adds a new dimension to genomics*. <https://doi.org/10.1038/s41592-020-01038-7>
- Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A., & McMaster, W. R. (2007a). Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: The *Leishmania* genome is constitutively expressed. *Molecular and Biochemical Parasitology*, 152(1), 35-46. <https://doi.org/10.1016/J.MOLBIOPARA.2006.11.009>
- Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A., & McMaster, W. R. (2007b). Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: The *Leishmania* genome is constitutively expressed. *Molecular and Biochemical Parasitology*, 152(1), 35-46. <https://doi.org/10.1016/J.MOLBIOPARA.2006.11.009>
- Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir, E. A. D., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., Zunder, E. R., Finck, R., Gedman, A. L., Radtke, I., Downing, J. R., Pe'er, D., & Nolan, G. P. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*, 162(1), 184-197. <https://doi.org/10.1016/j.cell.2015.05.047>
- Ley, K., Laudanna, C., Cybulsky, M. I., & S. N. (2007). *Combinatorial specificity Getting to the site of inflammation: the leukocyte adhesion cascade updated*. <https://doi.org/10.1038/nri2156>
- Li, G.-W., & Sunney Xie, X. (2011). *Central dogma at the single-molecule level in living cells*. <https://doi.org/10.1038/nature10315>
- Li, X., & Wang, C.-Y. (2021). *From bulk, single-cell to spatial RNA sequencing*. <https://doi.org/10.1038/s41368-021-00146-0>
- Liang, X., Gupta, K., Quintero, J. R., Cernadas, M., Kobzik, L., Christou, H., Pier, G. B., Owen, C. A., & Çataltepe, S. (2019). *Macrophage FABP4 is required for neutrophil recruitment and bacterial clearance in Pseudomonas aeruginosa pneumonia*. <https://doi.org/10.1096/fj.201802002R>
- Liang, X. H., Haritan, A., Uliel, S., & Michaeli, S. (2003). trans and cis splicing in trypanosomatids: Mechanism, factors, and regulation. In *Eukaryotic Cell* (Vol. 2, Issue 5, pp. 830-840). <https://doi.org/10.1128/EC.2.5.830-840.2003>
- Lieke, T., Nylén, S., Eidsmo, L., McMaster, W. R., Mohammadi, A. M., Khamesipour, A., Berg, L., & Akuffo, H. (2008). *Leishmania surface protein gp63 binds directly to human natural killer cells and inhibits proliferation*. <https://doi.org/10.1111/j.1365-2249.2008.03687.x>

- Lisack, J., Morriswood, B., & Engstler, M. (2022). *Response to comment on "Unexpected plasticity in the life cycle of Trypanosoma Brucei."*
<https://doi.org/10.7554/eLife>
- Loría-Cervera, E. N., & Andrade-Narvaez, F. (2020). The role of monocytes/macrophages in Leishmania infection: A glance at the human response. *Acta Tropica*, 207, 105456.
<https://doi.org/10.1016/J.ACTATROPICA.2020.105456>
- Louradour, I., Ferreira, T. R., Duge, E., Karunaweera, N., Paun, A., & Sacks, D. (2022). Stress conditions promote Leishmania hybridization in vitro marked by expression of the ancestral gamete fusogen HAP2 as revealed by singlecell RNA-seq. *ELife*, 11. <https://doi.org/10.7554/eLife.73488>
- Lu, L., Li, J., Wei, R., Guidi, I., Cozzuto, L., Ponomarenko, J., Guillem Prats-Ejarque, ·, & Boix, · Ester. (2022). *Selective cleavage of ncRNA and antiviral activity by RNase2/EDN in THP1-induced macrophages*. 79, 209.
<https://doi.org/10.1007/s00018-022-04229-x>
- Lun, A. T. L., Hopkins Bloomberg, J., Ji, H., Mccarthy, D. J., & Marioni, J. C. (2016). *A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor*.
<https://doi.org/10.12688/f1000research.9501.1>
- Lytal, N., Ran, D., & An, L. (2020). *Normalization Methods on Single-Cell RNA-seq Data: An Empirical Survey*. <https://doi.org/10.3389/fgene.2020.00041>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, 161(5), 1202-1214.
<https://doi.org/10.1016/J.CELL.2015.05.002>
- Mahmood, R., Hines, J. C., & Ray, D. S. (1999). Identification of cis and trans Elements Involved in the Cell Cycle Regulation of Multiple Genes in Crithidia fasciculata. In *MOLECULAR AND CELLULAR BIOLOGY* (Vol. 19, Issue 9).
<https://journals.asm.org/journal/mcb>
- Mandell, M. A., & Beverley, S. M. (2017). Continual renewal and replication of persistent Leishmania major parasites in concomitantly immune hosts. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5), E801-E810. <https://doi.org/10.1073/pnas.1619265114>
- Mann, S., Frasca, K., Scherrer, S., Henao-Martínez, A. F., Newman, S., Ramanan, P., & Suarez, J. A. (2021). *A Review of Leishmaniasis: Current Knowledge and Future Directions*. <https://doi.org/10.1007/s40475-021-00232-7/Published>
- Marcia Cardoso de Paiva, R., Grazielle-Silva, V., Santos Cardoso, M., Naemi Nakagaki, B., Pessoa Mendonça-Neto, R., Monte Cassiano Canavaci, A., Souza Melo, N., Massara Martinelli, P., Paula Fernandes, A., Duarte daRocha, W., R Teixeira, S. M., & Beverley, S. M. (2015). *Amastin Knockdown in Leishmania braziliensis Affects Parasite-Macrophage Interaction and Results in Impaired Viability of Intracellular Amastigotes*.
<https://doi.org/10.1371/journal.ppat.1005296>
- Marshall, S., Kelly, P. H., Singh, B. K., Pope, R. M., Kim, P., Zhanbolat, B., Wilson, M. E., & Yao, C. (2018). Extracellular release of virulence factor major surface protease via exosomes in Leishmania infantum promastigotes. *Parasites and Vectors*, 11(1). <https://doi.org/10.1186/s13071-018-2937-y>

- Martínez-Calvillo, S., Nguyen, D., Stuart, K., & Myler, P. J. (2004). Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryotic Cell*, 3(2), 506-517. <https://doi.org/10.1128/EC.3.2.506-517.2004>
- Marx, V. (2021). Method of the Year: spatially resolved transcriptomics. *Nature Methods*, 18(1), 9-14. <https://doi.org/10.1038/s41592-020-01033-y>
- Matheoud, D., Moradin, N., Bellemare-Pelletier, A., Shio, M. T., Hong, W. J., Olivier, M., Gagnon, É., Desjardins, M., & Descoteaux, A. (2013). *Leishmania* evades host immunity by inhibiting antigen cross-presentation through direct cleavage of the SNARE VAMP8. *Cell Host and Microbe*, 14(1), 15-25. <https://doi.org/10.1016/j.chom.2013.06.003>
- Matthews, K. R., & Larcombe, S. (2022). Comment on “Unexpected plasticity in the life cycle of *Trypanosoma brucei*.” 11, 74985. <https://doi.org/10.7554/eLife>
- Matthews, K. R., Tschudi, C., & Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes.
- Mattick, J. S. (2001). Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Reports*, 2(11), 986-991. <https://doi.org/10.1093/embo-reports/kve230>
- McConville, M. J., Turco, S. J., Ferguson, M. A. J., & Sacks, D. L. (1992). Developmental modification of lipophosphoglycan during the differentiation of *Leishmania major* promastigotes to an infectious stage. *EMBO Journal*, 11(10), 3593-3600. <https://doi.org/10.1002/j.1460-2075.1992.tb05443.x>
- McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*.
- Medina, L. S., Araújo Souza, B., Queiroz, A., Henrique Guimarães, L., Roberto Lima Machado, P., Carvalho, E. M., Edythe Wilson, M., & Schriefer, A. (2016). The gp63 Gene Cluster Is Highly Polymorphic in Natural *Leishmania (Viannia) braziliensis* Populations, but Functional Sites Are Conserved. <https://doi.org/10.1371/journal.pone.0163284>
- Medina-Acosta, E., Karess, R. E., Schwartz, H., & Russell, D. G. (1989). The promastigote surface protease (gp63) of *Leishmania* is expressed but differentially processed and localized in the amastigote stage. *Molecular and Biochemical Parasitology*, 37(2), 263-273. [https://doi.org/10.1016/0166-6851\(89\)90158-8](https://doi.org/10.1016/0166-6851(89)90158-8)
- Method of the Year 2013. (2014). <https://doi.org/10.1038/nmeth.2801>
- Mi, H., Muruganujan, A., Casagrande, J., & Thomas, P. (2013). Large-scale gene function analysis with the PANTHER classification system. <https://doi.org/10.1038/nprot.2013.092>
- Mishra, K. K., Holzer, T. R., Moore, L. L., & LeBowitz, J. H. (2003). A negative regulatory element controls mRNA abundance of the *Leishmania mexicana* paraflagellar rod gene PFR2. *Eukaryotic Cell*, 2(5), 1009-1017. <https://doi.org/10.1128/EC.2.5.1009-1017.2003>
- Mojica, F. J. M., Juez, G., & Rodríguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified P_{sfl} sites. *Molecular Microbiology*, 9(3), 613-621. <https://doi.org/10.1111/j.1365-2958.1993.tb01721.x>
- Mojtahedi, Z., Clos, J., & Kamali-Sarvestani, E. (2008). *Leishmania major*: Identification of developmentally regulated proteins in procyclic and metacyclic promastigotes. *Experimental Parasitology*, 119(3), 422-429. <https://doi.org/10.1016/J.EXPPARA.2008.04.008>

- Moon, K. R., David Van Dijk, †, Wang, Z., Chen, † William, Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G., & Krishnaswamy, ‡ Smita. (2017). *PHATE: A Dimensionality Reduction Method for Visualizing Trajectory Structures in High-Dimensional Biological Data*. <https://doi.org/10.1101/120378>
- Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., van den Elzen, A., Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G., & Krishnaswamy, S. (2019). *Visualizing structure and transitions in high-dimensional biological data*. <https://doi.org/10.1038/s41587-019-0336-3>
- Moore, K. J., & Matlashewski, G. (1994). *Intracellular Infection by f eishmania donovani Inhibits*. <http://journals.aai.org/jimmunol/article-pdf/152/6/2930/1066934/2930.pdf>
- Moore, L. L., Santrich, C., & LeBowitz, J. H. (1996). Stage-specific expression of the Leishmania mexicana paraflagellar rod protein PFR-2. *Molecular and Biochemical Parasitology*, 80(2), 125-135. [https://doi.org/10.1016/0166-6851\(96\)02688-6](https://doi.org/10.1016/0166-6851(96)02688-6)
- Moradin, N., Descoteaux, A., Beverley, S. M., & Sinai, A. P. (2012). *Leishmania promastigotes: building a safe niche within macrophages*. <https://doi.org/10.3389/fcimb.2012.00121>
- Mosaad, Y. M. (2015). *Clinical Role of Human Leukocyte Antigen in Health and Disease*. <https://doi.org/10.1111/sji.12329>
- Mosser, D., & Edelson, P. (1985). The mouse macrophage receptor for C3bi (CR3) is a major mechanism in the phagocytosis of Leishmania promastigotes. *J Immunol* . 1985 Oct;135(4):2785-9.
- Mottram, J. C., Coombs, G. H., & Alexander, J. (2004a). Cysteine peptidases as virulence factors of Leishmania. *Current Opinion in Microbiology*, 7(4), 375-381. <https://doi.org/10.1016/J.MIB.2004.06.010>
- Mottram, J. C., Coombs, G. H., & Alexander, J. (2004b). Cysteine peptidases as virulence factors of Leishmania. *Current Opinion in Microbiology*, 7(4), 375-381. <https://doi.org/10.1016/J.MIB.2004.06.010>
- Müller, L. S. M., Cosentino, O., Förstner, K. U., Guizetti, J., Wedel, C., Kaplan, N., Janzen, C. J., Arampatzi, P., Vogel, J., Steinbiss, S., Otto, T. D., Saliba, A.-E., Sebra, P., & Siegel, & T. N. (2018). *Genome organization and DNA accessibility control antigenic variation in trypanosomes*. <https://doi.org/10.1038/s41586-018-0619-8>
- Mundodi, V., Kucknoor, A. S., & Gedamu, L. (2005). Role of Leishmania (Leishmania) chagasi amastigote cysteine protease in intracellular parasite survival: Studies by gene disruption and antisense mRNA inhibition. *BMC Molecular Biology*, 6. <https://doi.org/10.1186/1471-2199-6-3>
- Murray, A., Fu, C., Habibi, G., & McMaster, W. R. (2007). Regions in the 3' untranslated region confer stage-specific expression to the Leishmania mexicana a600-4 gene. *Molecular and Biochemical Parasitology*, 153(2), 125-132. <https://doi.org/10.1016/J.MOLBIOPARA.2007.02.010>
- Myler, P. J., Audleman, L., Devos, T., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Fu ¶, G., Ivens, A., & Stuart, K. (1999). *Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes* (Vol. 96). www.ebi.
- Nalpas, N. C., Magee, D. A., Conlon, K. M., Browne, J. A., Healy, C., McLoughlin, K. E., Rue-Albrecht, K., Mcgettigan, P. A., Killick, K. E., Gormley, E., Gordon, S. v, & Machugh, D. E. (2015). RNA sequencing provides exquisite insight into the manipulation of the alveolar macrophage by tubercle bacilli OPEN. *Nature Publishing Group*. <https://doi.org/10.1038/srep13629>

- Nandan, D., Lo, R., & Reiner, N. E. (1999a). 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*. In *INFECTION AND IMMUNITY* (Vol. 67, Issue 8). <https://journals.asm.org/journal/iai>
- Nandan, D., Lo, R., & Reiner, N. E. (1999b). 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*. In *INFECTION AND IMMUNITY* (Vol. 67, Issue 8). <https://journals.asm.org/journal/iai>
- Nandan, D., & Reiner, N. E. (1995a). Attenuation of Gamma Interferon-Induced Tyrosine Phosphorylation in Mononuclear Phagocytes Infected with *Leishmania donovani*: Selective Inhibition of Signaling through Janus Kinases and Stat1. In *INFECTION AND IMMUNITY* (Vol. 63, Issue 11). <https://journals.asm.org/journal/iai>
- Nandan, D., & Reiner, N. E. (1995b). Attenuation of Gamma Interferon-Induced Tyrosine Phosphorylation in Mononuclear Phagocytes Infected with *Leishmania donovani*: Selective Inhibition of Signaling through Janus Kinases and Stat1. In *INFECTION AND IMMUNITY* (Vol. 63, Issue 11). <https://journals.asm.org/journal/iai>
- Názer, E., & Sánchez, D. O. (2011). *Nucleolar Accumulation of RNA Binding Proteins Induced by ActinomycinD Is Functional in Trypanosoma cruzi and Leishmania mexicana but Not in T. brucei*. <https://doi.org/10.1371/journal.pone.0024184>
- Ng, L. G., Hsu, A., Mandell, M. A., Roediger, B., & Hoeller, C. (2008). Migratory Dermal Dendritic Cells Act as Rapid Sensors of Protozoan Parasites. *PLoS Pathog*, 4(11), 1000222. <https://doi.org/10.1371/journal.ppat.1000222>
- Nishi, T., Shinzawa, N., Yuda, M., & Iwanaga, S. (2021). Highly efficient CRISPR/Cas9 system in *Plasmodium falciparum* using Cas9-expressing parasites and a linear donor template. *Scientific Reports* |, 11, 18501. <https://doi.org/10.1038/s41598-021-97984-z>
- Nunes, V. S., Moretti, N. S., da Silva, M. S., Elias, M. C., Janzen, C. J., & Schenkman, S. (2020). Trimethylation of histone H3K76 by Dot1B enhances cell cycle progression after mitosis in *Trypanosoma cruzi*. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1867(7). <https://doi.org/10.1016/j.bbamcr.2020.118694>
- Okwor, I., & Uzonna, J. (2016). Social and Economic Burden of Human Leishmaniasis. *Am. J. Trop. Med. Hyg*, 94(3), 489-493. <https://doi.org/10.4269/ajtmh.15-0408>
- Olivier, M., Atayde, V. D., Isnard, A., Hassani, K., & Shio, M. T. (2012). *Leishmania* virulence factors: focus on the metalloprotease GP63. *Microbes and Infection*, 14(15), 1377-1389. <https://doi.org/10.1016/J.MICINF.2012.05.014>
- Osorio, D., & Cai, J. J. (2021). Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics*, 37(7), 963-967. <https://doi.org/10.1093/bioinformatics/btaa751>
- Osorio, E. Y., Zhao, W., Espitia, C., Saldarriaga, O., & Hawel, L. (2012). Progressive Visceral Leishmaniasis Is Driven by Dominant Parasite-induced STAT6 Activation and STAT6-dependent Host Arginase 1 Expression. *PLoS Pathog*, 8(1), 1002417. <https://doi.org/10.1371/journal.ppat.1002417>

- Panigrahi, A., & O'Malley, B. W. (2021). Mechanisms of enhancer action: the known and the unknown. In *Genome Biology* (Vol. 22, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13059-021-02322-1>
- Park, M. H., & Wolff, E. C. (2018). Hypusine, a polyamine-derived amino acid critical for eukaryotic translation. In *Journal of Biological Chemistry* (Vol. 293, Issue 48, pp. 18710-18718). American Society for Biochemistry and Molecular Biology Inc. <https://doi.org/10.1074/jbc.TM118.003341>
- Passelli, K., Billion, O., & Tacchini-Cottier, F. (2021). The Impact of Neutrophil Recruitment to the Skin on the Pathology Induced by Leishmania Infection. *Frontiers in Immunology | Www.Frontiersin.Org*, 1, 649348. <https://doi.org/10.3389/fimmu.2021.649348>
- Picelli, S., Bjorklund, A., Faridani, O., Sagasser, S., Winberg, G., & Sandberg, R. (2013). *Smart-seq2 for sensitive full-length transcriptome profiling in single cells*. <https://doi.org/10.1038/nMeth.2639>
- Pires Id, M., Wright, B., Kaye Id, P. M., da Conceiç ão, V., & Churchill, R. C. (2019). *The impact of leishmaniasis on mental health and psychosocial well-being: A systematic review*. <https://doi.org/10.1371/journal.pone.0223313>
- Podinovskaia, M., & Descoteaux, A. (2015). Leishmania and the macrophage: A multifaceted interaction. In *Future Microbiology* (Vol. 10, Issue 1, pp. 111-129). Future Medicine Ltd. <https://doi.org/10.2217/fmb.14.103>
- Pollock, K., McNeil, K., Mottram, J., Lyons, R., Brewer, J., Scott, P., Coombs, G., & Alexander, J. (2003). *The Leishmania mexicana Cysteine Protease, CPB2.8, Induces Potent Th2 Responses*. <https://doi.org/10.4049/jimmunol.170.4.1746>
- Prasanna, P., & Upadhyay, A. (2021). Heat shock proteins as the druggable targets in leishmaniasis: Promises and perils. *Infection and Immunity*, 89(2). <https://doi.org/10.1128/IAI.00559-20>
- Proudfoot, L., Liew, F., & O'donnell, C. A. (1995). Glycoinositolphospholipids inhibit NO synthetis 745 Glycoinositolphospholipids of Leishmania major inhibit nitric oxide synthesis and reduce. *Eur. J. Immunol*, 25, 745-750. <https://doi.org/10.1002/eji.1830250318>
- Ramanathan, M., Porter, D. F., & Khavari, P. A. (2019). *Methods to study RNA-protein interactions*. <https://doi.org/10.1038/s41592-019-0330-1>
- Ramírez, C. A., Requena, J. M., & Puerta, C. J. (2013). Alpha tubulin genes from Leishmania braziliensis: Genomic organization, gene structure and insights on their expression. *BMC Genomics*, 14(1). <https://doi.org/10.1186/1471-2164-14-454>
- Ramos, C. S., Franco, F. A. L., Smith, D. F., & Uliana, S. R. B. (2004). *Characterisation of a new Leishmania META gene and genomic analysis of the META cluster*. <https://doi.org/10.1016/j.femsle.2004.07.037>
- Ranjan Dikhit, M., Das, S., Mahantesh, V., Kumar, A., Singh, A. K., Dehury, B., Rout, A. K., Ali, V., Sahoo, G. C., Kamal Topno, R., Pandey, K., Das, V. N. R., Bimal, S., & Das, P. (2018). The potential HLA Class I-restricted epitopes derived from LelF and TSA of Leishmania donovani evoke anti-leishmania CD8+ T lymphocyte response OPEN. *SCIEntIFIC RePoRTS* |, 8, 14175. <https://doi.org/10.1038/s41598-018-32040-x>
- Rashidi, S., Mansouri, R., Ali-Hassanzadeh, M., Mojtahedi, Z., Shafiei, R., Savardashtaki, A., Hamidizadeh, N., Karimazar, M., Nguewa, P., & Manzano-Román, R. (2021). *The host mTOR pathway and parasitic diseases pathogenesis*. <https://doi.org/10.1007/s00436-021-07070-6/Published>


- Ray, M., Gam, A. A., Boykins, R. A., & Kenney, R. T. (2000). *Inhibition of Interferon-g Signaling by Leishmania donovani*. <https://academic.oup.com/jid/article/181/3/1121/909895>
- Real, F., Mortara, R. A., Rabinovitch, M., & Jones, M. K. (2010). *Fusion between Leishmania amazonensis and Leishmania major Parasitophorous Vacuoles: Live Imaging of Coinfected Macrophages*. <https://doi.org/10.1371/journal.pntd.0000905>
- Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., R Illingworth, C. J., Billker, O., Berriman, M., & Lawniczak, M. K. (2018). *Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites*. <https://doi.org/10.7554/eLife.33105.001>
- Reverte, M., Onur, R. E., Jha, B., Desponds, C., Snä Kä Id, T., Prevel, F., Isorceid, N., Lyeid, L.-F., Owens, K. L., Lopes, U. G., Beverleyid, S. M., & Faselid, N. (2021). *The antioxidant response favors Leishmania parasites survival, limits inflammation and reprograms the host cell metabolism*. <https://doi.org/10.1371/journal.ppat.1009422>
- Rico, E., Rojas, F., Mony, B. M., Szoor, B., Macgregor, P., & Matthews, K. R. (2013). *Bloodstream form pre-adaptation to the tsetse fly in Trypanosoma brucei*. <https://doi.org/10.3389/fcimb.2013.00078>
- Rizzo, G., Vafadarnejad, E., Arampatzi, P., Silvestre, J.-S., Zerneck, A., Saliba, A.-E., & Cochain, C. (2020). *Single-cell transcriptomic profiling maps monocyte/macrophage transitions after myocardial infarction in mice*. <https://doi.org/10.1101/2020.04.14.040451>
- Rogers, M. B., Hilley, J. D., Dickens, N. J., Wilkes, J., Bates, P. A., Depledge, D. P., Harris, D., Her, Y., Herzyk, P., Imamura, H., Otto, T. D., Sanders, M., Seeger, K., Dujardin, J. C., Berriman, M., Smith, D. F., Hertz-Fowler, C., & Mottram, J. C. (2011). Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Research*, 21(12), 2129-2142. <https://doi.org/10.1101/gr.122945.111>
- Rogers, M. E., & Bates, P. A. (2007). *Leishmania Manipulation of Sand Fly Feeding Behavior Results in Enhanced Transmission*. <https://doi.org/10.1371/journal.ppat>
- Rogers, M. E., Chance, M. L., & Bates, P. A. (2002). *The role of promastigote secretory gel in the origin and transmission of the infective stage of Leishmania mexicana by the sandfly Lutzomyia longipalpis*. <https://doi.org/10.1017/S0031182002001439>
- Rogers, M., Kropf, P., Choi, B.-S., Dillon, R., Podinovskaia, M., Bates, P., & Müller, I. (2009). *Proteophosphoglycans Regurgitated by Leishmania-Infected Sand Flies Target the L-Arginine Metabolism of Host Macrophages to Promote Parasite Survival*. <https://doi.org/10.1371/journal.ppat.1000555>
- Roque, A. L. R., & Jansen, A. M. (2014). Wild and synanthropic reservoirs of Leishmania species in the Americas. In *International Journal for Parasitology: Parasites and Wildlife* (Vol. 3, Issue 3, pp. 251-262). Australian Society for Parasitology. <https://doi.org/10.1016/j.ijppaw.2014.08.004>
- Rosenzweig, D., Smith, D., Opperdoes, F., Stern, S., Olafson, R. W., & Zilberstein, D. (2008). *Retooling Leishmania metabolism: from sand fly gut to human macrophage; Retooling Leishmania metabolism: from sand fly gut to human macrophage*. <https://doi.org/10.1096/fj.07-9254com>
- Rougeron, V., de Meeûs, T., & Bañuls, A. L. (2017). Reproduction in Leishmania: A focus on genetic exchange. *Infection, Genetics and Evolution*, 50, 128-132. <https://doi.org/10.1016/J.MEEGID.2016.10.013>

- Sacks, D., & Kamhawi, S. (2001). *MOLECULAR ASPECTS OF PARASITE-VECTOR AND VECTOR-HOST INTERACTIONS IN LEISHMANIASIS 1*.
www.annualreviews.org
- Sacks, D. L., & Perkins, P. v. (1984). Identification of an infective stage of *Leishmania promastigotes*. *Science*, 223(4643), 1417-1419.
<https://doi.org/10.1126/science.6701528>
- Sacks, D., & Noben-Trauth, N. (2002). *THE IMMUNOLOGY OF SUSCEPTIBILITY AND RESISTANCE TO LEISHMANIA MAJOR IN MICE*. <https://doi.org/10.1038/nri933>
- Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). Data mining and Modelling for A comparison of single-cell trajectory inference methods. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-019-0071-9>
- Salei, N., Hellberg, L., Rg Kö Hl, J., & Laskay, T. (2017). *Enhanced survival of Leishmania major in neutrophil granulocytes in the presence of apoptotic cells*. <https://doi.org/10.1371/journal.pone.0171850>
- Salih, M., Fakiola, M., Lyons, P., Younis, B., Musa, A., Elhassan, A., Anderson, D., Syn, G., Ibrahim, M., Blackwell, J., & Mohamed, H. (2017). *Expression profiling of Sudanese visceral leishmaniasis patients pre- and post-treatment with sodium stibogluconate*. <https://doi.org/10.1111/pim.12431>
- Sánchez-Valdés, F. J., Padilla, A., Wang, W., Orr, D., & Tarleton, R. L. (2018). *Spontaneous dormancy protects Trypanosoma cruzi during extended drug exposure*. <https://doi.org/10.7554/eLife.34039.001>
- Sandoval Pacheco, C. M., Flores, G. V. A., Gonzalez, K., De, C. M., Gomes, C., Passero, L. F. D., Tomokane, T. Y., Sosa-Ochoa, W., Zúniga, C., Calzada, J., Saldaña, A., Corbett, C. E. P., Silveira, F. T., & Laurenti, M. D. (2021). Macrophage Polarization in the Skin Lesion Caused by Neotropical Species of *Leishmania* sp. *Hindawi Journal of Immunology Research*, 2021.
<https://doi.org/10.1155/2021/5596876>
- Sansbury, B. M., Hewes, A. M., & Kmieć, E. B. (2019). *Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair*. <https://doi.org/10.1038/s42003-019-0705-y>
- Santos, T. T. O., Machado, A. S., Ramos, F. F., Oliveira-da-Silva, J. A., Lage, D. P., Tavares, G. S. V., Mendonça, D. V. C., Cardoso, M. S., Siqueira, W. F., Martins, V. T., Ludolf, F., Reis, T. A. R., Carvalho, L. M., Freitas, C. S., Bandeira, R. S., Silva, A. M., Oliveira, J. S., Moreira, R. L. F., Fujiwara, R. T., ... Coelho, E. A. F. (2021). *Leishmania eukaryotic elongation Factor-1 beta protein is immunogenic and induces parasitological protection in mice against Leishmania infantum infection*. *Microbial Pathogenesis*, 151, 104745.
<https://doi.org/10.1016/J.MICPATH.2021.104745>
- Sasidharan, S., & Prakash Saudagar, &. (2021). *Leishmaniasis: where are we and where are we heading?* <https://doi.org/10.1007/s00436-021-07139-2/Published>
- Sasindran, S. J., Torrelles, J. B., & Bengoechea, J. A. (2011). *Mycobacterium tuberculosis infection and inflammation: what is beneficial for the host and for the bacterium?* 2(1). <https://doi.org/10.3389/fmicb.2011.00002>
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* VOLUME, 33. <https://doi.org/10.1038/nbt.3192>
- Saxena, A., Lahav, T., Holland, N., Aggarwal, G., Anupama, A., Huang, Y., Volpin, H., Myler, P. J., & Zilberstein, D. (2007). Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Molecular and*

- Biochemical Parasitology*, 152(1), 53-65.
<https://doi.org/10.1016/J.MOLBIOPARA.2006.11.011>
- Schalkwijk, J., Allman, E. L., M Jansen, P. A., de Vries, L. E., J Verhoef, J. M., Jackowski, S., M Botman, P. N., Beuckens-Schortinghuis, C. A., J Koolen, K. M., Bolscher, J. M., Vos, M. W., Miller, K., Reeves, S. A., Pett, H., Trevitt, G., Wittlin, S., Scheurer, C., Sax, S., Fischli, C., ... Dechering, K. J. (2019). Antimalarial pantothenamide metabolites target acetyl-coenzyme A biosynthesis in *Plasmodium falciparum*. In *Sci. Transl. Med* (Vol. 11).
<https://www.science.org>
- Schindelin, J., et al., & Cardona, A. (2012). *Fiji: an open-source platform for biological-image analysis*. <https://doi.org/10.1038/nmeth.2019>
- Schlein, Y., Jacobson, R. L., & Shlomai, J. (1991). *Chitinase secreted by Leishmania functions in the sandfly vector*. <https://doi.org/DOI:10.1098/rspb.1991.0097>
- Schmid-Hempel, P. (2009). *Immune defence, parasite evasion strategies and their relevance for “macroscopic phenomena” such as virulence*.
<https://doi.org/10.1098/rstb.2008.0157>
- Schneider-Poetsch, T., & Yoshida, M. (2018). *Along the Central Dogma-Controlling Gene Expression with Small Molecules*.
<https://doi.org/10.1146/annurev-biochem>
- Schuster, S., Lisack, J., Subota, I., Zimmermann, H., Reuter, C., Müller, T., Morriswood, B., & Engstler, M. (2021). Unexpected plasticity in the life cycle of *trypanosoma brucei*. *ELife*, 10.
<https://doi.org/10.7554/ELIFE.66028>
- Schwabl, P., Imamura, H., van den Broeck, F., Costales, J. A., Manguashca-Sánchez, J., Miles, M. A., Andersson, B., Grijalva, M. J., & Llewellyn, M. S. (2019). Meiotic sex in Chagas disease parasite *Trypanosoma cruzi*. *Nature Comms*. <https://doi.org/10.1038/s41467-019-11771-z>
- Scianimanico, S., Desrosiers, M., Dermine, J.-F. É. O., Phane Me, S. Â., Descoteaux, A., & Desjardins, M. (1999). *Impaired recruitment of the small GTPase rab7 correlates with the inhibition of phagosome maturation by Leishmania donovani promastigotes*. <https://doi.org/10.1046/j.1462-5822.1999.00002.x>
- Scott, P., & Novais, F. O. (2016). Cutaneous leishmaniasis: Immune responses in protection and pathogenesis. In *Nature Reviews Immunology* (Vol. 16, Issue 9, pp. 581-592). Nature Publishing Group.
<https://doi.org/10.1038/nri.2016.72>
- Seay, M. B., Heard, P. L., & Chaudhuri, G. (1996). Surface Zn-Proteinase as a Molecule for Defense of *Leishmania mexicana amazonensis* Promastigotes against Cytolysis inside Macrophage Phagolysosomes. In *INFECTION AND IMMUNITY* (Vol. 64, Issue 12). <https://journals.asm.org/journal/iai>
- See, P., Lum, J., Chen J, & Ginhoux F. (2018). A Single-Cell Sequencing Guide for Immunologists. *A Single-Cell Sequencing Guide for Immunologists. Front. Immunol*, 9, 2425. <https://doi.org/10.3389/fimmu.2018.02425>
- Séguin, O., & Descoteaux, A. (2016). *Leishmania, the phagosome, and host responses: The journey of a parasite. Cellular Immunology*, 309, 1-6.
<https://doi.org/10.1016/J.CELLIMM.2016.08.004>
- Sen, R., Bandyopadhyay, S., Dutta, A., Mandal, G., Ganguly, S., Saha, P., & Chatterjee, M. (2007). Artemisinin triggers induction of cell-cycle arrest and apoptosis in *Leishmania donovani* promastigotes. *Journal of Medical Microbiology*, 56(9), 1213-1218. <https://doi.org/10.1099/jmm.0.47364-0>

- Serafim, T. D., Coutinho-Abreu, I. v., Oliveira, F., Meneses, C., Kamhawi, S., & Valenzuela, J. G. (2018). Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nature Microbiology*. <https://doi.org/10.1038/s41564-018-0125-7>
- Serafim, T. D., Figueiredo, A. B., Costa, P., Marques-Da-Silva, E. A., & Gonçalves, R. (2012). *Leishmania* Metacyclogenesis Is Promoted in the Absence of Purines. *PLoS Negl Trop Dis*, 6(9), 1833. <https://doi.org/10.1371/journal.pntd.0001833>
- Shapira, M., McEwen, J. G., & Jaffe, C. L. (1988). Temperature effects on molecular processes which lead to stage differentiation in *Leishmania*. *The EMBO Journal*, 7(9), 2895-2901. <https://doi.org/10.1002/j.1460-2075.1988.tb03147.x>
- Shapiro, S., Naessens, J., Liesegang, B., Moloo, S., & Magundu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *International Laboratory for Research on Animal Diseases (ILRAD)*.
- Shio, M. T., Hassani, K., Isnard, A., Ralph, B., Contreras, I., Gomez, M. A., Abu-Dayyeh, I., & Olivier, M. (2012). Host Cell Signalling and *Leishmania* Mechanisms of Evasion. *Journal of Tropical Medicine*, 2012, 14. <https://doi.org/10.1155/2012/819512>
- Sibley, D. L., Krahenbuhl, J. L., Mike, G., Adams, W., & Weidner, E. (1986). *Toxoplasma* Modifies Macrophage Phagosomes by Secretion of a Vesicular Network Rich in Surface Proteins. <http://rupress.org/jcb/article-pdf/103/3/867/1053954/867.pdf>
- Sinclair-Davis, A. N., McAllaster, M. R., & de Graffenried, C. L. (2017). A functional analysis of TOEFAZ1 uncovers protein domains essential for cytokinesis in *Trypanosoma brucei*. *Journal of Cell Science*, 130(22), 3918-3932. <https://doi.org/10.1242/jcs.207209>
- Singh, O. P., Hasker, E., Sacks, D., Boelaert, M., & Sundar, S. (2014). *Asymptomatic Leishmania Infection: A New Challenge for Leishmania Control*. <https://doi.org/10.1093/cid/ciu102>
- Solletis, L., Ghorbal, M., Macpherson, C. R., Martins, R. M., Kuk, N., Crobu, L., Bastien, P., Scherf, A., Lopez-Rubio, J.-J., & Sterkers, Y. (2015). *Breaking eport First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites*. <https://doi.org/10.1111/cmi.12456>
- Sopwith, W. F., Debrabant, A., Yamage, M., Dwyer, D. M., & Bates, P. A. (2002). Developmentally regulated expression of a cell surface class I nuclease in *Leishmania mexicana*. *International Journal for Parasitology*, 32(4), 449-459. [https://doi.org/10.1016/S0020-7519\(01\)00372-1](https://doi.org/10.1016/S0020-7519(01)00372-1)
- Srikumar, S., Kröger, C., Hébrard, M., Colgan, A., Owen, S. v., Sivasankaran, S. K., S Cameron, A. D., Hokamp, K., & D Hinton, J. C. (2015). *RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium*. <https://doi.org/10.1371/journal.ppat.1005262>
- Stahl, P., et al., & Frisen, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294), 74-78. <https://doi.org/10.1126/science.aaf4374>
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., & Dudoit, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1). <https://doi.org/10.1186/s12864-018-4772-0>
- Stuart, T., & Satija, R. (2019). *Integrative single-cell analysis*. <https://doi.org/10.1038/s41576-019-0093-7>

- Subbian, S., Tsenova, L., Kim, M. J., Wainwright, H. C., Visser, A., Bandyopadhyay, N., Bader, J. S., Karakousis, P. C., Murrmann, G. B., Bekker, L. G., Russell, D. G., & Kaplan, G. (2015). Lesion-specific immune response in granulomas of patients with pulmonary tuberculosis: A pilot study. *PLoS ONE*, *10*(7). <https://doi.org/10.1371/journal.pone.0132249>
- Sudol, M., Chen, H. I., Bougeret, C., & Einbond, A. (1995). Characterization of a novel protein-binding module-the WW domain. *FEBS 15622 FEBS Letters*, *369*, 71. [https://doi.org/10.1016/0014-5793\(95\)00550-5](https://doi.org/10.1016/0014-5793(95)00550-5)
- Sunter, J., & Gull, K. (2017). *Shape, form, function and Leishmania pathogenicity: from textbook descriptions to biological understanding*. <https://doi.org/10.1098/rsob.170165>
- Supek, F., Bošnjak, M., Kunca, S., & Muc, S. (2011). Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), 21800. <https://doi.org/10.1371/journal.pone.0021800>
- Talamas-Rohana, P., Wright, S. D., Lennartz, M. R., & Russell, D. G. (1990). LIPOPHOSPHOGLYCAN FROM *Leishmania mexicana* PROMASTIGOTES BINDS LEUKOCYTE INTEGRINS TO MEMBERS OF THE CR3, p150,95 AND LFA-1 FAMILY OF. <http://journals.aai.org/jimmunol/article-pdf/144/12/4817/1049863/4817.pdf>
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, A. (2008). mRNA-Seq whole-transcriptome analysis of a single cell. *DETAILS ONLINE*, *6*(5). <https://doi.org/10.1038/NMETH.1315>
- Taylor, W. R. (2006). *Transcription and translation in an RNA world*. <https://doi.org/10.1098/rstb.2006.1910>
- Thind, A. S., Monga, I., Thakur, P. K., Kumari, P., Dindhoria, K., Krzak, M., Ranson, M., & Ashford, B. (2021). Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. In *Briefings in bioinformatics* (Vol. 22, Issue 6). NLM (Medline). <https://doi.org/10.1093/bib/bbab259>
- Tjhin, E. T., Spry, C., Sewell, A. L., Hoegl, A., Barnard, L., Sexton, A. E., Siddiqui, G., Howieson, V. M., Maier, A. G., Creek, D. J., Strauss, E., Marquez, R., Auclair, K., & Saliba, K. J. (2018). *Mutations in the pantothenate kinase of Plasmodium falciparum confer diverse sensitivity profiles to antiplasmodial pantothenate analogues*. <https://doi.org/10.1371/journal.ppat.1006918>
- Torres, F., Arias-Carrasco, R., Caris-Maldonado, J. C., Barral, A., Maracaja-Coutinho, V., & de Queiroz, A. T. L. (2017). LeishDB: A database of coding gene annotation and non-coding RNAs in *Leishmania braziliensis*. *Database*, *2017*. <https://doi.org/10.1093/database/bax047>
- Torres-Guerrero, E., Romano Quintanilla-Cedillo, M., Ruiz-Esmenjaud, J., & Arena, R. (2017). *Leishmaniasis: a review*. <https://doi.org/10.12688/f1000research.11120.1>
- Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S. A., Pinto, A., van den Abbeele, J., Ribeiro, R. M., Dias, S., Smith, T. K., & Figueiredo, L. M. (2016). Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. *Cell Host and Microbe*, *19*(6), 837-848. <https://doi.org/10.1016/j.chom.2016.05.002>
- Tsigankov, P., Gherardini, P. F., Helmer-Citterich, M., Späth, G. F., & Zilberstein, D. (2013). Phosphoproteomic analysis of differentiating *Leishmania* parasites reveals a unique stage-specific phosphorylation motif.

- Journal of Proteome Research*, 12(7), 3405-3412.
<https://doi.org/10.1021/pr4002492>
- Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T., & Tada, K. (1980). ESTABLISHMENT AND CHARACTERIZATION OF A HUMAN ACUTE MONOCYTIC LEUKEMIA CELL LINE (THP-1). *J. Cancer*, 26, 171-176.
<https://doi.org/10.1002/ijc.2910260208>
- Tzfati, Y., Abeliovich, H., Avrahami, D., & Shlomai, J. (1995). Universal Minicircle Sequence Binding Protein, a CCHC-type Zinc Finger Protein That Binds the Universal Minicircle Sequence of Trypanosomatids: PURIFICATION AND CHARACTERIZATION. *Journal of Biological Chemistry*, 270(36), 21339-21345. <https://doi.org/10.1074/JBC.270.36.21339>
- Ueno, N., & Wilson, M. E. (2012). Receptor-mediated phagocytosis of Leishmania: Implications for intracellular survival. In *Trends in Parasitology* (Vol. 28, Issue 8, pp. 335-344). <https://doi.org/10.1016/j.pt.2012.05.002>
- Utzinger, J., Becker, S. L., Knopp, S., Blum, J., Neumayr, A. L., Keiser, J., & Hatz, C. F. (2012). *Neglected tropical diseases: diagnosis, clinical management, treatment and control*.
<https://doi.org/10.4414/smwm.2012.13727>
- Uzcanga, G., Lara, E., Gutiérrez, F., Beaty, D., Beske, T., Teran, R., Navarro, J.-C., Pasero, P., Benítez, W., & Poveda, A. (2016). Critical Reviews in Microbiology Nuclear DNA replication and repair in parasites of the genus Leishmania: Exploiting differences to develop innovative therapeutic approaches. *Critical Reviews in Microbiology*, 43(2), 156-177.
<https://doi.org/10.1080/1040841X.2016.1188758>
- van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). *Trajectory-based differential expression analysis for single-cell sequencing data*.
<https://doi.org/10.1038/s41467-020-14766-3>
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., ... Velankar, S. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50.
<https://doi.org/10.1093/nar/gkab1061>
- Veinotte, L., Gebremeskel, S., & Johnston, B. (2016). CXCL16-positive dendritic cells enhance invariant natural killer T cell-dependent IFN γ production and tumor control. *Oncolmmunology*, 5.
<https://doi.org/10.1080/2162402X.2016.1160979>
- Venugopal, G., 1 , I. D., Bird Id, J. T., Washam Id, C. L., Roys Id, H., Bowlin Id, A., Byrum, S. D., & Weinkopff Id, T. (2022). *In vivo transcriptional analysis of mice infected with Leishmania major unveils cellular heterogeneity and altered transcriptomic profiling at single-cell resolution*.
<https://doi.org/10.1371/journal.pntd.0010518>
- Vickerman, K. (1985). DEVELOPMENTAL CYCLES AND BIOLOGY OF PATHOGENIC TRYPANOSOMES. In *British Medical Bulletin* (Vol. 41, Issue 2).
<https://academic.oup.com/bmb/article/41/2/105/296814>
- Vigneron, A., O'Neill, M. B., Weiss, B. L., Savage, A. F., Campbell, O. C., Kamhawi, S., Valenzuela, J. G., & Aksoy, S. (2020). Single-cell RNA sequencing of Trypanosoma brucei from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *PNAS*. <https://doi.org/10.1073/pnas.1914423117/-/DCSupplemental>

- Volf, P., Svobodova, M., & Dvorakova, E. (2001). *Bloodmeal digestion and Leishmania major infections in Phlebotomus duboscqi: effect of carbohydrates inhibiting midgut lectin activity*. <https://doi.org/10.1046/j.0269-283x.2001.00308.x>
- Walker, J., Vasquez, J. J., Gomez, M. A., 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. *Molecular and Biochemical Parasitology*, 147(1), 64-73. <https://doi.org/10.1016/J.MOLBIOPARA.2006.01.008>
- Wang, H., Sham, P., Tong, T., Pang, H., Wang, H., & Pang, H. (2020). Pathway-Based Single-Cell RNA-Seq Classification, Clustering, and Construction of Gene-Gene Interactions Networks Using Random Forests; Pathway-Based Single-Cell RNA-Seq Classification, Clustering, and Construction of Gene-Gene Interactions Networks Using Random Forests. *IEEE Journal of Biomedical and Health Informatics*, 24(6). <https://doi.org/10.1109/JBHI.2019.2944865>
- Ward, A. I., Lewis, M. D., Khan, A. A., McCann, C. J., Francisco, A. F., Jayawardhana, S., Taylor, M. C., & Kelly, J. M. (2020). In vivo analysis of trypanosoma cruzi persistence foci at single-cell resolution. *MBio*, 11(4), 1-13. <https://doi.org/10.1128/mBio.01242-20>
- Wauters, E., van Mol, P., Dinkarnath Garg, A., Jansen, S., van Herck, Y., Vanderbeke, L., Bassez, A., Boeckx, B., Malengier-Devlies, B., Timmerman, A., van Brussel, T., van Buyten, T., Schepers, R., Heylen, E., Dauwe, D., Doms, C., Gunst, J., Hermans, G., Meersseman, P., ... Lambrechts, D. (2021). *ARTICLE Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages*. <https://doi.org/10.1038/s41422-020-00455-9>
- Weih, F., Carrasco, D., Durham, S. K., Barton, D. S., Rizzo, C. A., Ryseck, R. P., Lira, S. A., & Bravo, R. (1995). Multiorgan inflammation and hematopoietic abnormalities in mice with a targeted disruption of RelB, a member of the NF- κ B/Rel family. *Cell*, 80(2), 331-340. [https://doi.org/10.1016/0092-8674\(95\)90416-6](https://doi.org/10.1016/0092-8674(95)90416-6)
- Weingessel, A. (2015). *Package "princurve" Title Fits a Principal Curve in Arbitrary Dimension*.
- Wenzel, U. A., Bank, E., Florian, C., Förster, S., Zimara, N., Steinacker, J., Klinger, M., Reiling, N., Ritter, U., & Zandbergen, G. (2012). *Leishmania major* parasite stage-dependent host cell invasion and immune evasion. *The FASEB Journal*, 26(1), 29-39. <https://doi.org/10.1096/fj.11-184895>
- Westrop, G. D., M Williams, R. A., Wang, L., Zhang, T., Watson, D. G., Marta Silva, A., & Coombs, G. H. (2015). *Metabolomic Analyses of Leishmania Reveal Multiple Species Differences and Large Differences in Amino Acid Metabolism*. <https://doi.org/10.1371/journal.pone.0136891>
- Wetzel, D. M., McMahon-Pratt, D., & Koleske, A. J. (2012). The Abl and Arg Kinases Mediate Distinct Modes of Phagocytosis and Are Required for Maximal *Leishmania* Infection. *Molecular and Cellular Biology*, 32(15), 3176-3186. <https://doi.org/10.1128/mcb.00086-12>
- Wetzel, D. M., Rhodes, E. L., Li, S., McMahon-Pratt, D., & Koleske, A. J. (2016). The Src kinases Hck, Fgr and Lyn activate Arg to facilitate IgG-mediated phagocytosis and *Leishmania* infection. *Journal of Cell Science*, 129(16), 3130-3143. <https://doi.org/10.1242/jcs.185595>

- Wheeler, R. J., Gluenz, E., & Gull, K. (2011). *The cell cycle of Leishmania: morphogenetic events and their implications for parasite biology* *mi_7479* 647..662. <https://doi.org/10.1111/j.1365-2958.2010.07479.x>
- Wheeler, R. J., Gluenz, E., & Gull, K. (2015). *Basal body multipotency and axonemal remodelling are two pathways to a 9+0 flagellum*. <https://doi.org/10.1038/ncomms9964>
- WHO. (2015). *Visceral leishmaniasis control strategies and epidemiological situation update in East Africa report of a WHO bi-regional consultation Addis Ababa, Ethiopia*. https://apps.who.int/iris/bitstream/handle/10665/190168/9789241509657_eng.pdf?sequence=1
- WHO. (2018a). *Department of control of neglected tropical diseases recognizing neglected tropical diseases through changes on the skin*. http://www.who.int/neglected_diseases/en
- WHO. (2018b). Global leishmania surveillance update, 1998-2016. *Weekly Epidemiology Record* 40(93), 521-540.
- WHO Expert Committee on the Control of the Leishmaniases. Meeting (2010 : Geneva), & World Health Organization. (2010). *Control of the leishmaniases : report of a meeting of the WHO Expert Committee on the Control of Leishmaniases, Geneva, 22-26 March 2010*. World Health Organization.
- Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., & Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. In *Genome Medicine* (Vol. 14, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13073-022-01075-1>
- Wilson, R., Bates, M. D., Dostalova, A., Jecna, L., Dillon, R. J., Volf, P., & Bates, P. A. (2010). *Stage-Specific Adhesion of Leishmania Promastigotes to Sand Fly Midguts Assessed Using an Improved Comparative Binding Assay*. <https://doi.org/10.1371/journal.pntd.0000816>
- Woelbing, F., Kostka, S. L., Moelle, K., Belkaid, Y., Sunderkoetter, C., Verbeek, S., Waisman, A., Nigg, A. P., Knop, J., Udey, M. C., & von Stebut, E. (2006). Uptake of *Leishmania major* by dendritic cells is mediated by Fcγ receptors and facilitates acquisition of protective immunity. *Journal of Experimental Medicine*, 203(1), 177-188. <https://doi.org/10.1084/jem.20052288>
- Worthey, E. A., Martinez-Calvillo, S., Schnauffer, A., Aggarwal, G., Cawthra, J., Fazelinia, G., Fong, C., Fu, G., Hassebrock, M., Hixson, G., Ivens, A. C., Kiser, P., Marsolini, F., Rickell, E., Salavati, R., Sisk, E., Sunkin, S. M., Stuart, K. D., & Myler, P. J. (2003). *Leishmania major chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene*. <https://doi.org/10.1093/nar/gkg469>
- Wright, P., & Roufaie, E. A. M. (1989). *Leishmania infection: surfaces and immunity*.
- Xin, Y., Kim, J., Ni, M., Wei, Y., Okamoto, H., Lee, J., Adler, C., Cavino, K., Murphy, A. J., Yancopoulos, G. D., Lin, H. C., & Gromada, J. (2016). Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12), 3293-3298. <https://doi.org/10.1073/pnas.1602306113>
- Xu, C., & Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, 31(12), 1974-1980. <https://doi.org/10.1093/bioinformatics/btv088>

- Yagoubat, A., Crobu, L., Berry, | Laurence, Kuk, N., Lefebvre, M., Sarrazin, A., Patrick Bastien, |, & Sterkers, Y. (2020). *Universal highly efficient conditional knockout system in Leishmania, with a focus on untranscribed region preservation*. <https://doi.org/10.1111/cmi.13159>
- Yao, C., Donelson, J. E., & Wilson, M. E. (2003). The major surface protease (MSP or GP63) of *Leishmania* sp. Biosynthesis, regulation of expression, and function. *Molecular and Biochemical Parasitology*, 132(1), 1-16. [https://doi.org/10.1016/S0166-6851\(03\)00211-1](https://doi.org/10.1016/S0166-6851(03)00211-1)
- Yasmin, H., Goodman, A. G., Bautista, J., Sanctis, D., Adhikary, A., Al-Ahdal, M. N., Roy, S., & Kishore, U. (2022). *Host-Pathogen Interaction in Leishmaniasis: Immune Response and Vaccination Strategies*. <https://doi.org/10.3390/immuno2010015>
- Yeo, G. H. T., Saksena, S. D., & Gifford, D. K. (2020). *Generative modeling of single-cell population time series for inferring cell differentiation landscapes*. <https://doi.org/10.1101/2020.08.26.269332>
- Yip, S. H., Sham, P. C., & Wang, J. (2018). Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. *Briefings in Bioinformatics*, 20(4), 1583-1589. <https://doi.org/10.1093/bib/bby011>
- Young, J., Dominicus, C., Wagener, J., Butterworth, S., Ye, X., Kelly, G., Ordan, M., Saunders, B., Instrell, R., Howell, M., Stewart, A., & Treeck, M. (2019). *A CRISPR platform for targeted in vivo screens identifies Toxoplasma gondii virulence factors in mice*. <https://doi.org/10.1038/s41467-019-11855-w>
- Young, J., & Kima, P. E. (2019). The *Leishmania* Parasitophorous Vacuole Membrane at the Parasite-Host Interface. In *YALE JOURNAL OF BIOLOGY AND MEDICINE* (Vol. 92).
- Zakharova, A., Albanaz, A. T. S., Opperdoes, F. R., Škodová-Sveráková, I., Zagirova, D., Saura, A., Chmelová, L., Gerasimov, E. S., Leštinová, T., Bečvář, T., Sádlová, J., Volf, P., Lukeš, J., Horváth, A., Butenko, A., & Yurchenko, V. (2022). *Leishmania guyanensis M4147 as a new LRV1-bearing model parasite: Phosphatidate phosphatase 2-like protein controls cell cycle progression and intracellular lipid content*. *PLoS Neglected Tropical Diseases*, 16(6), e0010510. <https://doi.org/10.1371/journal.pntd.0010510>
- Zappia, L., & Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience*, 7(7). <https://doi.org/10.1093/gigascience/giy083>
- Zhang, L., & Tarleton, R. L. (1999). *Parasite Persistence Correlates with Disease Severity and Localization in Chronic Chagas' Disease*. <https://academic.oup.com/jid/article/180/2/480/881037>
- Zhang, W. W., & Matlashewski, G. (2015a). CRISPR-Cas9-mediated genome editing in *Leishmania donovani*. *MBio*, 6(4). <https://doi.org/10.1128/mBio.00861-15>
- Zhang, W. W., & Matlashewski, G. (2015b). CRISPR-Cas9-mediated genome editing in *Leishmania donovani*. *MBio*, 6(4). <https://doi.org/10.1128/mBio.00861-15>
- Zhang, Y., Morgan, M. J., Chen, K., Choksi, S., & Liu, Z.-G. (1999). *Induction of autophagy is essential for monocyte-macrophage differentiation*. <https://doi.org/10.1182/blood>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., Mcdermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H.

(2017). *ARTICLE Massively parallel digital transcriptional profiling of single cells*. <https://doi.org/10.1038/ncomms14049>

Zheng, X. M., Moncollin, V., Egly, J. M., & Chambon, P. (1987). A general transcription factor forms a stable complex with RNA polymerase B (II). *Cell*, 50(3), 361-368. [https://doi.org/10.1016/0092-8674\(87\)90490-9](https://doi.org/10.1016/0092-8674(87)90490-9)