RNA-binding proteins required for differentiation and infectivity in *Leishmania mexicana* parasites.

Ewan Raphael Swales Parry

PhD

University of York

Biology

January 2022

Abstract

Leishmania species are single celled kinetoplastid parasites and the causative agents of the neglected tropical disease, leishmaniasis. *Leishmania* parasites differentiate between non-infective and infective forms that differ greatly, both in gross morphology and in their molecular biology. In kinetoplastids, transcription of large polycistronic mRNA transcripts and a scarcity of classical RNA polymerase II promoters indicate that regulation is predominantly post-transcriptional. RNA-binding proteins (RBPs) are a major component of this post-transcriptional regulation in trypanosomatids, regulating mRNA stability as well as splicing, mRNA editing, and translational efficiency. A previous study conducted by the Walrad lab produced a highly descriptive RNA-binding proteome (RBPome) of *Leishmania mexicana* across different lifecycle stages (De Pablos *et al.*, 2019). These data were analysed further to select a range of RBP candidates for a screen of null mutants to detect loss of fitness phenotypes. Barcoded RBP knockout lines were produced in *L.mexicana* using a CRISPR/Cas9 workflow, pooled and screened to assess their function in different lifecycle stages including *in vivo* infections in mice. Of the 67 RBPs for which deletion was attempted, many could not tolerate deletion and may have essential functions that warrant further investigation. Comparative barcode sequencing (bar-seq) of 31 null mutant lines, from multiple timepoints of the bar-seq screen, revealed stage-specific phenotypes caused by RBP deletion. Whole genome sequencing and replication of individual phenotypes were used to validate the screen. Selected RBPs were fused to small epitope tags and analysed by immunofluorescence and western blot, confirming their localisation and native size. The tagged RBP candidates will be used to identify bound RNAs with pulldown experiments in future work. The bar-seq screen data will inform future studies of RBPs in *Leishmania,* in particular those with phenotypes in the human infectious stages and may be linked to differentiation, infectivity or virulence.

Table of contents

List of tables

List of figures

Acknowledgements

I would like to thank both of my supervisors, Dr. Pegine Walrad and Prof. Jeremy Mottram, for their continuous support, for many productive and engaging discussions on my experiments and for helping to navigate extension and extra funding applications during the Covid19 pandemic. I also thank my Thesis Advisory Panel (TAP) members, Prof. Deborah Smith and Dr. Michael Plevin, for their support and valuable input to the project during TAP meetings. Thank you to the University of York Department of Biology for funding this PhD and to Monica Bandeira for her support of myself and the other postgraduate students.

The work presented here would not have been possible without the support, advice and help of my colleagues in the Walrad, Mottram and Kaye labs, past and present. Thank you for all the productive discussions, answering my many questions and even lending a hand operating UV-crosslinkers and transporting industrial quantities of PBS! Special thanks to Natalia Teles, Kirstin Spence and Rachel Neish for their help in the lab and to Tiago Ferreira, Vincent Geoghegan, Nicola Baker, Juliana Brambilla Carnielli Trindade, Rebecca Burge, Nathaniel Jones, Helen Ashwin and Katrien van Bocxlaer for their advice. I want to thank Sally James and Lesley Gilbert for lending their expertise in sequencing and for their help in the rush to submit samples before lockdown. Thanks also to Katherine Newling for her advice on analysis of sequencing data and related software. I'm very grateful to James Fox, and Debbie Pears for going above and beyond to keep the labs running smoothly, allowing us to focus fully on experiments.

Outside of work, I have been lucky enough to have the support of amazing friends, especially when things have been difficult: Caroline, Bex, Anika, Rebecca, Kirstin, Elise, Natalia, Oriane, Ridda, Sergios, Ed, Rachel, Kat, Shoumit, Nidhi, Gulab, Lili, Nina, Chris, Conor, Luke, Toby, Twm, Jacob, Matt, Hannah, Aby. Thanks for being so supportive, for all the good times and for distracting me from work when I really needed a break. I truly feel lucky to have met such wonderful people during my time at York. As always, I am grateful to my family for being supportive of me no matter what I decide to do and for giving me the opportunity and freedom to pursue the things I want to achieve; freedom that I realise many others do not have. Last, but by no means least, thank you Eve for all your patience with me and for being the person I have shared these PhD experiences with most, both the good and the bad. I am so lucky to have had all of your support and kindness during these crazy times.

Declaration

I declare that I am the sole author of this thesis which is a presentation of original work except for: Figure 5.1. Whole genome sequencing of *RBP* null mutant clones. which was produced by myself and Eva Kyriacou, and the related sequence analysis and alignment described in 2.1.6 which was carried out by Katherine Newling, Eva Kyriacou and myself. The work presented in Figure 5.1 has previously been presented for an award at the University of York (Eva Kyriacou, 2021). Figure 3.1 has been published in De Pablos *et al.*, (2019). All other work presented here has not previously been presented for an award at this, or any other, University. Data conversion and production of alignment files for data in Figure 5.2 was carried out by Katherine Newling. Work directly involving mice was carried out by Rachel Neish and Jayanthi Anand. Extraction of bone marrow macrophages was carried out in collaboration with Rachel Neish. Knockout of *LmxM.30.1650* was produced by Natalia Teles. All sources are acknowledged as references (Section 8).

1 Introduction

1.1 The Leishmaniases

1.1.1 Epidemiology

The leishmaniases are a group of diseases caused by *Leishmania* parasites. The epidemiology of leishmaniasis depends on many factors including: geographical location, host genetic background, *Leishmania* species, insect vector species and available treatment resources, just to name a few. When taking into account Daily Adjusted Life Years (DALYs), the leishmaniases as a group represent a major disease burden with ~700,000 DALYs estimated globally in 2019 (Alvar *et al.*, 2012; Abbafati *et al.*, 2020). Broadly speaking, the leishmaniases can be further divided into cutaneous leishmaniasis (CL), which predominantly affects the skin, and visceral leishmaniasis (VL) which affects the internal organs (viscera). From data compiled by the World Health Organisation (WHO) on worldwide incidence, it is reported that there are between 600,000 to 1 million new cases of CL and 50,000 to 90,000 cases of VL every year (World Health Organization, 2021). Sparse and poor-quality reporting from many endemic countries means that the true incidence and impact of leishmaniasis is probably severely underestimated (Singh *et al.*, 2006, 2010; Alvar *et al.*, 2008, 2012; Mosleh *et al.*, 2008; Ready, 2010; Chowdhury *et al.*, 2014; Salam, Al-Shaqha and Azzi, 2014; Kazerooni *et al.*, 2018). Underreporting can be caused by unstable political conditions, low budgets, poor allocation of resources and harsh climates with isolated populations, among other factors. For both VL and CL, incidence has reduced as treatment programs have progressed in the last ten years (Alvar *et al.*, 2012; Burza, Croft and Boelaert, 2018; Abbafati *et al.*, 2020).

The leishmaniases predominantly affect tropical to sub-tropical regions but have presented in 98 countries across five continents with up to one billion people at risk (Kieny and Pécoul, 2020; *WHO fact sheet: Leishmaniasis*, 2021; DNDi, 2021c). Over 90% of VL cases occur in six countries: India, Bangladesh, Sudan, South Sudan, Brazil and Ethiopia but cases of CL are more evenly distributed between the Americas, the Middle East, Central Asia and the Mediterranean (Alvar *et al.*, 2012). The data on incidence that are available have been used to produce mathematical models describing and predicting the epidemiological characteristics of these diseases under a range of different conditions (Le Rutte *et al.*, 2016; Shimozako, Wu and Massad, 2017; Shiravand *et al.*, 2018). These models have been crucial for informing decisions on intervention and treatment options, and become more robust as more data becomes available. However, building models that accurately predict the epidemiology of such complex parasite-vector-host interactions is extremely challenging and requires careful consideration of many factors.

Overall, whilst *Leishmania* parasites and their sand fly vectors are widespread, geographical factors are not sufficient in explaining the distribution of cases. Globally, the incidence of severe leishmaniasis is correlated with poor living conditions (Alvar *et al.*, 2012; Ready, 2014). Some basic homes can provide breeding sites for certain sand fly species, increasing the risk of infection (Ready, 2013). Those that work for long periods outdoors where *Leishmania* and sand flies are both present are more at risk, for example, subsistence farmers in sub-Saharan Africa and those that work in forests in South America (Ready, 2013;World Health Organization, 2010). Poor sanitation and infrequent removal of waste can produce breeding sites for sand flies and so increase the risk of infection (Ranjan *et al.*, 2005). In addition, those with poor living conditions often have little to no access to healthcare making early treatment of leishmaniasis less likely and so increasing the chance of severe disease and parasite transmission. As mentioned previously, there are several distinct clinical manifestations of *Leishmania* parasite infection that have very different physical characteristics resulting in dissimilar epidemiologies. The different *Leishmania* species, diseases, geographical locations and vectors have been reviewed in detail (Sharma and Singh, 2008; World Health Organization, 2010; Marcili *et al.*, 2014) and are discussed in 1.2.2. In some cases, as with *L. infantum* caused visceral leishmaniasis, prevalence decreases rapidly as living conditions improve but in others the situation is more complex (Ready, 2014).

In addition to the many human factors involved, those determined by the vectors themselves play a huge part in the epidemiology of leishmaniasis. The vectors of *Leishmania* parasites are sand flies of the subfamily *Phlebotominae*, predominantly from the genera *Phlebotomus* (in the Old World) and *Lutzomyia* (in the New World) (Killick-Kendrick, 1990; Sharma and Singh, 2008; Dostálová and Volf, 2012). Mathematical modelling of factors such as vector habitat, feeding habits and breeding locations can be used to understand what drives patterns in disease incidence. Most of these sand fly vectors exhibit marked local, seasonal, periodicity adding to the complexity of modelling *Leishmania* transmission and epidemiology (Ready, 2013). Non-human zoonotic reservoirs also play their part in the persistence of leishmaniasis with many sand flies exhibiting promiscuous biting habits; for

example the persistent infection of dogs with *L. major* in the Mediterranean or with *L. infantum* in Brazil (Ready, 2010; Shimozako, Wu and Massad, 2017). A range of other mammals including rodents across the Middle East and wild canids, or even sloths (Folivora) and anteaters (Vermilingua) in South America, can act as important reservoir hosts. The habits and the interaction of these wild hosts with humans are another set of factors that have to be considered when attempting to combat endemic leishmaniasis.

Overall, internationally co-ordinated efforts are improving the production and sharing of information on the incidence of leishmaniasis globally. As more data becomes available to those modelling the epidemiology of the leishmaniases, the picture we have of the disease becomes clearer and our ability to predict the impact of different control and treatment strategy improves. Information on the epidemiology of these diseases is by no means uniform, and although lacking in some areas, in others it has improved to the point where eradication of leishmaniasis is now a tangible goal.

1.1.2 Clinical symptoms

Symptoms presented by those infected with *Leishmania* parasites can be grouped into three major categories: cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL) and visceral leishmaniasis (VL). All of these basic groups of symptoms can present in a number of varied forms depending on the *Leishmania* species as well as the genetic background and immune system of the infected host. A fourth major group that is often overlooked includes those that are seropositive but asymptomatic. Historically, Old World cutaneous leishmaniasis symptoms have been documented since ancient times, with the first detailed reports from Avicenna in the 10th century who described cutaneous lesions, in what is now Afghanistan, as the 'Balkh sore'(Cox, 2002). Visceral leishmaniasis, also referred to as Kala azar was first identified in Indian patients in the 19th century, presenting with a fever that failed to reduce after antimalarial treatment (Cox, 2002). The history of leishmaniasis symptoms appearing in the New World has been documented at least as far back as the 16th century with similar diseases caused by New World specific parasites. The notable exception to this is New World VL caused by *L. infantum*, which was almost certainly brought over from Europe and Asia in multiple introduction events (Kuhls *et al.*, 2011).

Of the four categories mentioned, VL is considered to be the most serious as it is lethal if left untreated and affects young children particularly badly. Visceral leishmaniasis is caused by *L. donovani* in Africa and Asia and by *L. infantum* in the Mediterranean and South America. As the name suggests, the deadly symptoms of VL are caused by the parasite invading the viscera. Most commonly this includes proliferation of parasites in macrophages in the liver, spleen and bone marrow which leads to severe hepatosplenomegaly and immunosuppression (McGwire and Satoskar, 2014). Patients often present with a fluctuating fever as well as anaemia and leukopenia (sometimes classed more broadly as pancytopenia)(Rodrigues V., Da Silva and Campos-Neto, 1998; Varma and Naseem, 2010; Shah *et al.*, 2012; Flora *et al.*, 2014; McGwire and Satoskar, 2014; Ready, 2014).

VL is a major problem in children, possibly due to the incomplete development of the immune system. As with adults, the disease is fatal if left untreated but even after a young patient is admitted to hospital, treatment can be difficult with as much as 20% of children in hospital suffering from pulmonary infection in some areas (Rocha *et al.*, 2011). Weight loss or anorexia, abdominal distortion, nausea and jaundice are also common symptoms presented in children with VL, meaning many patients are very weak by the time they are admitted for treatment (Shah *et al.*, 2012).

An additional problem with VL is that even after successful treatment and recovery, a secondary presentation of symptoms known as post Kala-azar dermal leishmaniasis (PKDL) can manifest itself, causing further disease burden and potentially, further transmission (Ganguly *et al.*, 2015). PKDL patients typically present with symptoms around six months after treatment with the main symptoms being macular, maculo-papular or maculo-papular-nodular lesions covering a large area of the skin (Ismail *et al.*, 1999; Desjeux *et al.*, 2013). The severity and diffuse nature of these lesions makes them extremely difficult to treat. These lesions can last as long as 96 months and even after successful treatment to remove the parasites, healing of such widespread wounds can take a long time.

In contrast to VL, the symptoms of CL are seldom life threatening. Because of this, the effects of CL are often downplayed when in reality they contribute significantly to the very high DALY score that is attributed to the leishmaniases as a whole. CL is caused by a wide range of *Leishmania* species with some species dependent variation in symptoms (Handler *et al.*, 2015). Commonly, CL in the Old World is caused by *L. major* and *L. tropica* and by *L. mexicana, L. braziliensis, L. amazonensis, L. guyanensis and L. panamensis* in South and Central America. In general, *L. major and L.tropica* cause milder symptoms than the New World CL-causing *Leishmania* species (David and Craft, 2009). Typically, in an immunocompetent patient, these species cause a small cutaneous papule at the bitten site that progresses into a nodule. The nodule can then ulcerate, eventually healing with some permanent scarring (David and Craft, 2009). These symptoms affect the DALY score heavily because of the cost of treatment, social stigma and psychological harm that they cause. In several countries, women with CL or CL-caused scarring are heavily stigmatised, unable to marry and can lose access to their children (Velez *et al.*, 2001; Bennis *et al.*, 2017). Children can be severely affected by this social stigma as it is often assumed their lesions are contagious resulting in them being removed from education and restricted socially (Bennis *et al.*, 2017; Garapati *et al.*, 2018). Overall, cost of treatment and stigmatization can make this a debilitating disease with a lifelong negative effect on quality of life.

Some of the New World *Leishmania* species cause similar localised cutaneous lesions. For example *L. mexicana* infection often presents as a single localised, rounded ulcer on the ear, commonly referred to in South America as ‘chiclero´s ulcer’ (Low, 1919; Andrade-Narvaez *et al.*, 2005). It has been proposed that the abundance of cases with a single lesion on the ear is due to a combination of the exposure of that area and the biting habits of the vector (Eldin *et al.*, 2021). Lesions of this type very seldom progress into a more serious form of leishmaniasis.

Other New World *Leishmania* species in the *Viannia* subgenus, primarily *L. braziliensis* and *L. amazonensis*, can cause much more disfiguring mucocutaneous leishmaniasis (MCL). This form of the disease involves invasion of the mucous membranes and can be fatal if left untreated (David and Craft, 2009). Although most cases of leishmaniasis caused by the *Viannia* subgenus will present as a lesion that eventually heals, some eventually return with expansion into surrounding mucosal tissue directly or via the lymph system or bloodstream (Marra, Chiappetta and Vincenti, 2014). Infections of the nasal mucosa are the most common with many patients describing nasal congestion as an early symptom which progresses into the formation of nodules which later ulcerate (Daneshbod *et al.*, 2011). In later stages of the disease the disfiguration caused can be extremely severe and some patients develop similar symptoms to VL such as hepatosplenomegaly (Daneshbod *et al.*, 2011; Marra, Chiappetta and Vincenti, 2014). Whilst this presentation of leishmaniasis is seen outside of south America it is far rarer in the Old World (Shehzad and Abbas, 2010; Bari *et al.*, 2012; Madeddu *et al.*, 2014).

Although classically the symptoms of leishmaniasis have been divided into these three main forms, the full spectrum of the disease is much more broad. There are many more uncommon presentations of leishmaniasis, combinations of symptoms and co-infections that require investigation. Many of the differences in symptoms between patients are likely to be a result of the unique interplay between our immune system and the parasite. For example, a less common disseminated cutaneous leishmaniasis (DCL) can occur after infection with species that usually cause localised CL or MCL (*L. major, L. tropica, L aethiopica, L. braziliensis, L. panamensis, L. guyanensis and L. amazonensis*)(World Health Organization, 2010; Hashiguchi *et al.*, 2016; Dassoni, 2017; Membrive *et al.*, 2017). Although not as common as other presentations, DCL is severely debilitating as dense patches of nodules form across multiple areas of the body making treatment very difficult. The fact that this form is rare and not confined by cause to one *Leishmania* species group suggest that this form of disease may be linked to underlying immunological factors (Rosa and MacHado, 2011). DCL has been linked to patients that are immunodeficient, especially those that have co-infections like HIV or even other forms of leishmaniasis (Alborzi *et al.*, 2008; Rosa and MacHado, 2011). However, cases have been reported where no co-infections or underlying conditions have been detected (Hajjaran *et al.*, 2013). Another similar form, often termed ‘diffuse’ rather than disseminated, appears to be separable on the basis of mucosal involvement,which is never observed in the case of ‘diffuse’(Hashiguchi *et al.*, 2016). There are many more unusual symptoms of the leishmaniases that have yet to be examined or well documented.

One common feature of unusual, often unusually severe, leishmaniasis symptoms is the presence of underlying health problems or co-infections. Amongst the most concerning examples are those of co-infection with malaria or with HIV. Immunodeficiency appears to greatly increase the risk of serious disease after *Leishmania* infection making HIV co-infection particularly problematic (Alvar *et al.*, 2008; Burza *et al.*, 2014; Madeddu *et al.*, 2014; Marra, Chiappetta and Vincenti, 2014; Diro *et al.*, 2015; Lindoso *et al.*, 2016; Henn *et al.*, 2018). Similarly, the burden placed on the body from

Malaria co-infection impacts how well patients respond to leishmaniasis treatments and can result in severe clinical symptoms that are sometimes hard to correctly identify (Pinna *et al.*, 2016). *Leishmania-Plasmodium* co-infection is especially prevalent in sub-Saharan Africa (van den Bogaart *et al.*, 2012; Aschale *et al.*, 2019).

1.1.3 Leishmaniasis treatments

The leishmaniases, with many causative species and widely varying clinical symptoms, are challenging diseases to treat. To date, no singular approach has been found effective against all forms of the disease or in all endemic regions. In addition, most of the drugs available are toxic and do not fully eliminate the parasite. As yet no effective vaccine for leishmaniasis has been produced.

The first significant treatments for leishmaniasis were based on antimony as a cure for cutaneous lesions (Vianna, 1912). Soon after, intravenous injections of 'Tartar emetic' (antimony potassium tartrate) were administered to VL patients as they had been for several other parasitic diseases including human African trypanosomiasis (Di Cristina and Caronia, 1915 in Haldar, Sen and Roy, 2011) (Muir, 1915 and Rogers, 1915 in Rogers, 1939). 'Tartar emetic' was also adopted as a treatment for CL in the Americas (Low, 1919). Although this somewhat crude treatment was highly toxic (Severe respiratory distress and chest pain) and had many critics, it had a significant effect on the reduction of the mortality rate in VL from a 90% mortality rate to an 80% recovery rate (Gray and Trevan, 1931). The discovery of less toxic pentavalent antimonials (Sb(V)) for leishmaniasis treatment in 1945 was a major step forwards and has been the primary treatment ever since (Goodwin, 1945). Sb(V) is thought to enter the parasites via a phosphate transporter (Berman, Gallalee and Hansen, 1987; Rosen, 2002), possibly recognised by a *Leishmania* protein due to its structural similarity to a sugar moiety (Brochu *et al.*, 2003) but our understanding of this process is still lacking. The toxicity of Sb(V) to *Leishmania* is also poorly understood but it has been suggested that it acts as a prodrug with reduction in the parasite or macrophage to the more active Sb(III) (Wyllie, Cunningham and Fairlamb, 2004; Wyllie, Vickers and Fairlamb, 2008). It has also been suggested that SbV directly affects amastigote viability (Berman, Waddel and Hanson, 1985) and that it stimulates the host immune system to clear parasites by increasing levels of reactive oxygen species and nitric oxide (Basu *et al.*, 2006). None of these theories are necessarily mutually exclusive.

Since the discovery of Sb(V), other drugs with antileishmanial activity have been heavily sought after. Because Sb(V) treatment fails to completely remove *Leishmania* parasites, development of resistance has been an extensive problem, especially in India where Sb(V) is now far less effective than it once was (World Health Organization, 2010). Some alternative drugs have become available. Amphotericin B and Pentamidine, a polyene antifungal and an aromatic diamidine respectively, have both been tested with positive results for leishmaniasis treatment. Overall, despite displaying toxicity

and severe side effects at high doses, low doses of Amphotericin gave better results than pentamidine for treating VL (Mishra *et al.*, 1992). Several lipid formulations of Amphotericin were subsequently produced, the most widely used being the liposome packaged formulation, AmBisome. AmBiosome treatment of VL has been largely successful and is used both alone and in combination with other drugs (Smith *et al.*, 1995; Berman *et al.*, 1998; Balasegaram *et al.*, 2012). However, in patients with HIV or other immunosuppressed patients, AmBisome treatment has shown poor results, indicating a functioning immune system plays a role in the efficacy of this drug (Ritmeijer *et al.*, 2011). The alkylphospholipid drug, Miltefosine has also been successfully used to treat leishmaniasis, and is the only orally delivered drug that has seen widespread use. In general, Miltefosine has been more successful in South America than in the Middle East, Africa or Asia (Sundar *et al.*, 1998; Jha *et al.*, 1999) where resistance has proved problematic (Rijal *et al.*, 2013; Monge-Maillo, López-Vélez and Saravolatz, 2015). Like other systemic treatments Miltefosine is better suited to treating VL than CL, where local treatments (topical antimonials, cryotherapy and thermotherapy) are often used preferentially due to lower systemic toxicity (Monge-Maillo and López-Vélez, 2013a, 2013b). It is hard to determine how much Miltefosine efficacy depends upon *Leishmania* species due to geographical variation and a lack of standardisation in reporting methods (Ware *et al.*, 2021). Despite this, it has shown promise in many areas (e.g. Espada *et al.*, 2021) and if resistance is managed effectively, it may still see more widespread use (Ponte-Sucre *et al.*, 2017). As with other antileishmanial drugs, it is not fully understood how Miltefosine affects the parasite but it has been suggested that it interferes with *Leishmania* cytochrome oxidase C and induces apoptosis (which is a much debated topic in *Leishmania*)(Paris *et al.*, 2004; Luque-Ortega and Rivas, 2007; Verma, Singh and Dey, 2007). Resistance to Miltefosine, appears, at least in part, to be due to reduced uptake or increased efflux of the drug, possibly linked to mutations in the necessary *L. donovani* Miltefosine transporter (LdMT) or its associated factor LdRos3 (Pérez-Victoria, Castanys and Gamarro, 2003; Pérez-Victoria *et al.*, 2006; Seifert *et al.*, 2007). Mutations with links to Miltefosine resistance have since been discovered in other regions of the genome including the Miltefosine sensitivity locus (MLS) providing new understanding of the multifactorial nature of the problem (Verma, Singh and Dey, 2007; Coelho *et al.*, 2012; Kulshrestha *et al.*, 2014; Mondelaers *et al.*, 2016; Vacchina *et al.*, 2016; Carnielli *et al.*, 2018).

Lastly Paromomycin sulfate has been the only aminoglycoside drug with significant antileishmanial activity. Although it has been found to be somewhat ineffective in randomised trials (Salah *et al.*, 1995), further development of topical formulations has produced better results for curing CL (Armijos *et al.*, 2004) leading to its use as a combination therapy. Paromomycin is now being used in combination with AmBisome and some promising new oral combination formulations have passed animal trials for the treatment of VL (Parvez *et al.*, 2020)

Overall, today there are still very few effective treatments for leishmaniasis, and none of the available treatments are effective in all endemic countries. Worse still, the parasites are showing signs of resistance to many of these treatments; resistance that tends to increase the more a treatment is used.

This is often due to incomplete removal of parasites during the period of drug administration (Ponte-Sucre *et al.*, 2017). Whereas previous treatments for leishmaniasis have been largely decided at the discretion of localised medical professionals, long overdue clinical guidelines have been produced to standardise practices (Copeland and Aronson, 2015). As before, Sb(V) (usually Sodium stibogluconate) is the first line drug in most cases. Pentamidine is not widely used since it can cause insulin-dependent diabetes mellitus as a side effect and its efficacy has declined with increasing levels of parasite resistance (Mishra *et al.*, 1992)(Mishra *et al.*, 1992; Jha, 2006). Some success has been reported with Miltefosine, especially in South America, and AmBisome combination therapies more widely. The Drugs for Neglected Diseases initiative (DNDi) has been running many clinical trials involving combination therapies, such as the combination of thermotherapy with a short course of miltefosine, which is reported to be significantly better than thermotherapy alone (DNDi, 2021d). Another recent success with combination therapy is the use of sodium stibogluconate and Paromomycin to treat visceral leishmaniasis in east Africa which has reduced side effects and nearly halved the treatment time (Kimutai *et al.*, 2017). A promising combination therapy including CpG-D35 oligonucleotides as immunomodulators alongside the more traditional drugs is currently starting first-in-human clinical trials (DNDi, 2021a). Despite these and other recent successes (Chakravarty and Sundar, 2019; Roatt *et al.*, 2020), the rapidly developing resistance and relative cost (compared with Sb(V)) makes it clearer than ever that novel drugs are needed to combat the disease effectively. In particular, drugs with known mechanisms of action against well studied targets would be an ideal tool when trying to understand and circumvent the parasite's mechanisms of resistance.

1.1.4 Novel molecular targets and drugs

Part of the reason for the lack of novel drugs targeting *Leishmania* species is our lack of understanding of the genetics and molecular biology of these parasites. Being highly divergent from other eukaryotes means there is potential for finding divergent protein targets and biological pathways to disrupt, albeit less so than when targeting a virus, bacterium or fungus. However, their cellular complexity, extreme genomic plasticity and adaptability (as with *Trypanosoma* and *Plasmodium*) make them challenging targets to design drugs against. Recently, a rapidly expanding toolkit of *Leishmania* genome editing methods (section 1.2.5) has helped to greatly enhance our understanding of these parasites on a molecular level. The availability of CRISPR/Cas9 editing in several *Leishmania* species in particular has meant that drug target validation has become easier. Through target-based drug discovery, often starting with bioinformatic screens, some potential drug targets have recently been identified and studied (Opperdoes and Szikora, 2006; Walker *et al.*, 2012; Vakili *et al.*, 2018; Chávez-Fumagalli *et al.*, 2019). These targets include enzymes from the glycolysis and gluconeogenesis pathways (Opperdoes and Szikora, 2006) as well as several enzymes involved in the synthesis of precursors to trypanothione, a part of the oxidative stress pathway implicated in drug resistance (Walker *et al.*, 2012) and both topoisomerase II and dihydrofolate reductase (DHFR) (Gilbert, 2002; Chen *et al.*, 2008; Datta, Datta and Sen, 2008; Sharma *et al.*, 2012; Balaña-Fouce *et al.*, 2014; Harris, Mitchell and Morris, 2014). *Leishmania* kinases including Cdc2-

related kinase 3 (CRK3:CYC6), Casein Kinase 1 (CK1) and Mitogen-activated protein kinase 4 (MPK4), have also been investigated with varying levels of success and ongoing studies show promise in this area (Grant *et al.*, 2004; Saravanan *et al.*, 2010; Walker *et al.*, 2011; Rachidi *et al.*, 2014; Sanderson, Yardley and Croft, 2014). Although this has produced some valid avenues for drug design (even a few drugs that successfully inhibit their target), no clinically available anti-leishmanial treatments have resulted from this work to date. When looking for novel drug targets the primary concern is that the target protein is different enough from the closest structural homolog in the host as this difference directly impacts drug efficacy. As with some of the proteins mentioned earlier, this means targets belonging to pathways and mechanisms that are unique or highly divergent from those in the host cells are favourable. Further understanding of the molecular adaptations of *Leishmania* to their unique parasitic niche will help to uncover new avenues of drug discovery in this way. Bioinformatics can also be extremely useful for ranking large numbers of proteins by their relative conservation in *Leishmania* compared to mammalian hosts. To date, no genome-wide bioinformatic study of this kind has been attempted in *Leishmania* although genomic comparisons between kinetoplastids and other organisms have been made for some groups of proteins such as CCCH zinc-finger proteins, protein kinases and de-ubiquitinases (DUBs) (Kramer, Kimblin and Carrington, 2010; Damianou *et al.*, 2020; Baker *et al.*, 2021).

Alternatively, phenotypic drug screening, utilising high throughput screening of libraries of existing compounds against *Leishmania*, has become vastly more efficient as technologies have advanced and large pharmaceutical companies have been involved in collaboration. Traditionally, screening whole libraries of drugs against cultured *Leishmania* has been the most popular method for discovering drugs and the majority of new anti-leishmanials have been discovered in this manner. When these large libraries are composed of approved medical compounds, chances of arriving at effective drug treatments are greatly increased, but this often requires the involvement of an industrial partner. However, the subsequent deconvolution process for elucidating the target of any given compound can often be time consuming and fail to produce a result at all. This makes modification and iterative improvement of the drug difficult. Despite these drawbacks the ability to screen at high throughput currently makes this method preferable and it is largely studies of this kind that are working in coordination with the drugs for neglected diseases initiative (DND*i*) to expand the list of novel drugs and drug targets in *Leishmania* (Rajasekaran and Chen, 2015; Zulfiqar, Shelper and Avery, 2017; Brindha, Balamurali and Chanda, 2021). A list of current targets under investigation as well as a list of anti-leishmanial drugs that have already been approved is presented in detail in Rajasekaran and Chen (2015) and more recently discussed in Jones *et al.* (2018) and Brindha, Balamurali and Chanda (2021). Proteasome inhibitors specific enough to avoid inhibition of the mammalian proteasome, but broad enough to target several trypanosomatids, show particular promise (Khare *et al.*, 2016; Zhang and Lin, 2021). Cyclin-dependent kinases (CDKs) such as CRK12, CRK6 and CRK3 have also been suggested as drug targets in *Leishmania* (Walker *et al.*, 2011; Wyllie *et al.*, 2018). Another kinase that has been identified as a potential target is Casein Kinase 1 (CK1), which is susceptible to CK1-specific inhibitors which induce reduced viability of

axenic amastigotes and reduced numbers of intracellular amastigotes in infected macrophages (Allocco *et al.*, 2006; Rachidi *et al.*, 2014). Other drugs screened against amastigotes have recently been listed in (Dias-Lopes *et al.*, 2021). Several programs initiated by the DNDi (Leishmaniasis hit-to-lead, Screening leishmaniasis and crossover with the NTD Drug Discovery Booster) cover novel drug targets that are currently under pre-clinical investigation or are in clinical trials. Examples include the nitroimidazole derived DNDI-0690 and oxaborole class DNDI-6148 which are both showing promise in human clinical trials (DNDi, 2021b).

A major concern for *Leishmania* drug targets is that the target protein is both accessible and expressed in the amastigote stage of the parasites, as this is the stage that is predominantly treated. Currently there are many studies describing the expression of *Leishmania* proteins in promastigotes (Góngora *et al.*, 2003; Cuervo *et al.*, 2007; Alcolea, Alonso and Larraga, 2011; Walker *et al.*, 2012; Hajjaran *et al.*, 2018; Alcolea *et al.*, 2019; Tasbihi *et al.*, 2019; Sanchiz *et al.*, 2020). Other comparisons have been made between promastigotes and amastigotes (Thiel and Bruchhaus, 2001; El Fakhry, Ouellette and Papadopoulou, 2002; Bentel *et al.*, 2003; Foucher, Papadopoulou and Ouellette, 2006; Paape *et al.*, 2010; Pescher *et al.*, 2011; Carnielli *et al.*, 2014; Nirujogi *et al.*, 2014; de Rezende *et al.*, 2017; Nandan *et al.*, 2017). Lifecycle-stage-specific proteomes were presented for procyclic promastigotes, metacyclic promastigotes, axenic amastigotes and lesion-derived amastigotes in De Pablos *et al.* (2019) .Whilst promastigote proteomes are valuable for further studies on differentiation and transmission, in terms of drug discovery, proteomic and RNA sequencing data from the human infectious stages are the most useful. In addition to the amastigotes, metacyclic promastigotes are also infectious, but their role in disease has been poorly studied and, due to the rapid process of amastigogenesis (Courret *et al.*, 2002), they are often overlooked in studies of the interactions between the host immune system and the parasite.

Vaccines are also currently being investigated as a means of combating leishmaniasis. As with drug design, knowledge of *Leishmania* specific pathways and protein targets can aid vaccine design, although the focus for vaccines tends to be on parasite surface proteins due to their availability as antibody ligands. The earlier attempts at production of a *Leishmania* vaccine included use of whole killed parasites, fractionated *Leishmania* antigen or live attenuated parasites (Moafi *et al.*, 2019). While some have passed clinical trials and are used in cases of cutaneous leishmaniasis that are otherwise unresponsive to treatment, none of these have been used in a widespread way to grant protective immunity in humans. One vaccine based on three recombinantly produced *Leishmania* proteins, LEISH-F1 shows considerable promise as it can treat CL patients as well as induce protective immunity in healthy volunteers (Llanos-Cuentas *et al.*, 2010; Chakravarty *et al.*, 2011). Similar promising results were described following clinical trials of LEISH-F3, a vaccine based on the recombinant proteins nucleoside hydrolase (NH) from *L. infantum* and sterol 24-c-methyltransferase (SMT) from *L. donovani* (Coler *et al.*, 2015). Recently, a vaccine (ChAd63) based on the simian adenovirus delivery of KMP-11 and HASPB antigens from *L. donovani*, has entered clinical trials and performed well (Osman *et al.*, 2017). It is currently being evaluated in

phase II trials. General reviews of progress towards a preventative *Leishmania* vaccine and on methodology can be found in Ghorbani and Farhoudi, 2018; Moafi *et al.*, 2019 and Cecílio *et al.*, 2020.

1.2 *Leishmania* parasites

1.2.1 Origins and evolution

Leishmania parasites belong to the excavates, a supergroup of flagellated unicellular eukaryotes that were previously classified as belonging to the kingdom Protista. Modern molecular evidence has shown that the divergence within protists does not warrant their grouping as a eukaryotic kingdom. Within Excavata, the phylum Euglenozoa contains the class Kinetoplastea, organisms including *Leishmania* that contain the kinetoplast which carries mitochondrial DNA (Cavalier-Smith, 2016). Within the Kinetoplastea, the order Trypanosomatida contains over forty species of *Leishmania* and several other important human pathogens such as *Trypanosoma brucei* and *Trypanosoma cruzi* (Kaufer *et al.*, 2017). While many species of trypanosomatids are monoxenous and non-pathogenic, the parasitic *Trypanosoma* and *Leishmania* have a complex dixenous lifecycle that requires both a mammalian host and an insect vector for survival (discussed in detail in 1.3.2). The Kinetoplastids in general are highly divergent from the well-known eukaryotic model organisms with many major differences in their basic molecular biology. The evolution of Kinetoplastids is often hard to determine, but it is thought that parasitic agents like *Leishmania* evolved from monoxenous parasites or symbionts of insects, which in turn, shared a common ancestor with free-living flagellates (Cavalier-Smith, 2016; Kaufer *et al.*, 2017).

Given their early evolutionary divergence, it is not surprising that kinetoplastids have evolved to fit many different specific niches, from free living water dwelling flagellates to highly dependent monoxenous plant and insect pathogens or, in the case of *Leishmania,* obligate dixenous parasites that cause a major human disease (Simpson, Stevens and Lukeš, 2006; Stevens, 2008; Lukeš *et al.*, 2014). It is thought that parasitic Trypanosomatids like *Leishmania* diverged from other, free-living, relatives several hundred million years ago (Cavalier-Smith, 2016; Harmer *et al.*, 2018). The relationship between *Leishmania* species and their hosts is a complex one that has likely developed over the course of millions of years, as free-living parasites have adapted to a symbiotic or parasitic lifecycle in an insect and then in turn to a higher order animal. Not all parasitic *Leishmania* species with a dixenous lifecycle infect mammals as a secondary host. For example, *Leishmania tarentolae*, a useful strain for lab based studies, is not pathogenic to humans but is transmitted by the sand fly vector to lizards (Lainson and Shaw, 1987; Taylor *et al.*, 2010). Evolutionary studies have shown that the saurian and mammalian *Leishmania* are monophyletic, both deriving from the same common ancestor (Croan, Morrison and Ellis, 1997)

1.2.2 *Leishmania* lifecycle

As stated previously, the *Leishmania* species of importance to human disease are obligate parasites dependent on replication in both a mammalian and a sand fly host. The complex lifecycle of *Leishmania* is presented graphically in Figure 1.1. *Leishmania* amastigotes are first ingested by female sand flies of the genera *Phlebotomus* and *Lutzomyia* when they take a blood meal from an infected mammal (Killick-Kendrick, 1999; Ready, 2013). In the midgut of the sand fly, the change in environmental conditions such as temperature (decreasing) and pH (increasing) cause amastigotes to differentiate into procyclic promastigotes (Courret *et al.*, 2002; Barak *et al.*, 2005; Alcolea *et al.*, 2010). These flagellated, motile, and highly metabolically active cells are the first stage that are capable of replication in the sand fly. The procyclic promastigotes are initially confined to the bloodmeal, surrounded by peritrophic matrix but soon differentiate further into long nectomonads, which are able to escape the matrix as it is broken down by host enzymes (Dostálová and Volf, 2012). The long nectomonads differentiate into short nectomonads (also called leptomonds) which attach to the midgut microvilli to avoid being excreted with what remains of the bloodmeal (Pimenta *et al.*, 1994). Stage specific modifications to the terminal saccharides of the surface molecule lipophosphoglycan (LPG) were shown to be sufficient for allowing parasite detachment from the midgut epithelium (Pimenta *et al.*, 1992). Free swimming leptomonads then develop into the infective metacyclic promastigotes through the process of metacyclogenesis (da Silva and Sacks, 1987; Gossage, Rogers and Bates, 2003). Metacyclogenesis is triggered by a reduction in pH and low nutrient availability and remodels the parasites, producing a longer flagellum in comparison to the body length, increased motility as well as fundamental changes that determine infectivity (Sacks and Perkins, 1984; da Silva and Sacks, 1987; Bates, 2008). Although generally considered unidirectional, there is some evidence that metacyclogenesis may be reversible (at least for some individuals in the population) with the addition of a fresh blood meals inducing differentiation back into retro-leptomonad precursors (Serafim *et al.*, 2018). Metacyclic promastigotes, gradient purified based on their morphology, have been characterised by specific expression of the endoplasmic reticulum (ER) membrane associated marker SHERP (small hydrophilic ER-associated protein) (Späth and Beverley, 2001; Depledge *et al.*, 2010; Sádlová *et al.*, 2010). Another membrane associated protein, HASPB (hydrophilic acylated surface protein B), differentiates metacyclic promastigotes from other vector stages but is also expressed by amastigotes (Depledge *et al.*, 2010; Maclean *et al.*, 2012). While it is known that SHERP and HASPB are required for differentiation, many of the molecular mechanisms driving infectivity are yet to be investigated (Sádlová *et al.*, 2010; Doehl *et al.*, 2017). Metacyclic promastigote motility also appears to differ from the 'corkscrew' motion seen in procyclic promastigotes, instead employing a 'run and tumble' motion that is faster and holds a straight line for longer periods of time (Findlay *et al.*, 2021). In addition, metacyclic promastigotes were found to swim towards macrophage-derived stimuli, consistent with their role in establishing infection in host macrophages. Metacyclic promastigotes, along with their leptomonad precursors, aggregate in the anterior midgut at high density and are surrounded by a filamentous

proteophosphoglycan containing gel (promastigote secretory gel or PSG) that leads to a blockage of the sand fly gut, larynx and stomodeal valve (Rogers *et al.*, 2004; Hall *et al.*, 2021). The blockage leads to regurgitation of the promastigote parasites, mostly metacyclic promastigotes, during feeding which increases the number of parasites transferred to the mammalian host that are available to establish an infection (Schlein, Jacobson and Messer, 1992; Volf *et al.*, 2004). A poorly studied alternative lifecycle stage, thought to be derived from leptomonads, named haptomonads, attach to the epithelial wall of the larynx and the stomodeal valve and accumulate, potentially contributing to the regurgitation mechanism of transmission (Killick Kendrick, Molyneux and Ashford, 1974; Molyneux, Killick-Kendrick and Ashford, 1975; Killick-Kendrick, 1990; Schlein, Jacobson and Messer, 1992; Volf *et al.*, 2004; Serafim *et al.*, 2018; Hall *et al.*, 2021).

Once transmitted to the mammalian host, metacyclic promastigotes seek macrophages which also migrate to the site due to the inflammation. It seems that *Leishmania* can infect a range of cell types but mononuclear phagocytes are certainly a primary target of infection (Lai *et al.*, 2008). A small number of promastigotes can be found within neutrophils immediately after infection but are soon taken up by phagocytic macrophages where differentiation occurs (Bates *et al.*, 1992; Peters and Sacks, 2009). Parasite uptake is mediated by Fc and complement receptor binding followed by phagocytosis, although alternative mechanisms must exist regarding non-phagocytic cells (Guy and Belosevic, 1993; Peters *et al.*, 1995; Love, Kane and Mosser, 1998; Kima *et al.*, 2000; Kima, 2007). One specific pathway that has been described is caveolin-mediated endocytosis of *Leishmania donovani* into host macrophages (Kumar *et al.*, 2019). Following uptake, the intracellular parasite is surrounded by a membrane bound compartment in as little as 30 minutes (Bates *et al.*, 1992). The fusing of the promastigote containing compartment and the host cell lysosome forms a parasitophorous vacuole where the intracellular amastigotes can survive and replicate long term (Alexander and Vickerman, 1975; Chang and Dwyer, 1978). Human polymorphonuclear neutrophils have a significant role in the uptake of *L. donovani* as they are recruited in large numbers to the infection site and readily internalise the parasites (Pitale *et al.*, 2019). The same authors present evidence for *Leishmania* inducing neutrophil autophagy, leading to engulfment of the parasite-infected neutrophils by macrophages. There is evidence that *Leishmania* actively interfere with the conditions in the monocyte-derived phagosome to optimise it for their survival, for example reducing the amount of reactive oxygen species (ROS) present and increasing the pH (Matte, Arango Duque and Descoteaux, 2021). It appears that the ability to control phagosome conditions differs depending on the strain of *Leishmania* (da Silva Vieira *et al.*, 2019). When inside a parasitophorous vacuole, the harsh conditions promote amastigogenesis where differentiation to amastigotes occurs over the course of 24-72h (Kima, 2007; Kaye and Scott, 2011). Characteristic loss of the flagellum occurs in as little as five hours for some species (Bates *et al.*, 1992). Another characteristic feature of amastigotes is the surface protein δ-amastin, first identified in *T. cruzi*, which is essential for viability in *Leishmania* in host macrophages (Teixeira *et al.*, 1994; Jackson, 2010; de Paiva *et al.*, 2015). Replication of amastigotes whilst hijacking the host immune system underpins most of the disease symptoms caused by *Leishmania* parasites, although the many different presentations are far from

being well understood in the context of *Leishmania* molecular biology and the host immune system (Kima, 2007; Kaye and Scott, 2011). Following amastigote replication, the parasitophorus vacuole can swell to overwhelm the cell and eventually rupture releasing the parasites (Ridley, 1980). Released amastigotes can go on to infect further macrophages and recently it has been shown, at least for *L. amazonensis,* that amastigotes can also infect other host cells after being extruded from cells via blebs in the membrane that bud off and are phagocytosed again (Bates *et al.*, 1992; Love, Kane and Mosser, 1998; Rittig *et al.*, 1998; Lai *et al.*, 2008; Real *et al.*, 2014). Amastigotes are taken up by sand flies when they take a bloodmeal from an infected mammalian host, continuing the lifecycle.

1.2.3 *Leishmania* genetics

The genetics of *Leishmania* parasites, as with many of the fundamental aspects of their biology, is currently poorly understood. Unlike some other areas, comparison of *Leishmania* genetic mechanisms to those of other kinetoplastids does not often yield useful insight as they are highly divergent. Most *Leishmania* parasites have 36 chromosomes but *Leishmania mexicana* has 34 including fusion events between chromosomes 8 and 29 and between 20 and 36 (Britto *et al.*, 1998; Ivens *et al.*, 2005; Rogers *et al.*, 2011). Overall, *Leishmania* parasites have extremely plastic genomes compared to higher eukaryotes that possibly help them to adapt to a wide range of challenges and environments (Reis-Cunha, Valdivia and Bartholomeu, 2017). One aspect of this is the large degree of copy number variation observed for *Leishmania* chromosomes (Rogers *et al.*, 2011; Bussotti *et al.*, 2018). A consistent feature in *Leishmania mexicana* is the tetraploid nature of chromosome 30, but many other chromosomes become aneuploid with no obvious link to a temporal or environmental stimulus to date (Rogers *et al.*, 2011; Sterkers *et al.*, 2011). As well as copy number variation across whole chromosomes, mosaic aneuploidy has been described where cells within a clonal population contain chromosomes with differing ploidy caused by uneven pairing during mitosis (Sterkers *et al.*, 2011; Bussotti *et al.*, 2018; Damasceno *et al.*, 2020). The recent efforts of Negreira *et al.* (2021) presented the most complete picture of this phenomenon yet using single cell genome sequencing techniques to track the changing karyotypes of parasites from initially euploid, clonal lines. They found that some chromosomes are much more likely than others to achieve high copy numbers and that there is a general expansion in chromosome copy number during the promastigotes compared to amastigotes which are more often stably diploid. There are significant negative selection pressures to having high genomic plasticity in multicellular organisms, as drastic changes in the genome will inevitably lead to instability in some cells which may then impact the whole organism. In the case of single-celled *Leishmania,* while some individuals in the population may be negatively affected, faced with extreme selection pressures, a greater variability in gene expression and copy number across the population may be an advantage for survival (Barja *et al.*, 2017; Negreira *et al.*, 2021). After all, rapid asexual reproduction in *Leishmania* means that only a small number of cells need to survive for full population recovery.

In addition to widespread aneuploidy, *Leishmania* also contain circular sections of DNA (often referred to as episomes) which form by homologous recombination at direct repeated sequences (Ouellette *et al.*, 1991; Grondin, Roy and Ouellette, 1996; Ubeda *et al.*, 2014). While these structures are also found in mammalian cells, they are especially prevalent in *Leishmania* and appear to have a role as a mechanism for circumventing harsh selection pressures (Beverley, 1991). In particular, exposure to drug selection for extended periods of time appears to cause the induction of either deletions or amplifications (often in the form of episomes) of specific loci that are advantageous for survival in the selected drug (Garvey and Santi, 1986; Petrillo-Peixoto and Beverley, 1988; White *et al.*, 1988; Ouellette *et al.*, 1991; Ubeda *et al.*, 2008). It is worth noting that, once again, this is likely to be an effect selected for on a population level, where amplified or deleted loci in a few cells can confer a selective advantage under new conditions. The ability of *Leishmania* to transcribe directly from episomes, apparently without a promoter sequence and no obvious origin of replication, has been a useful tool for transient ectopic expression of selected proteins (Beverley *et al.*, 1984; Curotto de Lafaille, Laban and Wirth, 1992; Ubeda *et al.*, 2014; Clayton, 2016).

1.2.4 Gene regulation

In eukaryotic organisms, effective gene regulation is the key to producing complex and diverse cell-types able to adapt to local conditions and fulfil different roles, despite containing the same genetic information. In multicellular eukaryotes like ourselves, this complex network of regulation allows for the development of a vast array of cell lineages that co-exist within the tissues and organs of the body. In addition to the regulation of differentiation and development, gene regulation controls all of the local dynamic processes in the cytoplasm and organelles of the cell as it performs its specialised function. In single celled eukaryotes, such as *Leishmania* or the perhaps the best studied, baker's yeast (*Saccharomyces cerevisiae*), regulation of genes involved in subcellular processes drives all of the adaptive changes that affect the cell from the beginning to the end of the cell cycle. In *Leishmania* gene regulation must control the differentiation between several distinct lifecycle stages (Figure 1.1).

Gene regulation is often separated as transcriptional and post-transcriptional. Transcriptional regulation, consisting primarily of the assembly of transcriptional regulators and cofactors to gene promoters, was thought to be the main mechanism of gene regulation. Post-transcriptional gene regulation consists of a range of different molecular components including but not limited to: translational regulation, non-coding RNAs, splicing regulation, miRNAs and RNA-binding proteins. Over time it has become apparent that these post-translational mechanisms are both more diverse and more significant than previously thought. In *Leishmania*, post-transcriptional regulation is thought to play an especially large role in gene regulation as a whole (De Pablos, Ferreira and Walrad, 2016; Clayton, 2019). This dependence on post-transcriptional regulation has probably evolved to suit the unusual way genes encoded by *Leishmania* chromosomes are arranged compared to most eukaryotic models, with blocks of many protein coding genes clustered together, often all on the

same strand (Clayton, 2002; Ivens *et al.*, 2005; Rogers *et al.*, 2011). Polycistronic mRNAs like these were first described in the trypanosomatid relative *T. brucei* (Johnson, Kooter and Borst, 1987; Muhich and Boothroyd, 1988; Ullu, Matthews and Tschudi, 1993; Matthews, Tschudi and Ullu, 1994). These arrays superficially resemble bacterial operons, but while proximate genes are commonly co-transcribed, they are not necessarily functionally similar (Clayton, 2016). Because of the production of large polycistronic mRNAs containing many genes with varied functions, transcriptional regulation of individual genes is problematic.

Gene promoters are similarly divergent from those studied in the classical eukaryotic model organisms. Some promoters for RNA-polymerase I and III dependent transcripts have been identified but promoters for most operon-like polycistronic arrays, presumably driven by RNA-polymerase II, are elusive (Gay, Wilson and Donelson, 1996; Uliana *et al.*, 1996; Yan *et al.*, 1999; Stempliuk and Floeter-Winter, 2002; Boucher *et al.*, 2004; Thomas *et al.*, 2006). In addition, very few transcription factors or their binding sites have been identified. Recently, mechanisms driving *Leishmania's* unusual transcription are becoming clearer through the study of DNA base J. Base J is an unusual modified uracil, unknown outside of Euglenozoa, that replaces around 1% of thymine in the *Leishmania* genome, mostly at the telomeres (Genest *et al.*, 2007). While, the majority of base J is telomeric, about 1% was found to be localised to strand switch regions within the chromosome, where the transcription of long polycistronic coding sequence arrays usually terminates, as well as some transcription initiation sites (van Luenen *et al.*, 2012; Genest *et al.*, 2015). Single molecule real-time sequencing has shown that base J is inserted at specific sequences in these regions and is maintained as an epigenetic mark after DNA replication, preventing read-through at transcription termination sites (van Luenen *et al.*, 2012; Genest *et al.*, 2015; Jensen *et al.*, 2021). Most recently, the protein complex associated with base J and a novel base J binding protein have been characterised with a role as regulators of transcription termination in *Leishmania* that is independent of splicing and polyadenylation (Jensen *et al.*, 2021). Further investigation into the transcription initiation and transcription termination sites in this context is crucial for building a better understanding of the divergent way these parasites regulate gene expression.

One example of a characterised *Leishmania* promoter, possibly the best studied, is the spliced leader RNA (SL RNA) promoter that drives expression partially from an episome (Saito, Elgort and Campbell, 1994; Yu et al., 1998; Hitchcock et al., 2007). This is closely tied to another unusual genetic aspect of *Leishmania (*and other Kinetoplastids); a lack of introns. In fact, splicing of individual genes from polycistronic mRNAs requires an entirely different method to that employed by most other eukaryotes. Instead of cis-splicing where an intronic region within the mRNA loops back on itself to form a lariat structure, in *Leishmania,* individual genes are *trans*-spliced using a spliced leader RNA-sequence that is common to all protein coding transcripts (Clayton, 2002; Liang *et al.*, 2003; Hitchcock *et al.*, 2007; Kramer and Carrington, 2011; De Pablos, Ferreira and Walrad, 2016). The 5' intergenic region containing a 35nt spliced leader acceptor sequence was shown to be crucial for gene expression (Curotto de Lafaille, Laban and Wirth, 1992; Flinn and Smith, 1992).

The mechanism of *trans*-splicing is intimately linked with polyadenylation, both occurring simultaneously (LeBowitz *et al.*, 1993; Ullu, Matthews and Tschudi, 1993; Matthews, Tschudi and Ullu, 1994). The scarcity of introns means that splicing-dependent regulatory mechanisms such as the production of alternatively spliced transcripts, are not common in the *Leishmania* transcriptome. The addition of a common spliced leader sequence to the 5' end of all mRNA can be useful when separating the mRNA fraction from other more abundant forms of RNA. More recently, the majority of spliced leader sites and their usage can now be mapped easily using an online web server based tool, SLaP-mapper, which identifies these sequences in the raw read data from RNA-sequencing (Fiebig *et al.*, 2014).

Due to the lack of regular transcriptionally regulated promoters, several attempts have been made to explain gene expression in *Leishmania* through alternative mechanisms. One suggestion is that the ability to amplify and remove copies of specific genes can be used to regulate their expression (Ubeda *et al.*, 2008). However, it is unlikely that this means of expression is suitable or sufficient for rapid responses to changing conditions or stimuli. Layers of post-transcriptional regulation on top of these underlying mechanisms can provide the necessary fine tuning and regulation of genes that need to be active within a narrow time window (Clayton, 2002, 2016; De Pablos, Ferreira and Walrad, 2016). Epigenetic regulation of gene expression also appears to be important in kinetoplastids. Histone modification, particularly methylation and acetylation of histone tails, has been studied extensively in many organisms and is now being studied in detail in kinetoplastids (Staneva *et al.*, 2021). Interaction between reader and writer proteins such as bromodomain containing proteins and histone deacetylases with these epigenetic marks has been described at crucial strand switch regions that identify the start and end of many polycistronic transcripts (Jensen *et al.*, 2021;Jones *et al.*, 2021). Regulation of the chromatin structure propagated from these regions is thought to selectively provide or prevent access to the transcriptional machinery depending on cellular conditions in order to regulate transcript production (Fleck, Nitz and Jeffers, 2021).However, the predominant mechanism of gene regulation appears to be through the stabilisation and destabilisation of mRNA transcripts by RNA-binding proteins and other post-transcriptional regulatory mechanisms that are discussed in section 1.3.1.

1.2.5 Molecular and genetic manipulations in *Leishmania*

Because *Leishmania* have been less thoroughly studied than their *Trypanosoma* relatives, the methods available for investigating their molecular biology have been somewhat limited. Like trypanosomes, transfection of the free-living promastigote stages of *Leishmania* is relatively straightforward, with nucleofection being the most popular method. Intracellular amastigotes however, present a particular challenge to electroporation-based methods of transfection because of the multiple membrane barriers between the culture medium and the parasite nucleus. For this reason, and the comparative ease of culture, the majority of molecular studies in both of these parasites start

with the promastigote stage, despite later lifecycle stages (amastigotes in *Leishmania*) being the major cause of disease. Buffers and programs for nucleofecting trypanosmatids have been optimised and compared for modern equipment (Dean *et al.*, 2015). Recently, developments in the molecular tools available for studying *Leishmania* have accelerated studies involving reverse genetics (Duncan, Jones and Mottram, 2017).

The first methods used for genetically manipulating *Leishmania* included the transfection of episomal constructs for ectopic protein expression (reviewed in Beverley and Clayton, 1993). There are now many variants of the pNUS vectors, that were designed for use both in *Crithidia fasciculata* and *Leishmania* (Tetaud *et al.*, 2002). Any gene specific sequences must be added but all other necessary components such as UTRs, drug resistance genes and even fluorescent tags have been pre-designed so to reduce the setup time. Parasite lines containing episomal constructs can be maintained in a stable state with the use of drug selection. Several drugs that have been successfully used in *Trypanosoma* can also be used in *Leishmania* parasites such as Puromycin, Blasticidin, Neomycin, Hygromycin, Nourseothricin and Phleomycin (Cabañas, Vázquez and Modolell, 1978; Freedman and Beverley, 1993; Joshi *et al.*, 1995; Goyard and Beverley, 2000; Shalev *et al.*, 2013). However, long periods of successive passage and selection can lead to loss of episomes and stability can vary a lot between constructs. A more stable form of genetic manipulation is to integrate constructs into the parasite genome. The first, and most commonly used method for achieving this was to provide a donor DNA cassette flanked by long homology arms (≥500bp) that target the area of integration (Beverley and Clayton, 1993). Following successful transfection, integration then occurs due to complementary base pairing at the regions of homology followed by homologous recombination during DNA replication and repair. Drug selection can be used to select for positive transfectants and maintain unstable constructs, which is especially necessary if the resulting populations are not clonal. One issue with episomal expression is that it is hard to control the number of copies of an episome that are maintained in a cell, leading to very different levels of expression in different transfected cells.

Several constructs have been produced for integrating ectopic protein copies into a ribosomal locus in *Leishmania*, which can be useful for performing phenotypic rescue experiments in knockout cell lines (Mißlitz *et al.*, 2000; Garami and Ilg, 2001). The extremely active ribosomal locus can be advantageous as it produces high protein expression from integrated constructs, but this can also be problematic. Firstly, many gene functions are dosage dependent meaning that beyond a certain threshold of expression there can be significant negative effects or even lethality. This can result in false negative results when attempting to rescue phenotypes. Secondly, there are many ribosomal sequences with very high sequence similarity which can lead to multiple integrations and varying levels of protein expression between independent clones. A useful modular system of vectors and primers was also designed to allow more efficient production of tagged parasites (Dean *et al.*, 2015). A fusion PCR based approach makes it possible to produce the long homology arms needed for efficient recombination in *Leishmania* without having to clone each target region into a vector.

While homologous recombination has been used for many successful purposes in *Leishmania* parasites, there are some situations where it is highly inefficient. When manipulating essential genes for example, there is often a fitness cost or even lethality that means recovering successfully edited parasites is unlikely or impossible. To provide an alternative method capable of removing essential genes in an inducible manner, a DiCre system was developed for use in *Leishmania* (Duncan *et al.*, 2016). In this system, the gene of interest is cloned into a plasmid vector between two LoxP sites and flanked by homology to the region of interest. In the parasite genome, the 'floxed' CDS is integrated in place of one allele and a dimerised Cre-recombinase (Di-Cre) is integrated in place of the other using a different drug selection. Upon rapamycin induction the Cre-recombinase induces recombination at the flanking LoxP sites, inducing excision of the only remaining copy of the essential gene. Successful use of this system has validated essential genes in *Leishmania* and facilitated investigation of otherwise difficult to study targets (Damianou *et al.*, 2020).

The relatively recent introduction of the CRISPR/Cas9 system into *Leishmania* parasites has greatly improved upon existing protocols for genetic manipulation (Sollelis *et al.*, 2015; Zhang and Matlashewski, 2015; Beneke *et al.*, 2017; Martel *et al.*, 2017; Zhang, Lypaczewski and Matlashewski, 2017; Costa *et al.*, 2018). In *Leishmania mexicana*, the most commonly used system utilises a cell line that constitutively expresses both the Cas9 enzyme and T7 RNA-polymerase (Beneke *et al.*, 2017). Donor DNA and sgRNA-encoding DNA is amplified from pre-designed plasmids and from a standard reverse oligonucleotide respectively. Both DNA fragments are transfected into parasites expressing T7 RNA-polymerase that drives sgRNA production from the sgRNA-encoding DNA. The parasites also express Cas9 which is guided to the region of interest by the sgRNA, introducing double-stranded breaks that are repaired using the transfected donor DNA. Primers specific to the gene of interest including 30bp of homology (upstream and downstream in the case of knockout) and a forward sgRNA primer can be designed with a web-based server at www.LeishGedit.net. As with other systems, plasmid variants exist with different combinations of drug resistance as well as different fluorescent protein and small epitope tag sequences. The design of the LeishGedit system means that the same primers can be used for tagging and knockout, greatly increasing efficiency in large-scale screens. The CRISPR/Cas9 system is generally considered more efficient and precise than relying on homologous recombination. To date, knockout efficiency with the most modern systems of nucleofection has not been quantified but indications from routine lab use suggest that knockout efficiency is improved. Use of the CRISPR system in *Leishmania mexicana* in this study is discussed in 5.2 and 6.4.1.

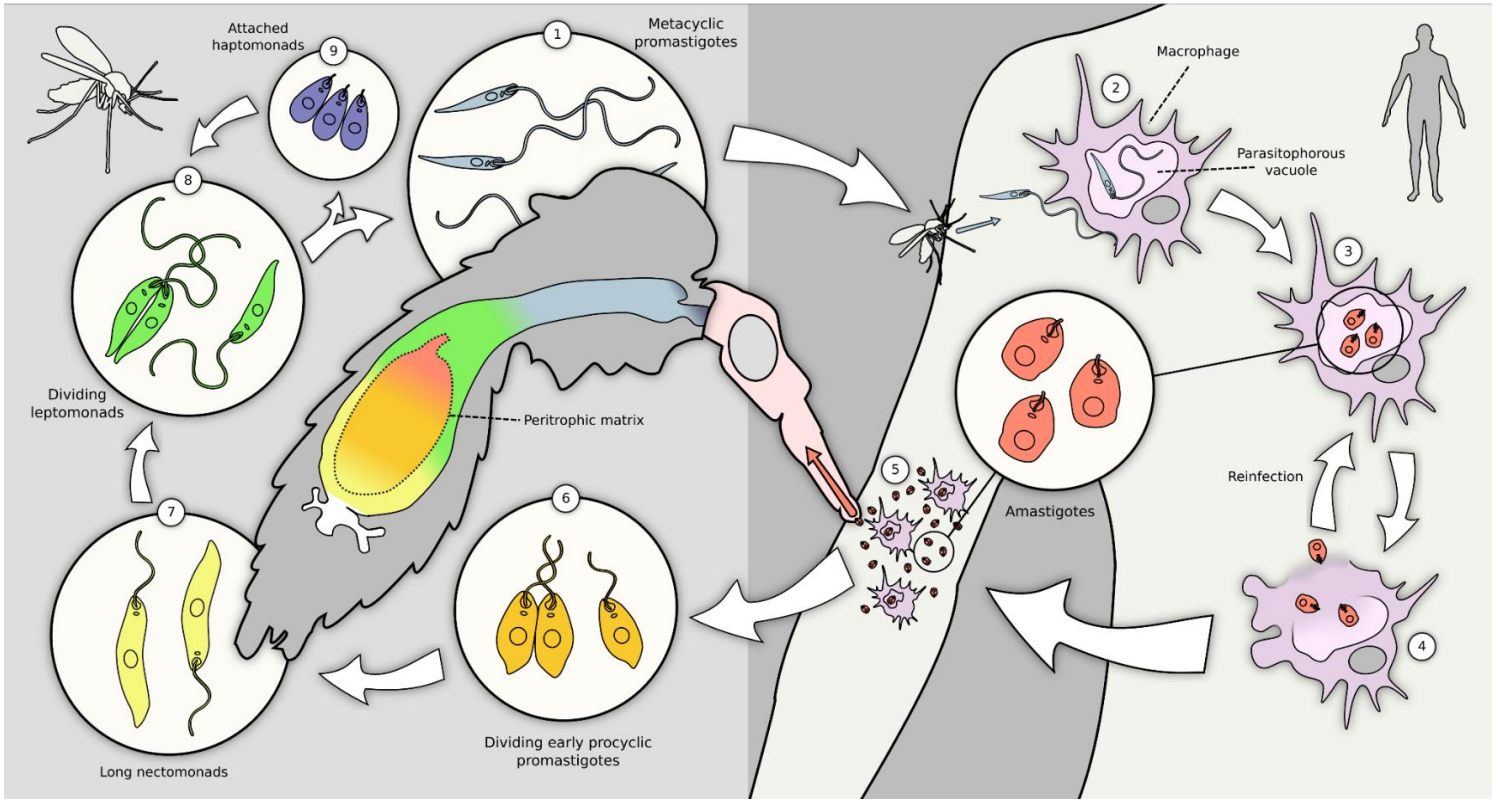


Figure 1.1. The dixenous lifecycle of *Leishmania* parasites. (**1**) Metacyclic promastigotes in the thoracic midgut of the sandfly are transferred to the mammalian host. (**2**) Metacyclic promastigotes are phagocytosed. (**3**) Conditions within the parasitophorus vacuole promote differentiation from the flagellated metacyclic promastigote to the smaller, immotile, amastigotes (amastigogenesis). (**4**) The intracellular amastigotes are highly replicative and may re-infect more macrophages. (**5**) When a sandfly feeds it ingests amastigotes, which are transferred to the abdominal midgut, surrounded by the peritrophic matrix (PM). (**6**) Midgut conditions promote differentiation into procyclic promastigotes, which divide and into early nectomonads. (**7**) The long nectomonads escape the PM and attach to the midgut microvilli. (**8**) Here they differentiate into short nectomonads called leptomonads which migrate further forward in the midgut. The leptomonads mature into metacyclic promastigotes which are infective (metacyclogenesis). (**9**) A small number of leptomonads become haptomonads, attach to the stomodeal valve and block it, increasing the chance of reflux during the next blood meal. (Dostálová and Volf [2012], Kamhawi [2006], De Pablos, Ferreira and Walrad [2016]).

1.3 RNA-binding proteins

1.3.1 An overview of eukaryotic RBPs

Eukaryotic gene regulation is generally thought of as more complex than the simplified systems seen in single celled prokaryotes. Gene expression is driven initially by the complex eukaryotic transcriptional regulation machinery. However, post-transcription, proteins that interact with the individual mRNA strands can affect protein expression in a variety of ways (Figure 1.2). It is important to point out that RNA-binding proteins are a functional group and not defined by any common evolutionary origin or similar structure. Protein kinases for example, have strong degree of sequence and structural conservation in the active site where their core function as a catalyst for phosphorylation is maintained (Knight *et al.*, 2007; Eswaran and Knapp, 2010; Baker *et al.*, 2021). In contrast, RNA-binding proteins include proteins with many different mechanisms for interacting with RNA (Hentze *et al.*, 2018; Van Nostrand, Freese, *et al.*, 2020).

As mRNA is being transcribed there are a range of basal proteins that are involved with protecting it from being degraded (Alberts *et al.*, 2014). As well as being inherently unstable compared to DNA, single stranded mRNA has a 5' and a 3' end that are targets for degradation. To prevent the loss of new transcripts, capping factors bind the 5' end and polyadenylation factors bind the 3' end of the transcript during the transcription process. At the 5' end this involves removal of a phosphate group followed by the addition of a modified guanine residue that is not easily cleaved by cellular exonucleases (Parker and Song, 2004). Cap binding proteins associate with the newly formed cap and provide further stability (Topisirovic *et al.*, 2011).

In most eukaryotes, cis-splicing of introns occurs during transcription, which involves many RNA-binding proteins, scaffold proteins and accessory factors that assemble into the spliceosome (Alberts *et al.*, 2014). As with cap-binding proteins, many of the proteins involved in splicing are highly conserved but mechanisms such as alternative splicing and the incorporation of tissue specific spliceosomal components can have far reaching regulatory effects on gene expression (Wang *et al.*, 2008; Chen and Manley, 2009). Truncated or non-functional protein products can be produced instead of full-length transcripts under the regulatory control of spliceosomal RNA-binding proteins (Lee and Rio, 2015). As RNA-polymerase II nears the end of the mRNA transcript, polyadenylation machinery associates with the 3' end and adds multiple adenosine repeats to form the poly-A tail. In a similar manner to cap-binding proteins, poly-A binding proteins associate with the 3' end of the transcript and protect it from exonuclease activity. Regulation of gene expression by protein binding at specific sites in the 3'UTR can lead to alternative polyadenylation of mRNAs which can regulate the stability, translation efficiency or subcellular location of the transcript (Glisovic *et al.*, 2008).

Mature mRNAs are exported from the nucleus to the cytosol through the nuclear pore when the complement of RNA-binding proteins signals that processing has been completed (Alberts *et al.*, 2014). In the cytosol, mRNA is translated into protein by ribosomes, which are themselves comprised of highly conserved and specialised RNA-binding proteins. Recruitment and assembly of ribosomal

subunits at the 5' end of mRNA transcripts is accomplished by core RNA-binding protein complexes but can be fine-tuned by accessory regulatory proteins (Glisovic *et al.*, 2008). The number of ribosomes bound, their translation speed and the availability of different tRNAs can all be regulated by RNA-binding proteins and directly affect the levels of protein produced from the encoded sequence (Dever, 2002). A large number of post transcriptional regulatory RBPs act by binding specific sequences in the 3'UTR of the mRNA and influencing the stability of the transcript via several mechanisms. Stabilisation or increasing the half-life of an mRNA can be achieved by blocking the degradative action of exo- and endonucleases (Alberts *et al.*, 2014). Transcripts that are stable for longer tend to express more protein. The localisation of mRNA can be affected by RNA-binding proteins through the recruitment of other proteins involved with mRNA trafficking or through exposing and covering localisation sequences directly (Glisovic *et al.*, 2008). Finally, regulatory RBPs can also destabilise mRNAs via several mechanisms including recruitment of exonucleases, miRNA mediated mRNA repression and targeting for RNA-decay at P-bodies (Alberts *et al.*, 2014). In addition to up- or down-regulating protein expression, RBPs play a crucial role in storing mRNAs in stress granules. In response to certain cellular conditions such as nutrient depletion, heat-shock or the presence of antigens such as LPS, some RBPs complex with mRNA to form dense, translationally inactive granules that stall protein expression until it is needed (Buchan and Parker, 2009; Singh *et al.*, 2015).

It was thought for a long time that regulatory RBPs provided a secondary level of control compared to transcription factors. As more evidence has been examined, it has become clear that while some RBPs do act to 'fine tune' the expression of transcriptionally controlled proteins, many also have major regulatory control over the transcripts they bind. Many groups of RBPs have now been studied in great detail. However, identifying the entire complement of RBPs in a eukaryotic genome is a feat that has only recently been attempted. Perhaps surprisingly, large-scale studies have shown that many proteins with diverse primary functions have secondary RNA-binding capacity (Baltz *et al.*, 2012; Castello *et al.*, 2012; Mitchell *et al.*, 2013; Hentze *et al.*, 2018; Van Nostrand, Freese, *et al.*, 2020). This includes many previously characterised proteins that were never recognised as RBPs since they lack canonical RBDs (Scherrer *et al.*, 2010; Tsvetanova *et al.*, 2010; Beckmann *et al.*, 2015).

1.3.2 RBPs in *T. brucei* and related kinetoplastids

Although kinetoplastid parasites are single-celled and diverged hundreds of millions of years ago from the eukaryotes that we use as model organisms, they still have many of the genetic hallmarks of higher eukaryotes and a complex regulatory network. However, as previously discussed, there are many differences in terms of gene regulation. The expression of polycistronic transcripts in *Leishmania* and other kinetoplastids means that post-transcriptional control of gene expression is thought to be the main mechanism controlling the expression of individual genes (Clayton, 2002; Kramer and Carrington, 2011; Kolev, Ullu and Tschudi, 2014; De Pablos, Ferreira and Walrad, 2016). For this reason, a considerable number of studies investigating RBP function have been

published on *T. brucei* and *T. cruzi*. Some of the first post transcriptional control mechanisms that were described in these species were those involved with phosphoglycerate kinase (PGK) B and C regulation (Blattner and Clayton, 1995). The three PGK genes were transcribed on a single polycistronic mRNA, their differential expression across lifecycle stages suggested the involvement of post-transcriptional regulation. While the 3'UTR sequence was identified as necessary for this regulation. AU-rich elements within the 3'UTR were identified as having a destabilising effect on the PGKB mRNA but many of the RBPs that bind them were not identified until later (Quijada *et al.*, 2002). The regulator of differentiation, RBP10, binds these AU-rich elements including at the PGKB, and the surface procyclin EP, 3'UTRs and destabilises these transcripts in a stage specific manner (De Pablos *et al.*, 2017; Mugo and Clayton, 2017). In *Trypanosoma cruzi*, another RNA binding protein *Tc*UBP1, was identified as binding to similar AU-rich sequences and once again having a destabilising effect on the target mRNA (Volpon *et al.*, 2005; Cassola *et al.*, 2015). These examples of single *trans*-regulatory RBPs having major control over gene expression seems to represent a common theme in these parasites.

The majority of kinetoplastid RBPs that have been studied actually stabilise their mRNA targets, in contrast to the previous examples. For example, several Zinc-finger proteins (ZFPs) have been investigated thoroughly, uncovering significant interplay between *trans*-regulators and even co-operative binding (Paterou *et al.*, 2009). A broad range of mRNA targets have been described for ZFPs including those that encode mitochondrial proteins in the case of *Tb*ZC3H20, heat-shock response proteins targeted by ZC3H11, and regulators of differentiation such as ZC3H18 (Benz *et al.*, 2011; Ling, Trotter and Hendriks, 2011; Droll *et al.*, 2013). A recent comparative study identified that ZC3H22 is procyclic specific in *T. brucei* and likely promotes cell division and proliferation due to the increased cell clumping and decreased expression of proliferation genes observed upon RNAi knockdown (Erben *et al.*, 2021). Several related ZFPs, the first RBPs to be directly implicated in the control of differentiation in these parasites, are discussed in section 0 of this thesis (Hendriks *et al.*, 2001; Hendriks and Matthews, 2005; Paterou *et al.*, 2009; Walrad *et al.*, 2009, 2012; Mörking *et al.*, 2012).

Pumilio/Fem-3 (Puf) domain containing RBPs have also been studied in kinetoplastids, interacting with RNA via large arrays of pumilio repeat domains. In *T. brucei*, the protein Puf9 stabilises transcripts during the S-phase of the cell cycle by binding to a specific U-rich sequence in the 3'UTR (Archer *et al.*, 2009). Puf3 has been described as associating with and regulating ribosomal protein-encoding transcripts (Erben *et al.*, 2021). This involvement in the basal machinery of transcription, translation and splicing is common amongst Puf proteins but not universal. For example, in *T. brucei*, Puf2 may be linked to transcriptional elongation, and Puf7 is involved with ribosomal RNA processing (Droll *et al.*, 2010; Jha *et al.*, 2014). Unusual Puf proteins KREPB4 and KREPB5 are involved with the *T. brucei* editosome, along with several other Puf-domain containing proteins (Carnes *et al.*, 2012). KREPB4 was later shown to be essential in both bloodstream and procyclic forms due to its role in recruitment of editosome component proteins that are necessary for successful

post-transcriptional RNA-editing in trypanosomes (McDermott and Stuart, 2017). In *T. cruzi*, the cytoplasmic Puf protein *Tc*Puf6 does not associate with the basal machinery but instead destabilises specific mRNA transcripts, possibly through its interaction with the helicase *Tc*Dhh1 and other elements of the RNA-degradation machinery (Dallagiovanna *et al.*, 2008). A thorough review of Puf proteins in *T. cruzi* is given in Caro *et al*. (2006).

One of the commonest and best studied RNA-binding domains is the RNA-recognition motif (RRM). The SR (Serine/Arginine) protein *Tb*RRM1 contains an RRM and is involved with the splicing machinery, demonstrating a stage specific selectivity for mRNAs and regulating chromatin structure under heat-shock conditions (Naguleswaran *et al.*, 2015). Recently, a homolog of the RRM-containing yeast translational regulator ScSgn1 was characterised in *T. cruzi* (Oliveira *et al.*, 2021). This protein binds to the poly-A tail of mRNAs involved with functions such as nucleic acid binding and endocytosis, possibly acting as a regulator of these processes under stress conditions due to its re-localisation to cytoplasmic granules. Another protein containing a single RRM, RBP9, is upregulated during the infectious bloodstream form of *T. brucei* and preferentially binds mRNAs encoding cytoskeletal proteins (De Pablos *et al.*, 2017; Erben *et al.*, 2021). Previously mentioned, RBP10 is an example of an RBP that is necessary and sufficient for the induction of differentiation (Mugo and Clayton, 2017). In *T. brucei* RBP10 binds specific U-rich motifs in the mRNA 3'UTR and destabilises them by targeting them for degradation. Depletion of RBP10 levels in procyclic form produced parasites incapable of differentiation, whereas overexpression induced differentiation to the bloodstream form. A concurrent study described similar lifecycle stage arrest after RBP10 depletion and presented bloodstream form transcripts being upregulated upon overexpression of both RBP10 and RBP9 (De Pablos *et al.*, 2017). Another important regulator of differentiation, the RRM -containing protein RBP6, has been overexpressed in *T. brucei* procyclic promastigotes which induced differentiation to infective metacyclic promastigotes (Kolev *et al.*, 2012). Although many RRM proteins contain a single RNA-binding domain, it is possible for repeats to occur or even a mixture of domains. For example *Tb*RGG2, an essential protein in both bloodstream and procyclic stages, contains two RRM motifs (Fisk *et al.*, 2008). Like the KREPB proteins described earlier, *Tb*RGG2 is involved with the mRNA-editing machinery, but is involved earlier in the process, helping to initiate editing and pass editing pause points (Ammerman *et al.*, 2010). Other examples include double RNA-binding domain proteins such as DRBD7, which binds mRNAs in bloodstream form and induces minor growth defects when depleted, or DRBD18 which facilitates selective nuclear export of the mRNAs it binds to (Wurst *et al.*, 2009; Erben *et al.*, 2021; Mishra *et al.*, 2021)

Other RBPs have been described in kinetoplastids that do not belong to such large structural families. One example of a more enigmatic RBP, Reg9.1, was identified from a genome-wide RNAi screen in which the authors identified RBPs that repress transcripts that are essential for the infectious stages while the parasites are non-infectious (Rico *et al.*, 2017). When Reg9.1 levels were depleted using RNAi, a subset of transcripts from the infectious, stumpy form were upregulated suggesting that Reg9.1 represses expression of these genes in procyclic parasites. Overexpression of the same protein

in bloodstream forms induced differentiation to procyclic forms and suggesting that the repression of these transcripts is a crucial part of stalling differentiation. Another interesting protein, the pentatricopeptide repeat protein KRIPP11, was characterised as specific binder of G-tract sequences found predominantly in mitochondrial transcripts, prior to mRNA-editing, in *T. brucei* (Kamba *et al.*, 2018). In fact, even some basal splicing factors have been shown to stabilise specific mRNAs in *T. brucei* (Gupta *et al.*, 2013).

In addition to the investigation of individual RBPs, as the technology for RNA-sequencing and proteomics has improved, large scale screens have been used to get a broader picture of both RNA-binding proteins and the regulatory sequences they interact with. The genome wide RNAi screen mentioned previously that identified Reg9.1 is a good example of this, providing information about many potential repressors of transcripts that are upregulated in the infectious stages of *T. brucei* (Rico *et al.*, 2017). In another large scale study, Erben *et al.* (2014) used a tethering assay to identify RBPs that stabilise or destabilise mRNA in *T. brucei*. A small bacterial N-protein was fused to ectopically expressed proteins in cell lines containing an inducible, lethal, *PGKB* reading frame, followed by the *boxB* sequence targeted by the N-protein. RBPs capable of destabilising the lethal transcript when bound to the *boxB* sequence conferred a selective advantage and increased parasite survival. This was achieved on a genome wide scale by incorporating sheared genomic DNA into the N-protein fusion vector which also contained the necessary sequences for tetracycline induction. One disadvantage of this is that many of the sheared genomic DNA sequences contain an incomplete CDS that misfolds or is not expressed and gives false negative results in the assay. This assay was later improved upon by Lueong *et al.* (2016) by producing a library of ORFs rather than sheared fragments. As well as identifying RBPs, characterised RBPs can be used to identify other factors that are involved with parasite differentiation, infectivity or virulence. For example, another recent RNAi screen was carried out on an RBP6 overexpressing *T. brucei* line that simulates differentiation in the tsetse fly vector (Toh *et al.*, 2021). Twenty-two genes were identified that were involved with different stages of this differentiation process including two previously uncharacterised cold-shock domain containing proteins CSD1 and CSD2 which potentially stabilise a range of mRNA transcripts.

Overall, our understanding of RBPs and their importance in kinetoplastids has come a long way, even helping to reveal their importance as post-transcriptional regulators in higher eukaryotes (Kramer and Carrington, 2011; De Pablos, Ferreira and Walrad, 2016). The range of transcriptional targets regulated by RBPs in the examples above demonstrate their ubiquitous involvement in kinetoplastid cellular processes. However, unlike higher eukaryotes, yeast, or bacteria, the interplay between transcriptional, post-transcriptional and post-translational regulation is still poorly described and is likely to form the basis for many future studies. Comparisons between post-transcriptional regulators in different kinetoplastids are now easily produced due to the availability of the CRISPR/Cas9 system and other advancements in the available molecular toolkit.

1.3.3 RBPs in *Leishmania*

In *Leishmania,* RBPs have not been studied for as long or as thoroughly as in trypanosomes. One of the main reasons for this is that the RNAi system, which has been used to great effect in functional studies in *T. brucei,* is lacking in *Leishmania major, L. mexicana* and *L. donovani* (Lye *et al.*, 2010; Atayde *et al.*, 2013). Initially, it was thought to be lacking in *Leishmania* entirely, but in the decade since its characterisation in species of the Viannia subgenus, very few studies (de Paiva *et al.*, 2015; Brettmann *et al.*, 2016) have utilised RNAi in for knockdown experiments in *Leishmania* (none involving RBPs to date). This is primarily due to the lack of an inducible RNAi system in these parasites. With the absence of RNAi, some studies have focused on bioinformatic comparisons of *Leishmania* and *Trypanosoma* RBPs. For example, Kramer, Kimblin and Carrington (2010) used the TriTryp database to identify all CCCH zinc-finger proteins in *Leishmania major, Trypanosoma brucei* and *Trypanosoma cruzi* that have been annotated. They also expanded the number of known CCCH proteins by using a consensus sequence to identify less conserved proteins. Interestingly, a similar number of non-redundant CCCH proteins were identified in *Leishmania major* (54) as in other eukaryotes despite having a much smaller genome.

Some of the few studies on RBPs in *Leishmania* before the introduction of the CRISPR/Cas9 system were focused on core processes such as tRNA transport or the editosome (Adhya *et al.*, 1997; Aphasizhev *et al.*, 2003). The first of these describes the membrane bound tubulin antisense binding (TAB) protein, which is associated with the D-stem loop of $tRNA^{Tyr}$ and is involved with import of tRNAs to the mitochondrion. The second paper investigated the proteins Ltp26 and Ltp28 (orthologues of *T. brucei* proteins gBP21 and gBP25), which were characterised as members of a 100kDa complex involved with the annealing of gRNA to mRNA targets during mRNA editing. Another group of RBPs that were studied before the introduction of CRISPR/Cas9 are the Poly(A)-binding proteins (PABPs). Despite being a well conserved protein in most other eukaryotes, the *L. mexicana* protein *Lm*PABP1 has a divergent RNA-binding domain (RBD4) and, when recombinantly expressed, could not rescue function in yeast (Bates, 2000). While all eukaryotes require PABPs to protect the mRNA poly(A) tail, the differences in transcription and splicing machinery in kinetoplastids likely explain the functional divergence of this protein. A subsequent study showed that, in *L. major,* PABP1 associates with the eukaryotic initiation factor 4G, directly linking it to translation initiation as in other eukaryotes (da Costa Lima *et al.*, 2010). PABP1 is expressed in procyclic parasites in *L. infantum* and binds the poly(A)-tail of mature mRNAs as in other organisms (Guerra *et al.*, 2011). However, PABP2 and the *Leishmania* specific PABP3 were not associated with *Lm*EIF4G3 but interacted with each other, suggesting common involvement in other processes and functional divergence from other characterised PABPs.

The *Leishmania* orthologues of PUF proteins have been studied in some detail in recent years. In *L. infantum,* tandem repeat domain containing proteins were identified as being highly antigenic, with later bioinformatic investigation and serological testing confirming the link (Goto *et al.*, 2006; Goto, Coler and Reed, 2007). Folgueira, Martínez-Bonet and Requena (2010), identified that Puf proteins

were consistent with this finding and investigated them further. They cloned and recombinantly expressed 10 *L. infantum* PUF proteins, which were then challenged with hamster sera, revealing *Li*PUF1 and *Li*PUF2 to be highly antigenic. Subsequent assays with human sera produced similar results for these two proteins. Another PUF protein has been linked to the destabilisation of *Leishmania* mRNAs through a unique class of retro-transposable elements (Azizi, Dumas and Papadopoulou, 2017). Short interspersed degenerated retroposons, or SIDERs, one and two are found in the 3'UTRs of *Leishmania* mRNAs where they promote rapid mRNA turnover (Bringaud *et al.*, 2007; Smith, Bringaud and Papadopoulou, 2009). Degradation of mRNA is caused by endonucleolytic cleavage and can occur both selectively and in a stage specific manner, differentially regulating transcripts (Müller *et al.*, 2010; Müller, Padmanabhan and Papadopoulou, 2010). In a tethering assay, PUF6 was shown to enhance mRNA degradation and reduced the half-life of SIDER2 containing transcripts. PUF6 null mutants also displayed accumulation of SIDER2 containing transcripts, consistent with the role of PUF6 in SIDER2 mediated decay.

Another of the large groups of eukaryotic RBPs are the DEAD-box helicase (DDX) proteins. They contain a core helicase domain with a Asp-Glu-Ala-Asp conserved domain that gives the group its name (Marchat *et al.*, 2015). These proteins are RNA-helicases which catalyse the separation of double-stranded RNA through the ATP-dependent mechanism of their core helicase domain. Many DDX proteins are involved with core cellular processes such as mRNA splicing, which requires the momentary disruption of RNA secondary structures. The DEAD-box helicase Hel67 was identified in *L. infantum* through UV-crosslinked pulldowns of sense and antisense large subunit gamma ribosomal RNA (sLSU- γ and asLSU-γ) (Padmanabhan *et al.*, 2012). This protein was implicated in the protection of asrRNA by blocking its degradation and cleavage during apoptosis-like cell death. Hel67 is the *Leishmania* homolog of the highly conserved helicase DDX3 in other eukaryotes. Further studies by the same group showed that the *Leishmania* homolog interacts with key components of the cellular stress response such as p97/VCP/Cdc48(Padmanabhan *et al.*, 2016).

The RRM containing RBP38 was investigated in *L. amazonensis* and found to bind both nuclear and mitochondrial DNA, in addition to its mitochondrial mRNA stabilising function described in trypanosomes (Sbicego *et al.*, 2003; Lira *et al.*, 2007). This dual function is not surprising as many of the structural characteristics necessary for RNA-binding are well suited to binding DNA with only minor adaptive changes needed. Two RBPs, SCD6 and RBP42, were more recently characterised in the closely related *L. braziliensis* (Nocua *et al.*, 2017). Both proteins were identified as interacting partners of the heat-shock protein Hsp70 during pulldown experiments (Ramírez *et al.*, 2013). The RNA-binding capacity of both proteins was validated for the whole 3'UTR of Hsp70 type two mRNA and for a specific AU-rich sequence within it (Nocua *et al.*, 2017). Further investigation of the interactomes of SCD6 and RBP42 showed significant overlap, suggesting a close functional relationship between the two proteins (Nocua, Requena and Puerta, 2021) . Both proteins are found in the cytoplasm of *L. braziliensis* promastigotes, but relocate to the nucleus during heat-shock along with a decrease in their mRNA levels.

As well as acting as post-transcriptional regulators themselves, many RBPs appear to be post-translationally regulated, providing other mechanisms for modulating protein abundance (De Pablos, Ferreira and Walrad, 2016; Romaniuk, 2016). One of the most common forms of post-translational modification, ubiquitination, was studied as a mechanism regulating mRNA cycling sequence binding protein (CSBP) in *L. donovani* (Bhandari and Saha, 2007). This protein contains two RNA-binding domains (CCCH zinc finger) and an Smr DNA endonuclease domain and is found in ribonucleoprotein granules in the stationary phase of promastigote culture (Bhandari *et al.*, 2011). The Smr domain was able to cleave both DNA and RNA when recombinant protein fragments were assayed but the full-length protein was specifically a riboendonuclease; the first example of this for an Smr domain containing protein. This riboendonuclease activity was downregulated in ubiquitinated CSBP samples compared to the un-modified protein, suggesting mono-ubiquitination as a mechanism of controlling RBP function.

Arginine methylation has also been implicated in the post-translational regulation of RNA-binding proteins. In *L. major*, protein-arginine methyltransferase seven (PRMT7) was found to catalyse RGG specific monomethylation of protein substrates (Ferreira *et al.*, 2014). While PRMT7 is not essential in promastigotes, null mutants displayed increased virulence in a mouse model of infection and overexpression of PRMT7 decreases virulence. A recent study explaining the phenotype of PRMT7 null mutants has linked the increased virulence to excessive neutrophil recruitment and host cell inflammatory response (Diniz *et al.*, 2021). A large-scale methyl-SILAC screen was used to identify proteins in *L. major* that were hyper- or hypo-methylated in the presence or absence of PRMT7 (Ferreira *et al.*, 2020). The targets of PRMT7 methylation were strongly enriched for RBPs, suggesting it has a role as a regulator of other *trans*-regulatory proteins. PRMT7 null mutants negatively impact the ability of the *L. major* Alba3 protein to bind and regulate mRNAs. Alba proteins have also been studied in *T. brucei*, where they are cytoplasmic, stage specific and regulate the translation machinery through interactions with the poly(A)-binding proteins and eIF4E4 (Alba3 specifically) (Mani *et al.*, 2011; Subota *et al.*, 2011). The structural and evolutionary relationships of *Leishmania* Alba proteins were discussed by da Costa *et al.* (2017) who also identified the structural and functional significance of the RGG-box motif in Alba1. Another RBP target of PRMT7 is the small cold-shock domain containing RBP16, a homolog of mammalian YBX2 (Ferreira *et al.*, 2020). In trypanosomes, RBP16 has been characterised as an accessory factor that regulates the editing and increases the stability of specific mitochondrial mRNAs (Hayman and Read, 1999; Pelletier *et al.*, 2001; Miller and Read, 2003; Ammerman, Fisk and Read, 2008; Fisk *et al.*, 2009). In *L. major*, when PRMT7 is absent there is a reduction in RBP16 levels during the stationary phase when N-terminally tagged and cytoplasmic but this effect is not seen in the mitochondrial native protein (Ferreira *et al.*, 2020). In addition to protein modifications as a means of RBP regulation, post-transcriptional modifications of RNA, similar to epigenetic modifications of DNA, can influence mRNA translation and stability. While this has not been heavily investigated yet in *Leishmania*, it appears to be present in most eukaryotes and is likely to play some role in fine-tuning gene regulation. Types of modification include the most studied m^6A methylation, m^5C methylation, cap methylation,

pseudouridylation, uridylation and RNA-editing (such as the mitochondrial RNA-editing in *T.brucei*) It is likely that many more post transcriptional and post-translational modifications involved with RBP regulation remain to be investigated in these parasites.

While individual RBP characterisations have revealed some details of post-translational regulation in *Leismania* parasites, huge gaps in the literature remain, especially in comparison to *T. brucei*. In recent years a few large-scale studies have started to build a bigger picture RBPs and their interactions with mRNA at different stages of the *Leishmania* lifecycle. Nandan *et al.* (2017) used UV-crosslinking and subsequent interactome capture with oligo(dT) magnetic beads to identify the mRNA bound proteome of *L. donovani* amastigotes. A stringent washing protocol followed by protein mass spectrometry analysis identified 79 RBPs with 49 that had no obvious homologs in the human genome. This highlights the huge evolutionary divergence between kinetoplastids and multicellular eukaryotes and the potential for finding pathways regulated by novel RBPs.

Although most studies of RBPs in *Leishmania* have focused on the stabilisation or destabilisation of mRNA transcripts through protein-RNA interactions, there are many other ways that RBPs can regulate these transcripts and their protein products. Before mRNA can be regulated in the cytoplasm and translated it must be exported from the nucleus, where it has been transcribed, via a nuclear pore. The export process itself requires many RNA-interacting proteins that can selectively regulate which transcripts are exported (for example DRBD18 in *T.brucei* [Mishra *et al.*, 2021]). Interestingly, unlike other eukaryotes, trypanosomes appear to export mRNA before transcription and splicing have completed (Goos *et al.*, 2019). Regulation of gene expression can also occur via the actions of proteins involved with the translational machinery. Direct interactions between translational machinery, other associated RBPs and mRNA have been less well studied in *Leishmania* than *T.brucei*, where many RBPs have been associated with the polysomes and active transcription (Klein *et al.*, 2015). Recently a crystal structure and pulldown of interacting proteins has been published for EIF4E5 of *L.infantum* including the previously described protein RBP43 (de Lima *et al.*, 2021). Lastly, RBPs can target mRNAs for storage in stress granules in response to external stimuli (Balagopal and Parker, 2009). Usually, mRNA localised to these granules is translationally inactive but stable, providing a mechanism for temporarily reducing expression of specific genes while others may be upregulated. In *Leishmania* SCD6 and RBP42 have been linked to stress granules as well as the cap binding protein LeishIF4E-3, which was found to be associated with stalled ribosomes under stress conditions (Nocua *et al.*, 2017; Shrivastava, Drory-Retwitzer and Shapira, 2018). The multitude of ways in which RBPs can regulate gene expression in *Leishmania* mean that the study of post-transcriptional regulation in these parasites can involve many different areas of molecular biology.

To gain a broader understanding of *Leishmania* RBPs as regulators of gene expression, the Walrad lab reported the mRNA-bound proteome (RBPome) of *L. mexicana* across four different lifecycle stages: procyclic promastigotes, metacyclic promastigotes, axenic amastigotes 24h post infection and lesion derived amastigotes four months post infection. UV-crosslinking followed by mRNA

pulldowns with oligo(dT) magnetic beads and quantitative mass spectrometry allowed for comparison of mRNA bound proteins between the lifecycle stages. Comparison with additional whole-cell and non-crosslinked proteomes across the four lifecycle stages were used to identify over 1400 mRNA associated proteins. While some of these proteins may not bind mRNA directly, their enrichment in the mRNA-bound proteomes suggests they belong to messenger ribonucleoprotein (mRNP) regulatory complexes. As expected, the majority of RBPs were most highly enriched in the mRNA bound samples during the proliferative procyclic promastigote stage. Importantly, on average, no temporal correlation was observed between the expression of transcripts and the abundance of the same proteins bound to total mRNA. The mRNAs associated with individual RBP candidates were also identified and showed stage specific differences independent of relative protein abundance.

1.4 Aims

Building on the results of the *L. mexicana* mRNA-bound proteome (De Pablos *et al.*, 2019), the aims of this study were to:

- Use bioinformatic approaches to analyse the RBPome data and select a range of RBPs to investigate further.
- With the CRISPR/Cas9 system available in *L. mexicana*, produce a library of barcoded RBP knockout cell lines.
- Screen barcoded RBP knockout cell lines for phenotypes across all major parasite lifecycle stages. Using a bar-seq method, assess relative fitness of all null mutants screened.
- Use whole genome sequencing of barcoded RBP knockout cell lines to assess the accuracy of the deletions performed.
- Validate the results of the bar-seq screen by characterising individual RBPs that presented phenotypes. Tagged RBPs produced for these experiments can be used for future investigation of interacting RNAs.

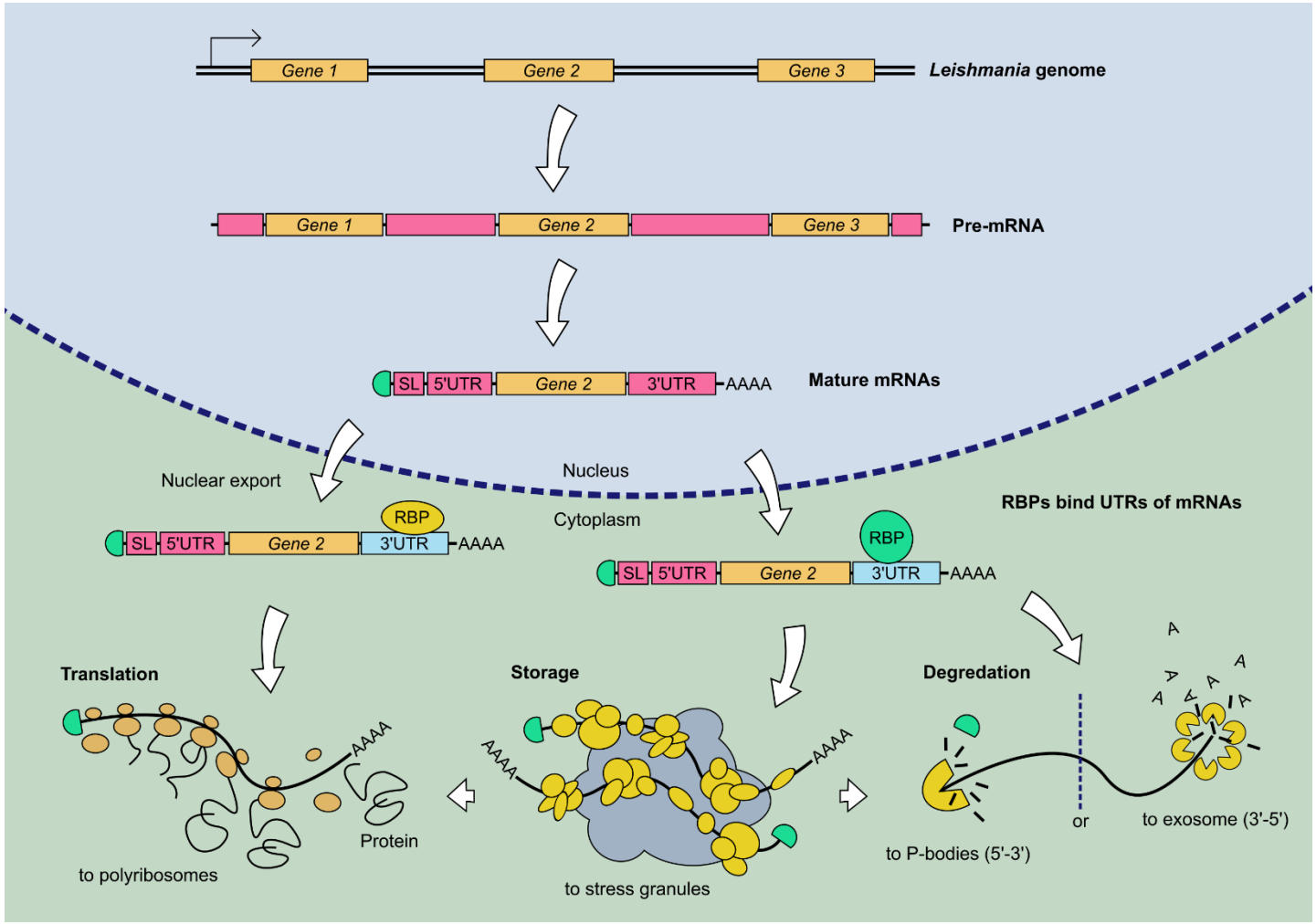


Figure 1.2. RBPs and the control of gene expression. In *Leishmania* parasites, genes are transcribed to produce polycistronic pre-mRNA. The pre-mRNA is trans-spliced, capped and polyadenylated by a range of RNA-binding proteins to produce individual mature mRNAs. Mature mRNAs are exported from nucleus to the cytoplasm where they interact with RBPs, predominantly through 3' UTR binding. RBPs can affect protein expression by altering mRNA translation, targeting mRNAs for storage in stress granules or by targeting mRNAs for degradation via several different pathways.

2 Materials and Methods

2.1 *In silico* methods

2.1.1 Identifying RBP targets

A list of characterised RBDs was produced using data from several large-scale analyses of RBPs in other organisms (Scherrer *et al.*, 2010; Tsvetanova *et al.*, 2010; Matia-González, Laing and Gerber, 2015; Sysoev *et al.*, 2016; Wessels *et al.*, 2016; Oliveira *et al.*, 2017; Romagnoli *et al.*, 2020; Van Nostrand, Freese, *et al.*, 2020), papers on RBPs in *T. brucei* and other related kinetoplastids (Kramer, Kimblin and Carrington, 2010; Erben *et al.*, 2014; Lueong *et al.*, 2016; De Pablos *et al.*, 2019). The search term 'RNA-binding' was also used to identify additional potential RBDs within the InterPro database followed by manual curation of the results (Apweiler *et al.*, 2001; Jones *et al.*, 2014; Finn *et al.*, 2017; Blum *et al.*, 2021).

All 1407 RBP identities from the LC/MSMS data for the *L. mexicana* XL-RBPome were located in TriTrypDB (Aslett *et al.*, 2009). The 'InterPro domain' search type, with 'intersect' selected, was used to search for 520 known RNA-binding domain types and subtypes from the InterPro database (Aslett *et al.*, 2009) within the XL-RBPome proteins. The same search was carried out on all protein identities in the *L. mexicana* genome (*L. mexicana* MHOM/GT/2001/U1103). Proteins with multiple RNA-binding domains from different categories were included in both categories. Data were presented graphically using Microsoft excel 2016, R-studio1.1.383 and Corel Draw 2017. Protein levels representing the triplicate mean for each lifecycle stage analysed by mass spectrometry were used when comparing protein level in different lifecycle stages. Numbers of peptide identities most abundant in each lifecycle stage were compared using Microsoft excel.

For analysis of RBPs in the XL-RBPome with unknown RBDs, the XL-RBPome amino acid sequences were downloaded from TriTrypDB and subset into individual FASTA files, each containing 100 sequences (Aslett *et al.*, 2009). These were uploaded to the *cat*RAPID signature server and results were outputted as .pdf files. Data were merged into a single '.csv' file in Microsoft Excel and analysed in R-Studio using ggplot2 and Tidyverse packages (RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.).

2.1.2 Dendrograms

Dendrograms were used to assess the homology of *Leismania* RBPs such as LmxM.15.0130 (DDX27) to their orthologs in model organisms as well as their paralogs in *Leishmania*. Amino acid sequences were retrieved from the TriTrypDB and stored in FASTA format (Aslett *et al.*, 2009). FASTA files were processed using MEGA7 (Kumar, Stecher and Tamura, 2016), aligned using a MUSCLE algorithm and trees were calculated using the 'maximum likelihood' with default parameters (MUSCLE: Gap Open = -2.90, Gap Extend = 0.00, Hydrophobicity Multiplier 1.20, Max

Memory = 2048MB, Max Iterations = 16, Cluster Method = UPGMA, Min Diag Length (Lambda) = 24. Maximum likelihood tree: Statistical Method = Maximum Likelihood, Test of Phylogeny = None, Substitutions Type = Amino Acid, Substitution Model/Method = Jones-Taylor-Thornton (JTT) model, Rates among Sites = Uniform Rates, Gaps/Missing Data Treatment = Use all sites, ML Heuristic Method = Nearest-Neighbor-Interchange (NNI), Initial Tree for ML = Automatic (Default – NJ/Bio/NJ), Branch Swap Filter = None, Number of Threads = 3.)

2.1.3 Primer design

Primers were designed using one of several different methods depending on their function. Specialised primer design such as CRISPR/Cas9 sgRNA or donor DNA primers are described under the relevant sub-sections. In general, small primers for analysing cloning steps, performing colony PCRs or checking for the presence or absence of an integrated resistance cassette in the *Leishmania* genome were designed manually using SnapGene software (from Insightful Science; available at snapgene.com) or Snapgene viewer as an aid. Parameters for manual design were: no shorter than 15bp, attempting to keep GC content near 50%, less than 5°C difference in melting temperature between primers where possible and including G/C clamps on either end of each primer. All primers designed were compared against the predicted template sequence using BLASTn or PrimerBLAST and only used if predicted to be at least 100x more likely to bind the target sequence than the next off target site (Altschul *et al.*, 1990; Ye *et al.*, 2012; Agarwala *et al.*, 2016).

2.1.4 KO morphology quantification

Images of *Leishmania* parasites, all taken with on a Zeiss Axio Observer microscope, were imported in '.czi' format into ImageJ (FIJI) so that scale data were preserved (Schindelin *et al.*, 2012). The scale of the image was checked manually before measuring. The segmented line tool was used to measure length of cells in the images from the tip of the cell body to the base of the flagellum, through the centre of the cell body. Cell widths were measured at 90° to the "length" measurement, at the widest point. Measurements were collected and imported into R-Studio for further analysis and data processing (RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.). Five fields of view were used for both WT and KO cells with over 300 cells measured for each data group.

2.1.5 RBP conservation: BLASTp

The following taxids were queried using the predicted amino acid sequences of all 67 RBPs included as candidates for CRISPR knockout screening: *Leishmania mexicana* MHOM/GT/2001/U1103 (taxid:929439), *Trypanosoma brucei brucei* TREU927 (taxid:185431), *Trypanosoma cruzi* strain CL Brener (taxid:353153), *Saccharomyces cerevisiae* S288C (taxid:559292), *Drosophila melanogaster* (taxid:7227), *Homo sapiens* (taxid:9606). NCBI BLASTp server was used with the following default parameters: Max target sequences = 100, Short queries = true, Expect threshold = 0.05, word size =

6, max matches in a query range = 0, Matrix = BLOSUM62, Gap costs = Existence:11Extension:1, compositional adjustments = conditional compositional score matrix adjustment (Altschul *et al.*, 1990). No filters or masks were applied. For each RBP, the match with the highest percentage sequence identity from each species proteome was plotted in R-Studio using the ggplot2 package to produce a heatmap with % colour intensity directly proportional to percentage sequence identity (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.)(Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.). If no positive results for a species were found in the top 100 matches the identity was assumed minimal and set to 0% for the sake of plotting.

2.1.6 Genome sequence analysis

Thirteen out of twenty-eight RBP null mutant cell lines from the knockout screen were sequenced using Illumina HiSeq. Analysis and processing of the sequencing reads provided by Novogene were carried out using a multi-node computing cluster (Viking) using the following modules: SeqKit (version 0.12.0), FastQC (version 0.11.7), MultiQC (version1.7), BWA (version 0.7.17), SAMtools (version 1.10) and mosdepth (version 0.2.8). Raw sequencing reads were subjected to quality checks using Seqkit, FastQC and MultiQC (for details see (Eva Kyriacou, 2021)). Sequence reads were ordered and mapped to the *L. mexicana* T7/Cas9 parental strain genome (JM6571) (Gluenz lab) using the BWA and SAMtools modules. Aligned sequences were output as '.bam' (BAM) files with corresponding '.bai' files for visualisation (work carried out by Katherine Newling and Eva Kyriacou [Eva Kyriacou, 2021]).

Aligned sequence reads were visualised in IGV (Integrated Genomics Viewer, version 2.9.2) and compared to the reference genome of JM6571 produced by the Gluenz lab using nanopore sequencing. Blasticidin *(BSD)* and Puromycin (*PUR)* resistance sequences were included with the reference genome for additional comparisons. Individual '.bam' files corresponding to RBP null mutant lines were loaded as separate tracks and compared at the locus of interest. Reads with distant partners were examined individually for each cell line to check for off target integrations. Relative read coverage at each locus of interest and the drug resistance regions was calculated using the mosdepth module comparing total reads mapped to the region of interest divided by the median coverage of the chromosome of interest. The read coverage of the *BSD* and *PUR* regions for each cell line were compared to median read coverage of the chromosome they were integrated into. Relative read coverages were presented graphically as a heatmap using the ggplot2 package in R-Studio (RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/., Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.).

2.1.7 TargetP and *cat*RAPID

Full amino acid sequences for *L. mexicana* RBP coding regions were downloaded from the TriTryp database and compiled into a single FASTA file including all RBP candidates from the knockout screen (Aslett *et al.*, 2009). The FASTA file was uploaded to the TargetP server where a deep learning algorithm (described in Armenteros *et al.*, 2019) was used to assign scores where >0.5 is a positive prediction of either a signal peptide (sp = secretory pathway involvement), a mitochondrial targeting peptide (mtp = mitochondrion or other plastid localisation), or any other targeting peptides (otp = all non-classical targeting peptides detected). Scores provided were visualised as a heatmap using R-studio with the ggplot2 package. An expanded FASTA file of all RBPs in the XL-RBPome was compiled and submitted in batches of 100 to the *cat*RAPID signature server for scoring RBPs with non-canonical RBDs for their RNA-binding potential (Livi *et al.*, 2015). Scores for all proteins were extracted and imported into R-studio in '.csv' format for analysis and visualisation. Scores were divided into two groups, those with known RBDs and those without, and compared statistically. Testing showed that the data had unequal variances but were normally distributed. The data were also unpaired, meaning a Welch's two sample t-test was suitable. Statistics were carried out in R-studio. The data were subset in R-studio into scores for all proteins containing each type of known RBD. These datasets were presented using the ggplot2 package.

2.2 Cell culture

2.2.1 Species and strains

The *L. mexicana* strain JM6571 was used throughout. All RBP knockouts were produced in this background considering the trade-off between CRISPR/Cas9 efficiency and parasite infectivity. This cell line has T7 RNA polymerase and *S.pyogenes* Cas9 integrated into the ribosomal locus to drive expression of sgRNAs and introduce guided dsDNA breaks in the genome (Beneke *et al.*, 2017). The *L. mexicana* M379 strain was used as an additional control in some experiments and *L. major* strain CC1 was used for the continuation of work on RBP16 and PRMT7.

2.2.2 Culture media and conditions

For experiments requiring promastigotes, *Leishmania* cells were maintained in M199 medium supplemented with 40 mM HEPES, 10% FBS, 100 U penicillin/ml, 100 µg/ml streptomycin, 100 µM adenine and 0.0005% hemin (Morgan, Morton and Parker, 1950). Grace's medium was used for differentiation into the metacyclic promastigotes (stationary phase) or axenic amastigotes (Grace, 1962). Grace's medium was prepared as follows: Added to one bottle of Grace's medium (powder): 600ml Milli-Q H_2O, 0.004166M (0.35g/L) $NaHCO_3$, 10% FBS, 100 U penicillin/ml, 100 µg/ml streptomycin and 10ml BME vitamins (Sigma, 100×). The medium was mixed and adjusted to pH 5.5. All media were filter sterilised before use using a 0.2µm pore filter. Cells were cultured horizontally in 25ml flasks at 26°C or in 175ml flasks under the same conditions for protein extraction or large-scale experiments.

2.2.3 Passaging parasites

Parasites were passaged at a minimum of once every four days to retain log phase culture and to stall differentiation. The standard passage dilution used was 1/100. A dilution of 1/1000 was used to remove dead cells when recovering clonal populations allowing for accurate PCR analysis.

2.2.4 Growth curves and cell counting

Growth curves were produced growing parasites in either Grace's medium or M199 using starting concentrations of either 1×10^6/ml or 5×10^5/ml. Cells were counted using a Beckman-Coulter Z2 series particle counter by diluting cell culture 1:100 in 10ml of Isoton (Beckman-Coulter) and mixing thoroughly before counting. In stationary phase, pipetting through a blunt syringe was sometimes used to break up rosettes that add error to the cell counts and, at high density, block the cell counter. For some applications, cells (e.g. bone marrow macrophages) were counted using a haemocytometer (improved Neubauer) and trypan blue stain. Cells were diluted 1:1 in 0.4% trypan blue solution (Lonza, solution prepared in 0.85% NaCl) and a 10ul aliquot was applied to the haemocytometer for counting under a light microscope. Stained cells were indicated as dead and ignored in counting.

2.2.5 *Leishmania* transfection

Cell cultures for transfection were grown to log phase (1×10^6-1×10^7/ml) in M199 medium. For nucleofection (with the Amaxa 2B system), a minimum of 5×10^6 cells were pelleted at 2000g for 10mins in a 15ml falcon tube. 3×cytomix transfection buffer was produced as follows: 200mM Na_2HPO_4, 70mM NaH_2PO_4, 15mM KCl, 150mM HEPES pH7.3 and was aliquoted and frozen at -20°C. 70µl of a defrosted aliquot of 3×cytomix was added to the appropriate amount of water and mixed with 3.15µl 10mM$CaCl_2$. This solution was used to resuspend the pelleted cells before being added to the PCR products requiring transfection, to give a total volume of 210µl/transfection in 1×cytomix. The transfection mix was quickly transferred to a nucleofection cuvette and pulsed once using the programme X001 (optimal in *Leishmania* as described in Dean *et al.*, 2015) on the Amaxa 2B nucleofector (Lonza). Prewarmed medium and a micro-Pasteur pipette was used to transfer the cells from the cuvette to a flask containing 10ml M199 medium at 26°C. Cells were incubated for at least six hours before adding drug for selection.

2.2.6 Amastigote differentiation

A flask containing 10ml of M199 medium was inoculated with 1×10^5 cells and grown for seven days. The culture was centrifuged at 2000g for 10min and resuspended in Grace's medium to a final concentration of 1×10^6/ml. The culture was incubated at 33°C (Bates *et al.*, 1992) with 5% CO_2 for 96h in vented flasks to allow complete differentiation to axenic amastigotes.

2.2.7 Metacyclic promastigote purification

A solution of 20% Ficoll in Milli-Q water was mixed 1:1 (2ml+2ml/extraction) with DMEM to make 10% Ficoll. 4ml of the 10% Ficoll was added to the bottom of a 15ml falcon tube and 6ml of stationary culture resuspended in DMEM was layered on top gently so that the layers did not mix. The tubes were centrifuged at 1300g for 10min to separate metacyclic promastigotes which were harvested in the top 9mls (8mls for more stringency) and transferred to a new 50ml falcon. The cells were mixed with 20ml of DMEM to wash and then pelleted at 2000g for 10min. The supernatant was removed, the pellet resuspended in 10ml of Grace's medium and transferred to a new flask to recover, producing a culture enriched in metacyclic promastigotes (Späth and Beverley, 2001).

2.2.8 Macrophage isolation culture and infection

For the RBP bar-seq screen and other related experiments, bone marrow macrophages were isolated from BALB/c mouse femurs and stored as follows. Filtered resting medium (DMEM with D-glucose 4.5g/L, sodium pyruvate 0.11g/L, 1mM L-glutamine) was pre-warmed to 37°C. Tibia/femur bones were washed in 70% ethanol in a petri dish and cleaned of all flesh using a scalpel. The bones were cut carefully at 90° leaving a circular cross section and the warm DMEM was pipetted through with a syringe to flush out the bone marrow-derived macrophages. Live macrophages were counted using a haemocytometer and 0.4% trypan blue solution diluted 1:1. Macrophages were diluted and plated in sterile petri dishes at a concentration of 5×10^5/ml in 10ml of active medium (DMEM as previously with 1% Penstrep, 20% FBS and 30% L-cell (stimulating factor secreted by L929 cells, as described by Weischenfeldt and Porse [2008]). Incubation at 37°C triggered differentiation into activated macrophages over seven days with medium being replaced twice during this period. Macrophages were harvested for freezing with a cell scraper and multiple washes of warm filtered DMEM. Cells were counted again using trypan blue as before and were aliquoted at either 5×10^6 or 1×10^7 cells/cryovial in 90% foetal bovine serum and 10% DMSO.

For macrophage infections with *L. mexicana*, cryovials of bone marrow-derived macrophages isolated previously were carefully thawed in a 37°C water-bath and transferred to a 50ml falcon tube. 10ml of pre-warmed DMEM (1mM L-glutamine) was added, allowing the cells to equilibrate. Cells were centrifuged at 1200rpm (200g) for 6-10min, after which the supernatant was removed and replaced with active medium. 2.5×10^5 cells were added per well in a six well plate and incubated at 37°C until they were fully adherent and displayed 'activated' morphology (24h after plating). Macrophages were infected with stationary phase parasites at a ratio of 6:1. Each pool of Δ*RBP* clones was used to inoculate a six well plate, three wells for harvesting after 24h and three for harvesting at 72h. After 6h macrophages were washed thoroughly in warm DMEM (5x) to remove attached promastigotes before covering the cells once more with resting medium (5ml/well in 6 well plates). Cells were scraped as before to detach and three wells were collected for each RBP null mutant pool and transferred to a 15ml falcon tube. The remaining cells were collected by further washing and scraping the well and added to the falcon tube before being centrifuged at 3700g for

10min. The resulting pellet was resuspended in 2ml of PBS to wash and pelleted once more at 17,900g for 5min. The supernatant was removed and the harvested, and pelleted cells were stored immediately at -80°C for DNA RNA or protein extraction later.

2.2.9 Drug selection

Leishmania cells were selected for various genetic modifications using a range of drugs. Where a combination of drugs was required for selection, such as all RBP null mutant clones, Blasticidin and Puromycin were used due to their compatibility with the T7/Cas9 expressing line JM6571. Drugs were added directly to the culture medium and mixed thoroughly. For selection after electroporation or nucleofection, the cells were given a minimum of four hours to recover before drugs were added. Hygromycin and SAT were not used for selection of T7-RNA polymerase or Cas9 in JM6571 cells after CRISPR editing was performed (Joshi *et al.*, 1995). When selecting for parasites expressing drug resistance markers, drugs were added to media of healthy cultures at the following concentrations: Blasticidin = 10µg/ml, Puromycin = 50µg/ml, Hygromycin = 50µg/ml, Nourseothricin = 10µg/ml, G418/Neomycin = 15µg/ml.

2.2.10 Pooling Δ*RBP* clones

All 28 RBP null mutant clones and three control cell lines were grown in 10ml of M199 medium in T25 flasks until log phase ($>1\times10^6$ and $<1\times10^7$). Lines were passaged several times depending on growth rate to ensure all lines were at a comparable stage for pooling. A fixed number of cells (3.22×10^6) from each cell line were mixed in a 50ml falcon. Mixing was repeated five times to give six independent pools which were transferred into 100ml Grace's medium in a T175 flask after a 10min centrifugation at 3200rpm to give a final concentration of 1×10^6/ml.

2.2.11 Storage of *Leishmania* lines

Leishmania cells were frozen while proliferating, in log phase, using 10% DMSO and 40% FBS in M199 (freezing medium filtered using a 0.2µm pore filter). 1ml of culture was added to 1ml of freezing medium in a cryovial and mixed well before freezing at -80°C in a Mr Frosty container (Nalgene). Cell lines were stored both at -80°C and in liquid nitrogen. To recover cells from frozen, the 2ml aliquots were defrosted gently on ice, immediately transferred to 10ml of M199 medium in a 15ml falcon tube and centrifuged at 3200rpm for 10mins to pellet. The supernatant containing DMSO was removed and the cells were resuspended in fresh M199 or Grace's medium and incubated at 26°C to recover.

2.3 Cloning

2.3.1 Bacterial transformation

Bacteria used for standard transformations with plasmid DNA were either DH5α or XL1 blue *E. coli*. First, 50ul of supercompetent cells were thawed on ice. Plasmid DNA (0.1–50ng) was added to thawed cells and mixed gently. An empty vector was transformed separately as a negative control. Cells were incubated on ice for 30 minutes and then heat-pulsed in a 42°C water-bath for 45 seconds. Tubes were incubated on ice for 2 minutes before the addition of 500µl of SOC medium/LB broth for an outgrowth of 1hr at 37°C (shaking at 250rpm). LB-Agar plates with antibiotic were used for streaking out both the transformation mixture and a 1/10 dilution. Bacterial plates were incubated at 37°C overnight. Plasmids that had stability issues due to recombination between repeated regions were transformed into stbl3 *E. coli* (Invitrogen). Transformation with stbl3 bacteria was carried out as above with minor differences as specified in the manufacturer's protocol.

2.3.2 Mini preparation

For extraction of amplified plasmid DNA from bacterial pellets, a Nucleospin mini preparation kit (Machery Nagel) was used according to the manufacturer's protocol. In brief: A 5ml culture of plasmid containing *E. coli* was pelleted at 15,000g for 3mins. The supernatant was removed and 250µl of buffer A1 was added, before vortexing to resuspend the pellet. 250µl of buffer A2 was added and the mixture inverted gently 6-8 times and incubated at room temperature for 5min until the lysate cleared. 300µl of buffer A3 was added and the mixture inverted gently to mix. The mixture was pipetted into a Nucleospin column and centrifuged at 11,000g for 1min before discarding the flow-through. The membrane was washed with 500µl of buffer AW by centrifuging again at 11,000g for 1min and then washed with buffer A4 using the same conditions. The membrane was dried by centrifuging at 11,000g for 3min. Milli-Q water was pre-heated to 70°C and 25µl was added directly to the membrane before resting for 5min. To elute, the column was centrifuged a final time at 11,000g for 1min. The final step was repeated with a fresh 25µl of Milli-Q water to increase the yield.

2.3.3 Midi preparation

For the extraction of greater quantities of plasmid DNA from bacterial pellets a Hi-speed plasmid midi preparation kit (Qiagen) was used with either the manufacturer's protocol or the following optimised protocol for stbl3 *E. coli*. In brief: 50-150ml of overnight culture containing plasmid transformed *E. coli* was pelleted at 10,000g for 15min in 50ml falcon tubes. Pellets were frozen at -20°C. All pellets were combined in 6ml of buffer P1 to which 6ml of P2 was added. The tube was mixed by inversion throughout a 5min incubation at room temperature. The resulting mixture turns blue. 6ml of prechilled buffer P3 was added and mixed thoroughly until the blue colour disappears. The resulting mixture was transferred to a QIA filter cartridge and incubated at room temperature for 10min. A Hi-Speed tip was equilibrated with 4ml of buffer QBT, allowing it to flow through the

resin until it no longer dripped. The incubated mix was filtered into the equilibrated Hi-Speed tip and allowed to drain through completely. A 20ml wash of buffer QC was added to the Hi-Speed tip and allowed to drain through completely. 5ml of buffer QF was warmed to 56°C and used to elute DNA from the tip into a 15ml falcon tube. The DNA was precipitated by mixing with 3.5ml of 100% isopropanol and incubating for 5min. The liquid containing DNA precipitate was passed through a QIA precipitator using a syringe, retaining the DNA. 2ml of 70% ethanol was slowly passed through the precipitator using a syringe to wash the DNA. The precipitator was dried by passing air from an empty syringe until no more liquid could be removed. Finally, 500µl of buffer TE heated to 56°C was used to elute the DNA from the precipitator. The eluate was transferred back through the precipitated two more times to increase the yield.

2.3.4 Bacterial clone selection and colony PCR

When amplifying plasmids for cloning purposes, *E. coli* transformed using protocol **2.3.1** were removed from 37°C incubation and inspected for colony growth. To check for those containing a successful cloning step, 5-20 clones were picked using a sterile pipette tip to scrape half a colony. The half-colonies that were picked were dispersed in 30µl of Milli-Q water before being heated to 95°C for 10min to lyse the bacterial cells. A 25µl PCR reaction was set up as follows: 2µl of the lysate used as DNA template, Taq polymerase (NEB, 0.125µl), 10x standard Taq buffer (2.5µl), dNTPs (10uM stock, 0.5µl), 10µm forward primer (0.5µl), 10µm reverse primer (0.5µl), Milli-Q H_2O (20.825µl). Individual colonies were selected based on PCR results and picked by scraping half a colony with a pipette tip. For mini- and midi-preparations the tip was then transferred directly to LB; either 5ml or 50ml respectively.

2.4 DNA/RNA methods

2.4.1 RNA extraction from *Leishmania*

RNA extractions were carried out using the Direct-zol RNA MiniPrep kit (ZYMO Research) with minor adjustments to the supplied protocol. Between 1-5x10^7 *Leishmania* cells were pelleted at 13,000g for 5min and washed two times with sterile PBS before being resuspended in 500µl of Trizol reagent (Ambion/Life technologies) and frozen at -80°C. An equal volume of ethanol was added to the defrosted sample, mixed and the whole sample added to a Zymo-Spin IIC Column in a collection tube and centrifuged at 13,000g for 30s. The flow-through was discarded. RNA samples were then treated with Turbo DNAse (Invitrogen). The manufacturer's protocol was followed but incubated for at least 30mins at room temperature to completely remove DNA (DNA presence checked by PCR). 400µl of RNA PreWash was added to the column and centrifuged, as before, discarding the flow-through. The wash step was repeated for stringency. 700µl of RNA Wash Buffer was added to the column and centrifuged as before. A clean collection tube was added and the column centrifuged at 13,000g for at least 3min to dry the membrane. The column was transferred to an RNAse free tube

and 40µl of Milli-Q water at 50°C was added to elute the RNA. This step can be repeated to increase yield but at the cost of having a lower final concentration of eluted RNA.

2.4.2 DNA extraction from *Leishmania*

Leishmania parasites were extracted from culture and centrifuged at 13,000g in a 1.5ml Eppendorf to form a pellet with 1-2×10^7 cells. The pellet was washed in 200µl of PBS before being resuspended in the same volume and frozen at -20°C. Several different kits were tested with a comparison of three options being produced in Figure 4.6. The first two are column-based purifications (Qiagen: DNeasy kit and Omega: Blood and Tissue kit) and the final method (PCR-Bio) is a faster protocol based on centrifugation.

2.4.2.1 DNeasy Blood & Tissue Kit (Qiagen)

The DNeasy tissue extraction protocol states that a maximum of 5×10^6 cells should be used but for *Leishmania*, increasing this to 1-2×10^7 increased the DNA yield considerably without a noticeable reduction in purity. 20µl of proteinase K was added to the defrosted cells with 200µl of buffer AL and mixed by vortexing vigorously before incubating for a minimum of 10min at 56°C. 200µl of ethanol (100%) were added, mixed with the sample and loaded into a DNeasy Mini spin column with a collection tube. The spin column was centrifuged at 6000g for one minute and the flow-through discarded. 500µl of buffer AW1 was used to wash the column, centrifuging as before and discarding the flow-through. Placing the column in a new collection tube, 500µl of buffer AW2 was added and centrifuged as before. The membrane was dried by centrifuging for 5min at 17,900g and discarding the flow-through. The spin column was carefully transferred to a fresh 1.5ml Eppendorf for elution. DNA was eluted in a minimum of 15µl Milli-Q water heated to 70°C, repeated with a fresh aliquot of 15µl Milli-Q water to give a total of 30µl. The membrane was incubated with water for 5min instead of 1min suggested in the protocol.

For the RBP bar-seq screen several precautions were taken. Tubes for sample collection were sterile RNAse/DNAse free and were kept separate to the lab environment to reduce the probability of barcode contamination. For each time point of cultured cells, two samples of 2ml were taken after gently mixing the culture flask. The cells were pelleted at 13,000rpm for 5min, supernatant was removed and the pellets were frozen immediately at -80°C. Samples derived from infected macrophage culture were processed using the same protocol. For isolating total DNA from infected footpads, frozen tissue from -80°C was kept cold on dry ice. Chilled forceps and scalpel were used to break the sample into fine pieces on a glass slide to improve the subsequent digest step. The sample (~15mg) was mixed with 180µl of buffer ATL and 30µl of proteinase-K from the DNeasy kit and incubated overnight (at least 12h) at 37°C. The standard DNeasy animal tissues spin column protocol was then followed with the following changes. During step six the column was dried at 17,900g for

5min. In step seven 30µl molecular grade water was heated to 70°C and incubated on the column for 5min before being centrifuged at 6000g for 1min as usual.

2.4.2.2 E.Z.N.A.® tissue DNA kit (Omega Bioscience)

Starting with 1-2×10^7 cells in 200µl of PBS as described earlier, 25µl of proteinase K solution was added along with 220µl of BL buffer and vortexed to mix. The mixture was incubated at 70°C for at least 10min, vortexing again during incubation. After 220µl of 100% ethanol were added to the mix it was vortexed thoroughly. The sample was transferred to a Hi-Bind DNA mini-column with collection tube and centrifuged at 13,000g for 1min before discarding the flow-through. 500µl of buffer HBC was added and centrifuged at 13,000g for 30s discarding the collection tube after and replacing it. 700µl of DNA wash buffer was added and centrifuged as in the previous step. This washing step was repeated and the empty column was centrifuged at 13,000g for 5min to dry the membrane. 25µl of Milli-Q water at 70°C was added directly to the membrane and incubated for 5min before eluting by centrifugation at 13,000g for 1min. The last step was repeated with a fresh aliquot of water to increase the yield in a final volume of 50µl.

PCRBIO Rapid Extract PCR Kit

First, 1-2×10^7 cells were pelleted by centrifuging at 13,000g for 5min and were added to 20µl of buffer A and 10µl of buffer B diluted in 70µl of Milli-Q water. The sample was vortexed to mix and incubated at 75°C for 5min vortexing twice during incubation, and at 95°C for 10min to deactivate the proteinase. 500µl of Milli-Q water were added and mixed before centrifuging the sample at 13,000rpm for at least 1min to pellet the debris. The supernatant containing DNA was then transferred to a clean Eppendorf tube.

2.4.3 Ligation

T4 DNA ligase was used to join DNA fragments at an appropriate ratio using the following reaction mixture: vector DNA 50ng, insert DNA at 3:1 (insert:vector) ratio, T4 DNA ligase buffer (10X) 2µl, T4 DNA ligase 1µl, nuclease-free water to 20µl. The 20µl reaction was left at 4°C for 20min and then an ice-bath at 16°C overnight.

2.4.4 Sanger sequencing

Samples were prepared for Sanger sequencing following the manufacturer's instructions using the Mix2seq kit (Eurofins). Analysis was carried out using SnapGene software (from GSL Biotech; available at snapgene.com). In brief: 15µl purified DNA template was added to a Mix2seq barcoded tube at the following concentration:

o Plasmid DNA up to 15 kbp: 50 - 100 ng/µl

o PCR products: 150-300 bp: 1 ng/µl; 300-1000 bp: 5 ng/µl; > 1000 bp: 10 ng/µl

A 2µl sample of the sequencing primer at 10pmol/µl (10µM) was mixed with the sample. The tubes containing a minimum of 17µl were sealed with a sterile cap and sent for collection and processing.

2.4.5 Genomic DNA preparation for Illumina HiSeq

Genomic DNA was extracted as in **2.5.1.1** or **2.5.1.2.** Those extracted as in **2.5.1.3** were further purified using the columns from **2.5.1.2** as the PCRBio method produced DNA with protein impurities when assessed by Nanodrop. Concentrations of all clones included were adjusted to 30ng/µl in 30µl to give 900ng of DNA per clone. Whole genome sequencing and quality control was carried out by Novogene using the Illumina HiSeq platform. Novogene sequencing consisted of a NovaSeq S4 300 cycle run (150 bp paired end), and demultiplexed FASTQ files were returned. Processing of FASTQ files is described in **2.1.6**.

2.4.6 Barseq library preparation for Illumina HiSeq

PCRs amplifications of the barcode region from each experimental timepoint were produced using the protocol in 2.5.4. All PCRs were cleaned using protocol 2.5.5 and assessed by nanodrop. Addition of Nextera Illumina sequencing adaptors was carried out with an additional PCR from 3µl of each amplification as template. This was performed using primers from the Illumina Nextera XT index kit, with a unique combination of i5 and i7 barcodes for each sample, in a 7 cycle PCR reaction using NEBNext Q5 Hot Start HiFi PCR mastermix (NEB). Indexed samples were then purified using 0.9X AMPure XP beads, quantified using the QuantIT reagent (ThermoFisher), and pooled at approximately equimolar ratios. Magnetic beads were used to purify the PCR products with Nextera adaptors added. The beads were washed twice in 80% ethanol. Samples were eluted in TE buffer. All 54 samples were pooled in equal proportion. The sample pool quality was assessed using a Bioanalyzer (Agilent) followed by in house quality control at Novogene. Amplicon sequencing was carried out by Novogene using the Illumina Hiseq platform.

2.4.7 Whole genome sequencing

DNA extractions were carried out using the E.Z.N.A.® tissue DNA kit (Omega Bioscience) as detailed in 2.4.2.2. Genomic DNA (900ng) of RBP null mutant clones was sent to Novogene for library preparation and sequencing in a volume of 30µL at a concentration of 30ng/µL. Sequencing was carried out with a NovaSeq S4 300-cycle run (150 bp paired end), and demultiplexed FASTQ files were returned for analysis.

2.5 PCR

2.5.1 sgRNA cassette

Reactions were carried out in 20µl set up as follows: F-primer (100µM) 0.4µl, R-primer (100µM 1143/OL6137) 0.4µl, dNTPs (10mM), Q5 polymerase 0.2µl, Q5 buffer (5×), Milli-Q water 14.6µl. PCR conditions: 98°C for 30s, 35× (98°C for 10s, 60°C for 30s, 72°C for 15s), 72°C for 10min, held at 16°C. The overlap between F and R primer means template is not required.

2.5.2 Donor cassette

Reactions were carried out in 40µl reaction set up as follows: F-primer 0.8µl, R-primer 0.8µl, DNA template 30ng/µl (pGL2662 [Blasticidin-R], pGL2667 [Puromycin-R) 0.8µl, dNTPs (10mM), Q5 polymerase (NEB, Cat.No. M0491L) 0.4µl, Q5 polymerase buffer (5×) 8µl, Milli-Q water 28.4µl. PCR conditions: 94°C for 5mins, 45×(94°C for 30s, 65°C for 30s, 72°C for 2min15s), 72°C for 7min, hold at 12°C.

2.5.3 RBP analytic

Genomic DNA was extracted using a spin column kit after a proteinase K digestion (DNeasy [Qiagen], E.Z.N.A.® Tissue DNA Kit [OMEGA]). At least two PCRs were performed: one to assess resistance cassette integration at the correct genomic locus and the other to assess the presence or absence of the CDS of the target gene (Figure 4.4). A standard reverse primer to either *BSD* or *PUR* sequences was used with a gene specific upstream forward primer, roughly 500bp upstream of the ATG, to check cassette integration. For the gene PCR two primers were designed to give an amplicon of roughly (varying depending on optimal primer design for the gene of interest) 500bp from within the CDS. Primers were designed to have a Tm of ~60°C when used at 200nM with Taq DNA polymerase (NEB) (See Table 7.2.). PCR conditions: 95°C for 30s, 15×(95°C for 30s, touchdown:70°C-55°C[-1°C/cycle] 30s, 68°C for 60s), 15x(95°C for 30s, 55°C for 30s, 68°C for 60s), 68°C for 5min, held at 12°C. Results were analysed on 1% agarose gels visualised with SYBERsafe (ThermoFisher).

2.5.4 BarSeq amplification

PCR amplification of the barcode region of pooled RBP null mutant clones was carried out with primers 1766 and 1768 (Figure 4.4, Table 7.2.). 50µl PCR reactions were set up as follows; 5×CG rich buffer 10µl, 5×Hifi reaction buffer 10µl, Milli-Q water 25.25µl, gDNA template 2µl, dNTPs (10µM stock) 0.15µl, F-primer 1µl, R-primer 1µl, Q5 polymerase (NEB M0491L) 0.25µl. PCR conditions were as follows: 98°C for 5min, 28×(98°C for 30s, 60°C for 30s, 72°C for 10s), 72°C for 5min, hold at 16°C . If the final concentration after PCR column purification was below 10ng/µl, PCRs were repeated and pooled during purification. L0_1- M168_6 were amplified for 25 cycles. Samples A24_1 – A72_6 were amplified with 30 cycles and samples FP3W_1 – FP6W_6 were amplified with 28 cycles.

2.5.5 PCR column purification

For all standard applications a NucleoSpin gel and PCR clean-up kit (Machery Nagel) was used according to the manufacturers protocol. Milli-Q water heated to 70°C was used instead of elution buffer depending on downstream applications. For column purification of barcode amplification PCRs, the QIAquick® PCR & Gel Cleanup Kit (Qiagen) was used for greater purity. Five volumes of buffer PB was added to one volume of PCR sample and mixed before transferring the whole mixture to a QIAquick column and centrifuging at 13,000g for 30s. The flow-through was discarded and 750µl of buffer PE was added to the column before centrifuging to wash the membrane and discarding the flow-through. The membrane was dried by centrifuging for 3min at 13,000g. The DNA was eluted with 15µl of buffer EB heated to 70°C, incubated on the membrane for 5min and centrifuged into a clean Eppendorf at 13,000g for 1min. Elution was repeated with a fresh 15µl aliquot to increase yield.

2.6 Protein methods

2.6.1 Protein extraction from *Leishmania*

Cells were counted and a minimum of 3×10^7 were centrifuged at 2000g for 10mins and the supernatant was removed. The pellet was resuspended in 1ml of filtered PBS and centrifuged as before to wash. The supernatant was removed and, for each 1×10^7 cells, 20µl of Laemmli buffer (Laemmli, 1970) was added at 95°C and incubated at this temperature for 5min before being stored at -20°C.

2.6.2 Western blotting

SDS-PAGE gels for western blotting were made for purpose using the method outlined in Sambrook and Russell, 2001 with a BioRad 1.5mm glass buffer dam and Protogel acrylamide solution (National Diagnostics). For most purposes acrylamide gels were made up with a 12 or 15% resolving gel (for 15ml: H_2O 4.9ml, 30% acrylamide mix 6ml, Tris-Cl [1.5M, pH 8.8] 3.8ml, SDS [10%], 0.15ml, 10%

ammonium persulfate 0.15ml, TEMED 0.006ml) and a 5% stacking gel (for 5ml: H_2O 3.4ml, 30% acrylamide mix 0.83, tris-Cl [1.5M, pH 8.8] 0.63ml, SDS [10%], 0.05ml, 10% ammonium persulfate 0.05ml, TEMED 0.005ml). Gels were run in a mini protean tank (Biorad) at 100V for ~2hr or until the dye front reached the end of the casing. For protein transfer to Immun-Blot PVDF membrane (pore size 0.2µm, Biorad), a Novex semi-dry transfer system (Invitrogen) was used at 20V for 1hr. Ponceau stain (0.4% Ponceau-S in 1% acetic acid used as in Salinovich and Montelaro, 1986 visualised protein transfer from gel to membrane. Ponceau stained membranes were imaged using the ChemiDoc MP imaging system (Biorad) with the blot/UV/stain-free sample tray and the 'Ponceau' setting. Membranes were de-stained with Milli-Q water with three 3-minute washes. Membranes were blocked for 1hr in 15ml of blocking buffer (5% milk in TBS-t [0.05% tween]). Primary antibodies were used as in Table 7.1 made up in TBS-t (0.05% tween) with 1% milk. Secondary antibodies most commonly used were anti-rabbit HRP (NIF824) and anti-mouse HRP (NIF825), both used at 1:10,000 dilution in 1% milk TBS-t (0.05% tween) incubated with the membrane for 1hr shaking. Three ten-minute washes were carried as before using TBS-t (0.05% tween). ECL prime western blotting agent (Amersham) was used to image membranes in a ChemiDoc MP imaging system (Biorad) using blot/UV/stain-free sample tray and the 'chemiluminescence' setting.

2.7 Immunofluorescence

2.7.1 Staining fixed *Leishmania*

For locating the mitochondrion in live cells, Mitotracker stain was used as a preliminary step to the standard immunofluorescence protocol. When mitochondrial visualisation was not required, the protocol starts with the fixation step. M199 medium was pre-warmed to 26 ºC and Mitotracker deep red 633 (Invitrogen) added at 500nM in enough medium for 100µl of medium/Mitotracker mix per cell pellet. $2x10^6$ cells were pelleted by centrifuging at 2000g for 10min. The supernatant was removed, and cells gently resuspended in 100µl of warm Mitotracker-containing medium. These were incubated at 26 ºC for 30mins to allow uptake. Mitotracker must be added to live cells where a pH gradient across the mitochondrial membrane facilitates its uptake. Manufacturer's recommend 50-500nM but a paper describing use in *T. brucei* suggest 1µM (Field *et al.*, 2004). Here, 500nM was used. ~2ml of PBS was added and the resulting mix pelleted at 2000g for 10mins. The supernatant was removed before adding paraformaldehyde in PBS (4%, pH 7.5) to cells which are fixed for 20min at room temperature. Added PBS and centrifuged at 2000g for 10min. Supernatant was removed then the pellet resuspended gently in PBT (2% TritonX100 + glycine (0.1M or 10mg/mL) in PBS), incubated at room temperature for 20min then centrifuged again at 2000g for 10min and resuspended in 100µl of PBS. Prepared 'Superfrost plus' slides (ThermoFisher) with an ImmEdge PAP pen (Vector Laboratories) or paraffin wax, marking a roughly 5mm diameter circle for each sample. 10µl of fixed parasites were applied to the slide which was left for 15-20min until parasites have adhered but not so that the droplet dries completely. Wash steps were carried out on slide three times in 1%BSA in PBS. Blocking solution was removed and replaced with 10µL PBT

(0.1%TritonX100, 1%BSA in PBS) either with or without primary antibody (e.g. Sigma anti-HA mouse monoclonal 1:500, Table 7.1). Slides were incubated for 1hr in a humidity chamber. The sample was washed three times in PBT (0.1%Triton), once in PBS. 10µL of secondary antibody (e.g. Alexa 594 goat anti-mouse/rabbit, Table 7.1) [1:2000] in PBT (0.1% TritonX100, 1% BSA in PBS) was added and slides were incubated 1hr in humidity chamber again. Slides were washed two times in PBT (0.1%TritonX100) and once in PBS. After drying, ~10µL droplet of Vectashield (DAPI included: Vector Laboratories) was applied to the sample along with a coverslip which was sealed with nail varnish. Western blots were stripped of antibody using Restore PLUS western blot buffer (ThermoFisher) before a second blocking step and addition of additional primary and secondary antibodies (e.g. anti-NMT loading controls, Table 7.1).

3 Analysis of the RBPome and selection of RBPs for gene deletion

3.1 Introduction

RNA-binding proteins have been relatively poorly characterised in *Leishmania* species, with the majority of knowledge being derived from experimentation in the related kinetoplastid *Trypanosoma brucei* (De Pablos, Ferreira and Walrad, 2016). The isolation of the mRNA-bound proteome (RBPome) represents a major development in our understanding of the interaction between *trans*-regulatory RBPs and their mRNA targets in *L. mexicana.* By comparing protein levels across the four isolated lifecycle stages (procyclic promastigotes, metacyclic promastigotes, macrophage-derived amastigotes and lesion-derived amastigotes) it was possible to identify many stage specific RBP-mRNA interactions (De Pablos *et al.*, 2019). Increased protein expression or mRNA-binding in a single lifecycle stage can suggest stage-specific functions but further experiments are required to confirm this.

In this chapter, the UV crosslinked RBPome (XL-RBPome) was searched with a list of known RNA-binding domains from better studied organisms to identify distantly related RBPs in *Leishmania*. At the start of this project, the annotations of *Leishmania* genomes in the TriTryp database were relatively incomplete (Aslett *et al.*, 2009). Since then, many improvements and updates have been made, especially in linking hypothetical genes to their annotated orthologs in other related kinetoplastids. Despite this, there are still a large number of genes with no more than an automatically generated name based on areas of sequence homology. The automatically generated name often refers to conserved protein domains found within the predicted protein sequence. The type of RNA-binding domain can suggest the role of an RBP in the parasite and which types of RNA it may interact with. Additional analysis of the amino acid sequences of *Leishmania* RBPs also revealed potential mitochondrial or secretory pathway localisations for RBPome proteins as well as gaining insight into their relative evolutionary conservation. Research on orthologs of these RBPs in *T. brucei* also was carried out to gain further insight into the potential functions of the different RBPome proteins.

In order to investigate the role of RBPs as *trans*-regulators and their potential involvement in differentiation and infectivity, the work presented in this chapter was used as a foundation for selecting RBPs for gene deletion and screening in Chapter 3.10. Analysis of the RBPome by De Pablos *et al.*, (2019) was crucial for selection of candidates. Further analysis was carried out on the RBPome and is presented here, facilitating an informed selection of candidates using several criteria.

The aims of these analyses were to:

- Determine which mRNA associated proteins in *L. mexicana* had characterised RNA-binding domains.
- Compare both RNA-binding and protein expression in different lifecycle stages.
- Test whether bioinformatics can be used to predict RNA-binding capacity for *L. mexicana* RBPome identities without known RBDs.
- Collate published information and currently available bioinformatic tools to provide information on *L. mexicana* RBPs that will be useful for further investigations.
- Select a range of RBPs from different protein families to produce a library of *L. mexicana* RBP knockout lines that can be screened for phenotypes.

3.2 Criteria for RBP selection

When selecting RBPs to study as regulators of important processes in *Leishmania*, the focus was placed on those that are likely to play a role in differentiation, infectivity or virulence. The CRISPR/Cas9 system developed for use in *L. mexicana* has been adapted to improve efficiency (Beneke *et al.*, 2017; Baker *et al.*, 2021) but higher throughput knockout screening with commercially developed, bespoke construct libraries has yet to be optimised. With this in mind, it was important to choose a realistic number of RBPs to be investigated in the time frame of a three-year PhD. In total 67 RNA-binding proteins were selected for the knockout screen. Proteins that are likely to be found in mRNP regulatory complexes were preferred over those with no evidence for their functional involvement in regulation. The presence of an RBP in either the UV-crosslinked or non-crosslinked RBPome was a major criterion for selection. 61 out of 67 attempted knockouts were for genes that were included in the UV crosslinked RBPome (De Pablos *et al.*, 2019). Identities that were enriched in the UV-crosslinked or non-crosslinked samples compared to the whole cell proteome were favoured along with those that appeared to be expressed or bind mRNA in a stage specific manner in the human infectious stages (metacyclic promastigotes, macrophage derived amastigotes and lesion derived amastigotes). Constitutively expressed and proteins with high evolutionary conservation across eukaryotes were largely excluded from the bar-seq screen as they have less potential for identifying druggable targets and for understanding *Leishmania* specific processes. Similarly, proteins named as likely constitutive components of core basal machinery of transcription, translation or splicing were mostly excluded. A small number of proteins were included in the screen due to their relevance to previous work on post-translational regulation of RBPs, for

example methylation (Ferreira *et al.*, 2020). A flow-chart of the complete candidate selection process is presented in Figure 3.7.

3.3 RBPs with characterised RBDs

To gain insight into the function of the mRNA bound proteins isolated in the *L mexicana* RBPome, gene IDs from the UV-crosslinked RBPome were entered into the TriTryp database search function and searched for known RNA-binding domains (RRM, Pumilio, CCCH zinc-finger, DEAD-box helicase, cold-shock domain and others) (Supplementary Table 7.3)(Aslett *et al.*, 2009). The same search was also carried out in the whole *L. mexicana* genome to analyse RBPome coverage. InterPro domain codes represented the most complete functional annotations in the *L. mexicana* genome so were used for this analysis over alternatives such as PFAM or GO terms (which are partly derived from PFAM/InterPro domains [Carbon *et al.*, 2021,Ashburner *et al.*, 2000]). Literature on RNA-binding proteins screens was used to identify the most common types of RNA-binding domain for inclusion in this search (Scherrer *et al.*, 2010; Tsvetanova *et al.*, 2010; Matia-González, Laing and Gerber, 2015; Sysoev *et al.*, 2016; Wessels *et al.*, 2016; Oliveira *et al.*, 2017; Romagnoli *et al.*, 2020; Van Nostrand, Freese, *et al.*, 2020). Additional RNA-binding or nucleic acid binding domain codes were added by searching the InterPro database. All proteins bound to mRNA, in the UV-crosslinked RBPome and containing a known RNA-binding domain are represented in Figure 3.1 (De Pablos *et al.*, 2019). Domains that are often found in proteins linked to the basal machinery of transcription, translation or splicing have been grouped in the category 'basal'. While this removes many low priority proteins, it does not identify all RBPs in the basal machinery as the function of many RBPs cannot easily be predicted based on homology. For example, many DEAD-box helicases (not removed) are involved in splicing or translation but others are regulators of specific mRNA transcript stability. The commonest and best described domains are labelled, and all others have been grouped in the category 'other'.

L. mexicana produces proteins containing most of the common RNA-binding domains: RRM, DEAD-box helicases, CCCH zinc fingers, Puf and KH, but appears to have many mRNA-bound proteins with either novel RNA-binding domains or classical RBDs with sufficient sequence divergence to avoid detection. These five common RBDs: RRM, DEAD-box helicases, CCCH zinc fingers, Puf, and KH were enriched in terms of percentage of containing proteins in the XL-RBPome compared to the whole genome. CCCH zinc finger proteins were less successfully isolated in the XL-RBPome compared to the other common RBDs. Proteins containing RBDs related to the basal machinery of transcription, translation and splicing were highly enriched in the XL-RBPome, likely due to their higher abundance and their binding of a broad spectrum of mRNAs. Where a genomic analysis of proteins describes all predicted coding sequences, isolation of a proteome using a pulldown protocol includes some biases for abundant proteins, strength of binding, or in the case of UV-crosslinked samples, number of interacting points (which become covalent RNA-protein bonds). Despite this, a diversity of RNA-binding domains was well represented in the XL-RBPome with proteins containing all common RNA-binding motifs.

3.4 Conservation of *L. mexicana* RBPs

Many RNA-binding proteins that are involved in core cellular processes exhibit high evolutionary conservation across eukaryotes. Alternatively, proteins that have diverged to become involved in *Leishmania* or kinetoplastid-specific pathways are likely to have less sequence similarity to their closest homolog in other eukaryotic models. To investigate relative conservation of the different RBPs selected for CRISPR knockout screening, predicted protein sequences for all 67 candidate RBPs were compiled from TritrypDB in FASTA format (Aslett *et al.*, 2009). The NCBI BLASTp server was used to compare all candidate RBP amino acid sequences to the predicted protein sequences from the genomes of the following species: *Leishmania mexicana, Trypanosoma brucei, Trypanosoma cruzi, Saccharomyces cerevisiae, Drosophila melanogaster, Homo sapiens* (2.1.5). A heatmap of percentage sequence identity to the protein with the lowest e-value in the predicted proteomes of each of these species shows considerable variation in conservation between different RNA-binding proteins (Figure 3.2). The whole sequence of each predicted RBP was compared to include differences that are found outside the RBD. Only two proteins, LmxM.34.4950 and LmxM.36.5820, had no matches outside of *Leishmania.* LmxM.29.3370 was the only protein with a homolog in *T. brucei* but no matches in any of the other organisms. 12 RBPs had no BLASTp matches outside of trypanosomatids, possibly indicating involvement in clade specific cellular processes. The Y-axis was ordered from highest mean percentage identity to lowest across all species considered and reveals that some of the most conserved RBP are the DEAD-box helicases including LmxM.21.1552, LmxM.36.1850, LmxM.31.0400, LmxM.15.0130, LmxM.05.0360. This is consistent with their tendency for involvement with essential eukaryotic cell processes (Erben, Chakraborty and Clayton, 2013; Kellner *et al.*, 2015; Marchat *et al.*, 2015; Ozgur *et al.*, 2015; Radhakrishnan *et al.*, 2016; Zheng *et al.*, 2017). It should be noted that, while this information about RBP conservation is useful, it is not a prediction of direct orthologs. In practice, the top percentage identity hits from *T. brucei* or *T. cruzi* are often directly orthologous when synteny is checked on chromosome maps but at lower levels of homology and in the more distant, *Homo sapiens*, *Saccharomyces cerevisiae* and *Drosophila melanogaster* proteomes, further analysis would be needed to trace the closest ortholog accurately. Despite being closely related, many RBPs with the lowest e-values from BLASTp comparison in *T. brucei* had less than 50% identity to their *Leishmania* orthologs.

3.5 RBPs without characterised RBDs

There were many proteins associated with mRNA that do not appear to contain a known RNA-binding domain (Figure 3.1). The majority of these are listed in the TriTryp database as 'Hypothetical Conserved' proteins (Aslett *et al.*, 2009). Since RNA-binding domains were identified automatically based on sequence identity, it is likely that many proteins in *Leishmania* are so divergent that the domains are no longer recognised and annotated. In addition to these, many of these proteins may have previously uncharacterised RNA-binding domains or maybe closely associated proteins from the same mRNP complexes but not directly binding RNA. In an attempt to assess their propensity to

bind RNA, predictive computational methods were explored. The *cat*RAPID signature server developed by the Tartaglia lab (Cirillo, Agostini and Tartaglia, 2013; Livi *et al.*, 2015) was used to assess the RNA-binding capacity of the entire UV-crosslinked RBPome (Figure 3.3a). Overall, being a dataset isolated from mRNA, if the algorithm is predicting RNA-binding, then the median score for the dataset should be above 0.5 (the threshold for a positive RNA-binding prediction). In the case of the *L. mexicana* XL-RBPome, the median score was 0.45 and the mean was 0.47 suggesting that RNA-binding is not predicted as well by *cat*RAPID signature as it is in other species. The difference between well characterised RBD containing proteins and other mRNA associated factors was also investigated. RNA-binding scores were higher on average for proteins containing known RNA-binding domains than those that did not. A Welch Two Sample t-test reported a significant difference in means of these two groups: t = -4.31, df = 178, p-value = 2.64e-05. However, the variance in scores for each group are high. This means that for any individual hypothetical protein, this method doesn't reliably differentiate RBPs from non-RBPs in *Leishmania* parasites. Comparative analysis of mean RNA-binding scores for proteins grouped by RBD presence shows that some domains are more easily recognised as RNA-binding by *cat*RAPID signature than others (Figure 3.3b). RRM domains, for example, have been well studied in many model organisms and *Leishmania* proteins containing them had a mean score of 0.64.The lack of conservation between the RNA-binding proteins from model organisms that the algorithm has been trained with and those from *Leishmania mexicana* is likely to explain the relatively poor results. Because of the difficulties assessing the RNA-binding potential of these proteins and a lack of structural and functional information, proteins without a recognised RNA-binding domain were not prioritised for the RBP bar-seq screen.

3.6 Relative abundance of RBPs with characterised RBDs in four lifecycle stages

The mass spectrometry data from the XL-RBPome allowed for the quantitative comparison of RBP levels between the different lifecycle stages (De Pablos *et al.*, 2019). Both the whole XL-RBPome and the subset containing known RBDs were sorted by highest mean protein level in each of the four stages analysed (procyclic promastigotes, metacyclic promastigotes, macrophage derived amastigotes and lesion derived amastigotes). Figure 3.4a highlights the lifecycle stage in which the protein identities in the UV-crosslinked RBPome are eluted from mRNA in the highest quantity. More proteins are most abundant in the isolation from procyclic culture than from the other lifecycle stages, followed by lesion derived amastigotes and macrophage derived amastigotes. This trend correlates RBP abundance with the most metabolically active and replicative lifecycle stages. Conversely, fewest proteins were most abundant in the metacyclic promastigotes which are less transcriptionally and translationally active, as well as less replicative (Gossage, Rogers and Bates, 2003; Moreira *et al.*, 2014; Dillon *et al.*, 2015; Kloehn *et al.*, 2015; De Pablos *et al.*, 2019). If only proteins with known RBDs are analysed in the same way, the majority are most abundant when isolated from procyclic promastigotes (67%). Proteins that were most abundant when isolated from

the human infectious stages make up only a third of the RBD containing proteins assessed (Figure 3.4b).

3.7 Subcellular localisation of homologous RBPs

When choosing *L. mexicana* RBPs for the CRISPR/Cas9 bar-seq screen, it was our aim to include many proteins that are likely to be *trans*-regulators of differentiation or infectivity. Often, these regulatory RBPs bind mRNA outside the nucleus where it is translated. This is in contrast to many RBPs involved in the core machinery of transcription and splicing, which are often highly conserved, specific and localised (at least transiently) to the nucleus (Stern *et al.*, 2009; Názer, Verdún and Sánchez, 2011; Fernández-Moya *et al.*, 2012; Gupta *et al.*, 2014; Wongsombat *et al.*, 2014; Das *et al.*, 2015; Naguleswaran *et al.*, 2015; Dean, Sunter and Wheeler, 2017; Oliveira *et al.*, 2021). Although this appears to be a trend, there are undoubtedly some regulators of specific mRNAs that are nuclear and closely involved with transcriptional or splicing machinery meaning they cannot be overlooked. While there is limited published work on RNA-binding protein localisations in *Leishmania*, many RNA-binding proteins have been tagged in the distantly related kinetoplastid *Trypanosoma brucei* in association with the TrypTag project (Dean, Sunter and Wheeler, 2017). To gain insight into the expected subcellular localisation of the selected *L. mexicana* RBPs, a table of the observed localisations from *T. brucei* was produced (Figure 3.5). It is important to note the large degree of evolutionary separation (several hundred million years [Cavalier-Smith, 2016; Harmer *et al.*, 2018]) between *Leishmania* and *Trypanosoma* mean that orthologous proteins don't necessarily perform the same functions or have the same localisation. However, being related species (kinetoplastida), where this information is available, it can be valuable and provide useful functional insight to build upon. This is especially true for RBPs that are well conserved in kinetoplastids (Figure 3.2). A majority of the selected candidates have a cytoplasmic distribution in *T. brucei*. Many RBPs (12/67 examined from the TrypTag database [Dean, Sunter and Wheeler, 2017]) appear to have different or additional localisations depending on whether they were N or C-terminally tagged, highlighting the need for small unobstructive tags and tagging attempts at both ends of the amino acid sequence (Ferreira *et al.*, 2020). RBPs where the direct ortholog was not obvious, tagging was not attempted and tagging failures are coloured in grey.

3.8 TargetP predictions.

Another way of predicting the location and function of an uncharacterised protein is to identify organelle targeting sequences in the amino acid sequence. Targeting sequences are ubiquitous in eukaryotic proteins, are usually located at the N-terminus and specifically target the recruitment of the protein to the secretory pathway, the mitochondria or other plastids depending on the specific sequence. Some of the commonest and best described targeting sequences are the N-terminal targeting peptides. Presence of a targeting peptide can reveal a lot of information about the subcellular localisation and the function of a protein. Knowing the location of a targeting peptide is also important so that disruption can be avoided when tagging. Discovering cleavage sites is crucial

because tags added upstream of the site can be removed in the cell by proteolytic cleavage. One of the most successful bioinformatic prediction pipelines for detecting targeting sequences can be accessed through the TargetP2.0 server (Armenteros *et al.*, 2019). The amino acid sequences for candidate proteins for CRISPR/Cas9 knockout were compiled from TriTrypDB and uploaded in FASTA format to the TargetP2.0 server which uses deep learning and neural networks to accurately predict several types of targeting sequences (Aslett *et al.*, 2009). The resulting scores for signal peptide sequences (secretory pathway involvement) and mitochondrial localisation signals are presented in Figure 3.6. By default, TargetP considers scores of greater than 0.5 to be high confidence matches. Three gene products, *LmxM.36.5820, LmxM.36.5845* and *LmxM.28.0825*, had a high probability of containing a mitochondrial localisation signal. Several other RBPs can be seen to have a relatively high score but fall short of the threshold of 0.5. Due to evolutionary divergence between *Leishmania mexicana* and most other organisms that these algorithms have been trained with, it is worth considering these as potentially containing a mitochondrial localisation signal.

For signal peptide detection only one protein was identified with high confidence amongst the 67 candidate RBPs: LmxM.36.1620. It is important to note that although the presence of the sequence is high confidence this does not alone suggest that the protein is secreted, as signal peptides can also target proteins to organelles involved in the secretory pathway and the lysosome(Armenteros *et al.*, 2019). As with the mitochondrial localisation predictions, several other proteins with scores narrowly below 0.5 may require caution when tagging at the N-terminus to avoid disrupting localisation, but the majority of RBPs were given low scores (>0.1).

3.9 Conclusions

A list of 67 proteins suitable as targets for an RBP bar-seq screen was compiled using the criteria discussed as guidelines in section 3.2. The background information presented in this chapter was used to inform this selection. A small proportion of the proteins in the XL-RBPome for *L. mexicana* were found to contain known RBDs, with many different families of RBPs identified. The relative conservation of candidate RBPs was assessed using BLASTp and RBPs were identified that contain kinetoplastid or even *Leishmania* specific regions. Bioinformatic analysis of proteins associated with mRNA (De Pablos *et al.*, 2019) but lacking RBDs was largely unsuccessful as a tool for predicting RNA-binding capacity in *Leishmania*. However, capability of predicting RNA-binding for well conserved domains suggests that re-training of this algorithm with RBPs from divergent species may improve the results in these parasites. The proteins from the XL-RBPome were analysed further to identify the lifecycle stage in which they were isolated at highest concentration. TargetP was used to detect any potential localisation signals in the predicted protein sequences for target RBPs and the Tryptag project was used to identify subcellular localisations for all orthologs in *T. brucei*. Of the proteins selected using this background information, 67 were taken forward and screened in Chapter 3.10. All 67 proteins are listed in Table 3.9.1.

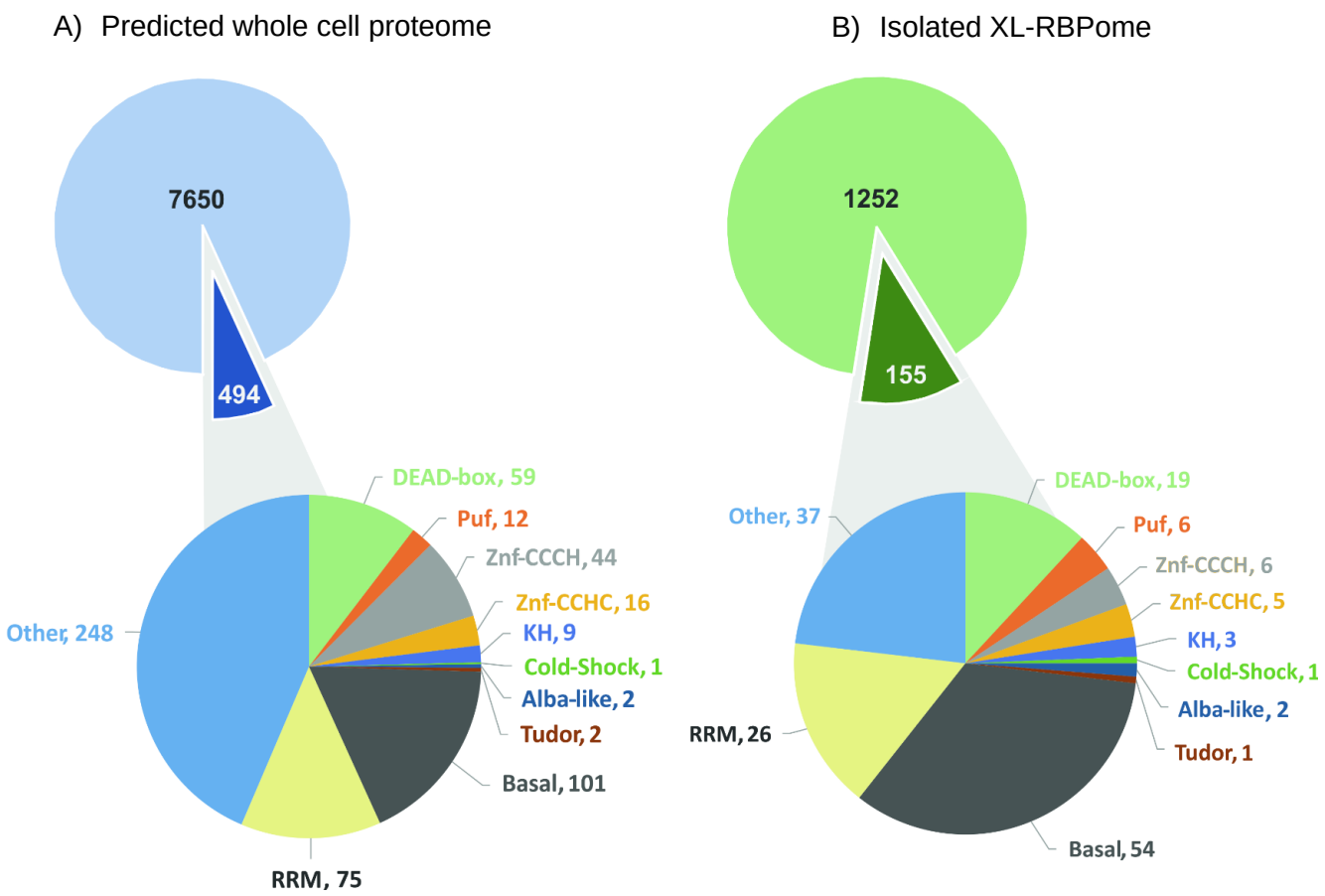


Figure 3.1. Proteins containing common RBDs in the *L. mexicana* genome and RBPome. The number of proteins identified with characterised RNA-binding domains (RBDs) in (A) the predicted *L. mexicana* proteome (Tritrypdb.org) and (B) the XL-RBPome (mRNA associated and UV-crosslinked). Characterised RBDs were detected in 494 out of 8144 genes in the total proteome and 155 out of 1407 proteins in the XL-RBPome (Pablos *et al.* 2019). Domain classes labelled as 'Other' represent proteins with additional, potential RBDs not labelled in this diagram (for the full list see supplementary Table 7.3). 'Basal' proteins have homology to the translational and splicing machinery. The RBPome RBD total is higher than the number of RBPs identified because of the presence of multiple domain classes in single proteins.

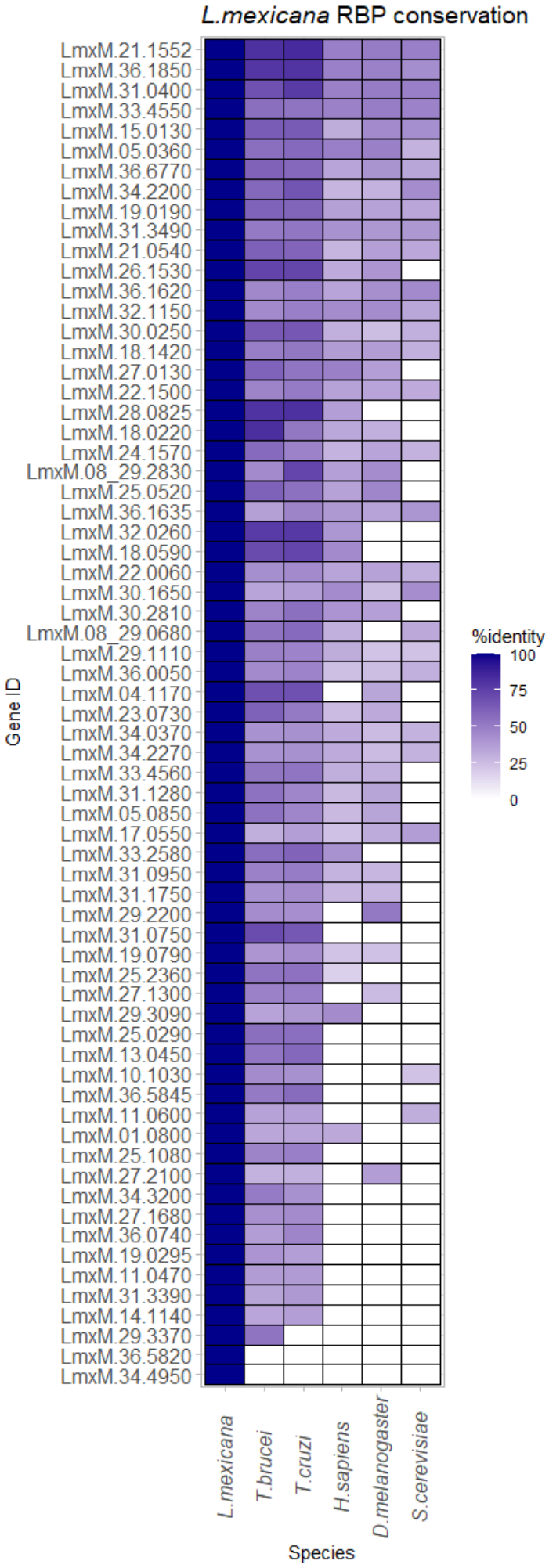


Figure 3.2. BlastP results to show *L. mexicana* RBP conservation. The amino acid sequences for all 67 RBPs included in the barcoded knockout screen were compiled for analysis of conservation. BLASTp was used to identify the top hit for each RBP in the genomes of: *L.mexicana*, *T.brucei*, *T.cruzi*, *H.sapiens*, *D.melanogaster* and *S.cerevisae*. The percentage identity comparing the top hit from each genome to the L.mexicana RBP sequences is presented as a heatmap. RBP gene IDs are arranged on the Y-axis from the most conserved to least conserved where the most conserved has the highest mean percentage identity across all six species. Higher degrees of conservation can be found in comparisons between the kinetoplastids alone.

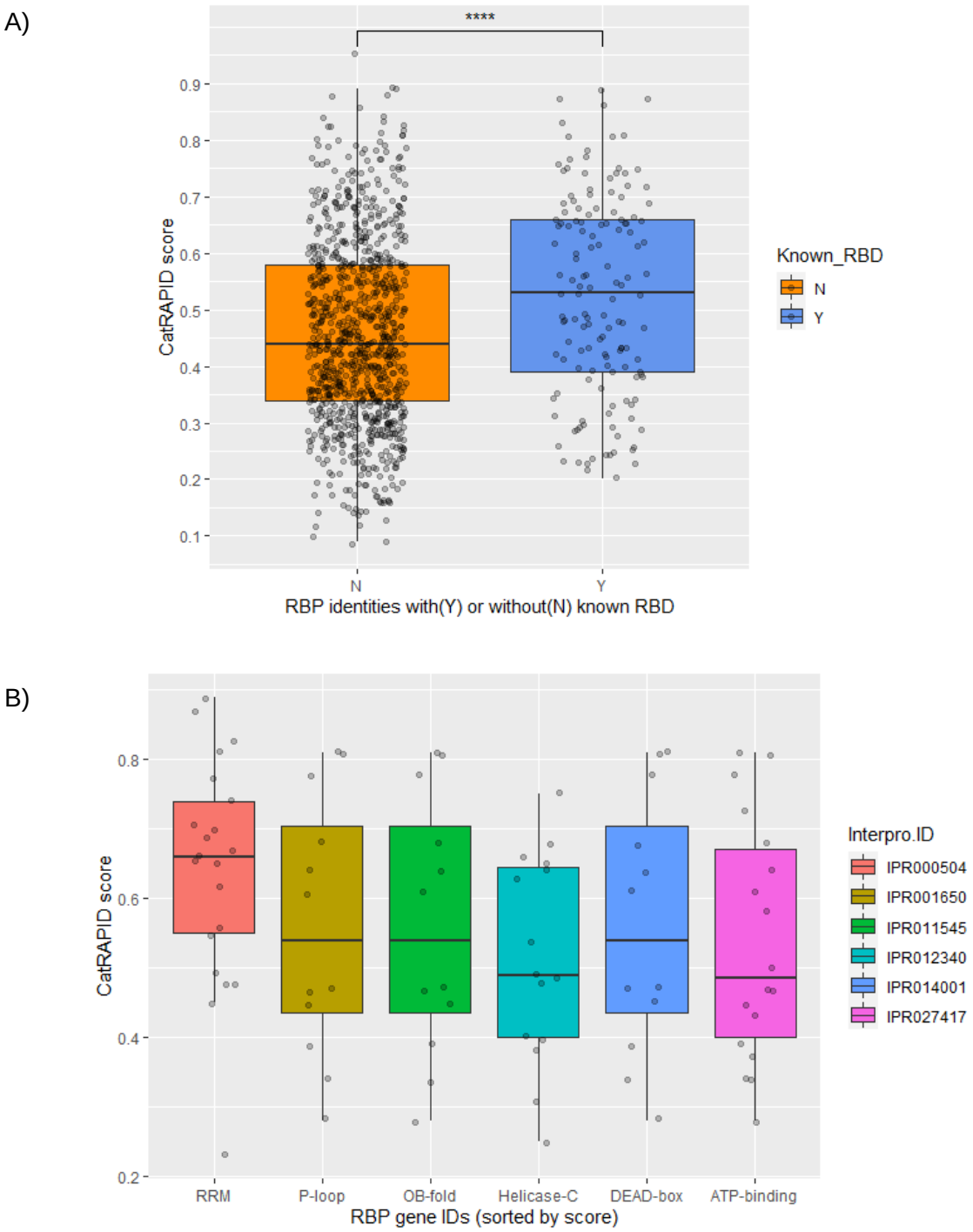
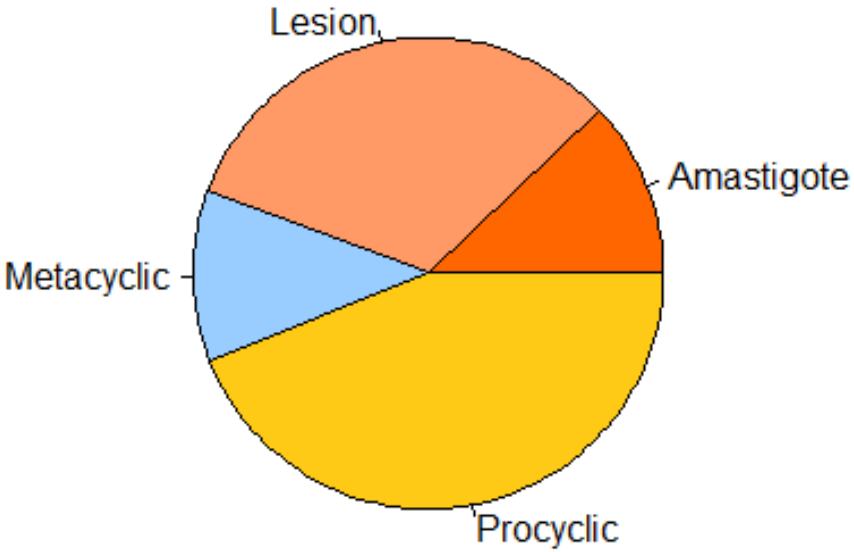


Figure 3.3. Predicting RNA-binding propensity in RBPs with no characterised RBD. A) All amino acid sequences for the XL-RBPome (mRNA-associated and UV-crosslinked) were compiled from TriTrypDB in FASTA format and processed using the CatRAPID signature server and default settings (Livi *et al*., 2015). CatRAPID scores were subset by RBPs containing (Y) or not containing (N) a common RBD from analysis in Figure 3.3.1. Boxplots with overlaid scatter plots show the trend around the mean as well as individual data points. The means of the two groups were found to be significantly different using a Welch's two sample t-test: t = -4. 32, df = 178, p = 2.64 x10^{-5}. B) When proteins containing individual domains were grouped and CatRAPID scores were compared, some RBDs were scored highly (RRM) whereas others such as DEAD-box domains were barely scored higher than a non-RNA-binding domain such as ATP-binding.

A) XL-RBPome



B) XL-RBPome: RBD-containing

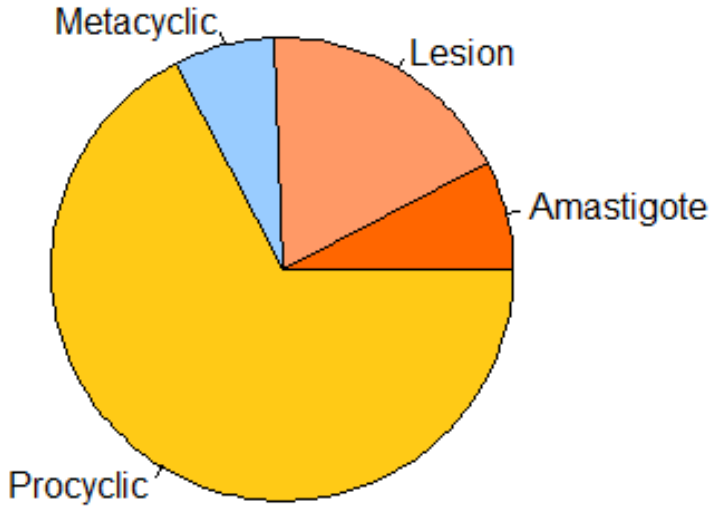


Figure 3.4. Relative abundance of RBPs in different *L.mexicana* lifecycle stages. A) The relative quantities of protein eluted from mRNA and analysed by quantitative mass spectrometry were compared between lifecycle stages. For each protein, the stage with the 'highest mean condition' was used to count the number of proteins most abundant on mRNA isolated from that stage. B) Proteins with no known RBD from analysis in Figure 3.3.1 were removed from the dataset. The reduced dataset was then graphed again for comparison. The majority of RBPs in *Leishmania mexicana* are most associated with mRNA in the non-infectious procyclic stage. Labels refer to procyclic promastigotes, metacyclic promastigotes, amastigotes from cultured macrophages and lesion derived amastigotes from a mouse model of infection, four months post infection (De Pablos *et al.* 2019).

Localisation N	Gene (Name)	Localisation C
Cytoplasm	LmxM.01.0800 (ZFC2H2)	Kinetoplast
Cytoplasm	LmxM.04.1170 (DRBD3/4)	
Nucleolus	LmxM.05.0360 (DDX)	Nucleolus
Basal body	LmxM.05.0850 (C2H2)	
Nucleolus	LmxM.08_29.0680 (TRRM3)	Nucleolus
Cytoplasm	LmxM.08_29.2830 (RBP6)	Cytoplasm / Nucleus
Nucleolus	LmxM.10.1030 (RBP11)	Nucleolus
Nucleolus	LmxM.11.0470 (PUF10)	Nucleolus
	LmxM.11.0600 (NOP47)	
	LmxM.13.0450 (Alba1)	
	LmxM.14.1140 (Cold-shock domain)	Kinetoplast
Nucleolus	LmxM.15.0130 (DDX27?)	Nucleolus
	LmxM.17.0550 (RBP23)	Cytoplasm
Nucleoplasm	LmxM.18.0220 (RBP29)	Endocytic / Cytoplasm
	LmxM.18.0590 (unknown RRM)	
Cytoplasm	LmxM.18.1420 (PUF2)	Cytoplasm
Cytoplasm	LmxM.19.0190 (DNA/RNA binding?)	
	LmxM.19.0295 (ZC3H40)	Cytoplasm
Cytoplasm	LmxM.19.0790 (SRP)	Cytoplasm
	LmxM.21.0540 (La-domain, Winged Helix)	Nuclear lumen
	LmxM.21.1552 (SUB2)	Nucleoplasm
	LmxM.22.0060 (TF-like protein)	
Cytoplasm	LmxM.22.1500 (DDX)	Cytoplasm
	LmxM.23.0730 (RBP10)	
	LmxM.24.1570 (DRBD13)	
Cytoplasm	LmxM.25.0290 (RBP43)	
Cytoplasm	LmxM.25.0520 (RBP3)	Flagellum
Cytoplasm	LmxM.25.1080 (RBP28)	Cytoplasm
	LmxM.25.2360 (Puf8)	Nucleolus
	LmxM.26.1530 (RBP8)	
Cytoplasm	LmxM.27.0130 (ZFP3)	Glycosome
	LmxM.27.1300 (ZC3H41)	
	LmxM.27.1680 (L-domain like)	
	LmxM.27.2100 (TRRM1)	Nucleoplasm
Cytoplasm	LmxM.28.0825 (RBP16)	
	LmxM.29.1110 (TRRM2)	Nucleoplasm

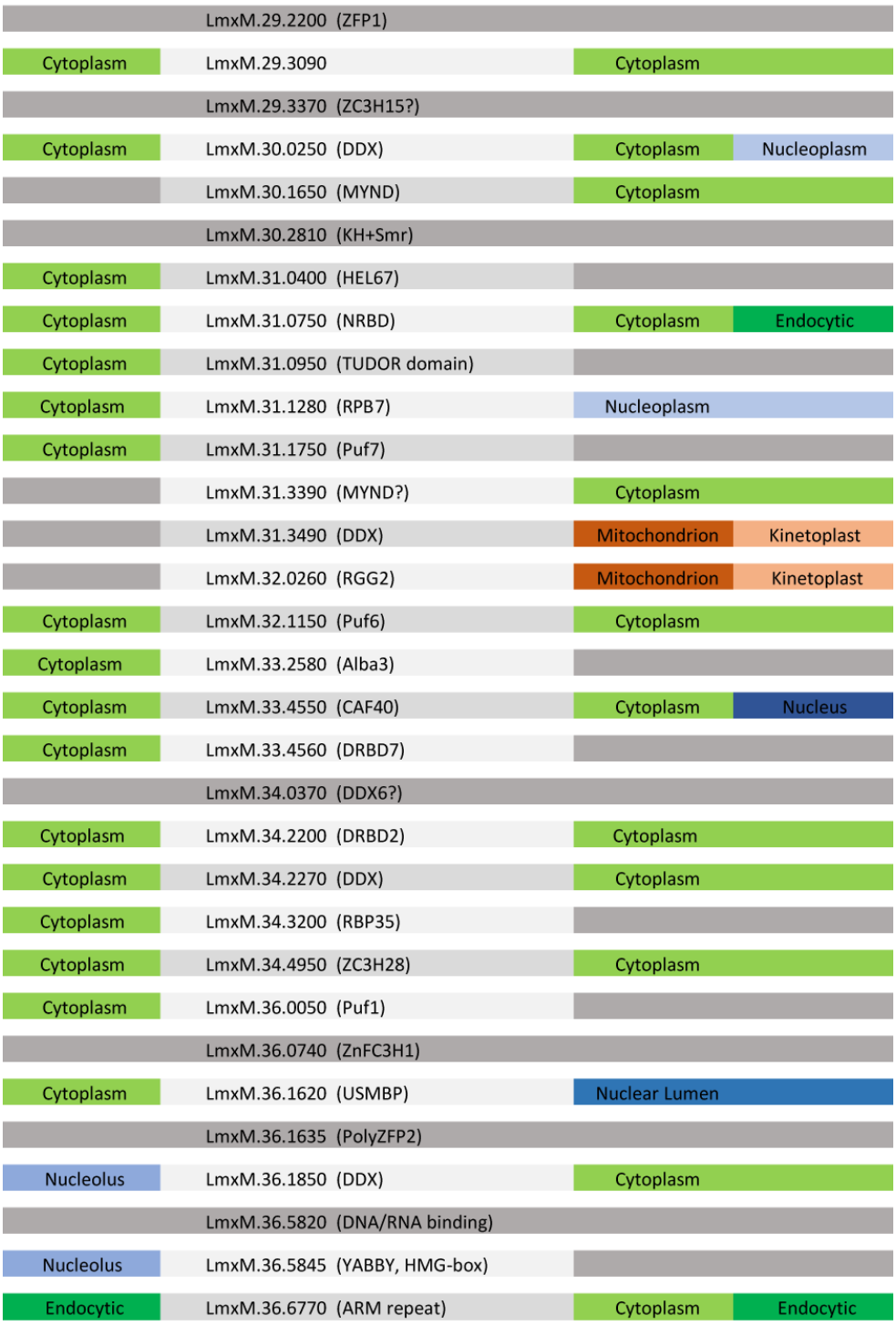


Figure 3.5. TrypTag localisation of RBP orthologs. The localisation of *T.brucei* orthologs of all 67 RBPs selected for the *L.mexicana* bar-seq screen are presented in a graphical table format for ease of use (Dean *et al.*, 2017). Details and sub categories have been simplified. Proteins for which tagging was not attempted or was unsuccessful are shown in grey. Proteins for which tagging was only attempted at one terminus or failed at one terminus are shown partially in grey.

A) Secretory signal prediction scores

B) Mitochondrial localisation scores

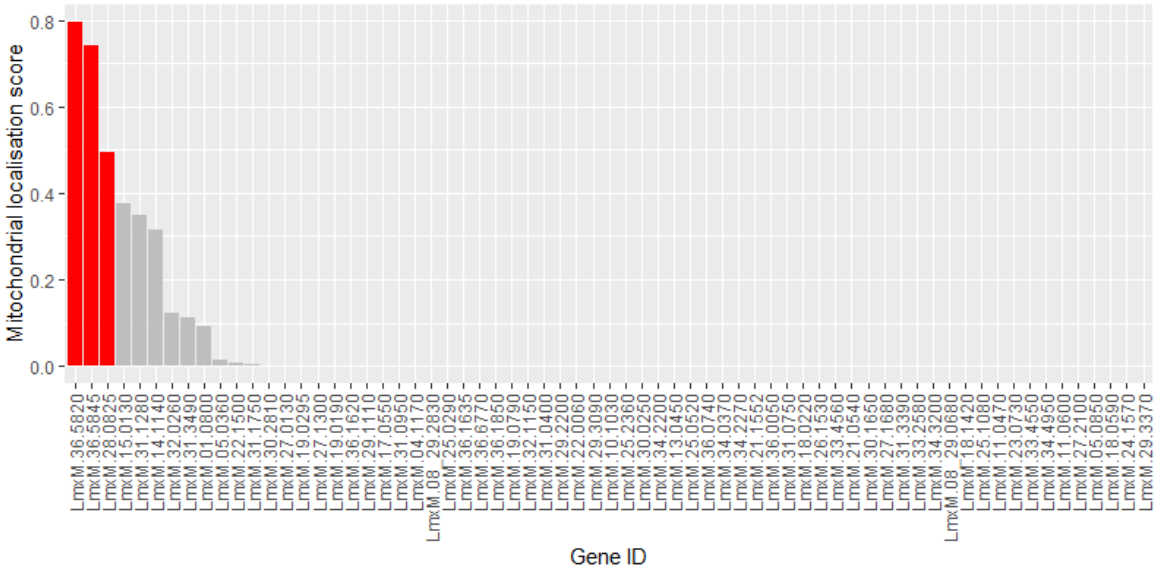


Figure 3.6. Predictions of targeting sequences in *Leishmania* RBPs. Predicted amino acid sequences for all 67 RBPs included in the knockout screen were processed using the TargetP2.0 server (Armenteros *et al.*, 2019) to detect signal sequencing suggesting secretory pathway involvement (A) or mitochondrial localisation (B). Protein identities with a positive score (>0.5) are highlighted in green for the secretory pathway and red for mitochondrial localisation. All proteins are listed from highest to lowest score on the X-axis.

Table 3.1.*Leishmania mexicana* RBPs selected for knockout. All RBPs selected for knockout are listed along with the lifecycle stage they were isolated in at highest abundance in the XL-RBPome (mRNA-associated and UV-crosslinked). Names for RBPs are provided where they have been used in previous studies in *Leishmania* or where the syntenic ortholog of a *Leishmania* RBP has been named in either *T.brucei* or *T.cruzi*. Interpro codes for predicted domains within each predicted protein sequence is provided in the final column.

Name	Gene ID	Highest mean condition:XL-RBPome	RBD
	LmxM.01.0800	Metacyclic	IPR013087(Zinc finger C2H2 type domain profile)
DRBD3/4	LmxM.04.1170	Procyclic	2x IPR000504 (RRM)
	LmxM.05.0360	Metacyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
	LmxM.05.0850	LD Amastigote	PF13913 (zf-C2HC)
TRRM3	LmxM.08_29.0680	Procyclic	IPR000504(RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
RBP6	LmxM.08_29.2830	Procyclic	IPR000504 (RRM)
RBP11	LmxM.10.1030	Procyclic	IPR000504 (RRM)
PUF10	LmxM.11.0470	Procyclic	IPR001313(Pumilio-family RNA binding repeat)
NOP47	LmxM.11.0600	Procyclic+LD Amastigote	3x IPR001878 (ZF_CCHC)
Alba1	LmxM.13.0450	Procyclic	IPR002775 (Alba)
	LmxM.14.1140	Procyclic	IPR002059 (Cold Shock)
	LmxM.15.0130	Metacyclic	IPR011545(Dead-box helicase)
RBP23	LmxM.17.0550	Procyclic	IPR000504 (RRM)
RBP29	LmxM.18.0220	Procyclic	IPR000504 (RRM)
	LmxM.18.0590	All stages	IPR000504 (RRM)
PUF2	LmxM.18.1420	Procyclic	IPR001313(Pumilio-family RNA binding repeat)
	LmxM.19.0190	Procyclic	IPR013729 (MBF1), IPR001387 (Helix-turn-helix)
ZC3H40	LmxM.19.0295	Procyclic	IPR000571 (zf-CCCH)
	LmxM.19.0790	Procyclic	IPR031545 (SRP_TPR_like), IPR013699 (SRP72 RNA-binding domain)
	LmxM.21.0540	Procyclic Amastigote + LD Amastigote	IPR006630 (La),IPR000504 (RRM)
SUB2	LmxM.21.1552	LD Amastigote	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
	LmxM.22.0060	Lesion	8X IPR000967 (zf-NF-X1)
	LmxM.22.1500	Procyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
RBP10	LmxM.23.0730	NA	IPR000504 (RRM)

DRBD13	LmxM.24.1570	Procyclic	2x IPR000504 (RRM)
RBP43	LmxM.25.0290	NA	IPR035979 (RBD)
RBP3	LmxM.25.0520	Procyclic	IPR000504 (RRM)
RBP28	LmxM.25.1080	Procyclic	IPR000504 (RRM)
Puf8	LmxM.25.2360	Procyclic	IPR033133 (PUM_HD)
RBP8	LmxM.26.1530	Procyclic	IPR000504 (RRM)
ZFP3	LmxM.27.0130	NA	IPR000571 (Zinc finger C3H1-type profile)
ZC3H41	LmxM.27.1300	Procyclic	IPR004088(KH domain)
	LmxM.27.1680	NA	
TRRM1	LmxM.27.2100	Procyclic	3x IPR000504 (RRM)
RBP16	LmxM.28.0825	Procyclic	IPR002059('Cold-shock' DNA-binding domain)
TRRM2	LmxM.29.1110	Procyclic	3x IPR000504 (RRM)
ZFP1	LmxM.29.2200	NA	IPR000571 (zf-CCCH)
	LmxM.29.3090	Procyclic	
	LmxM.29.3370	LD Amastigote	IPR000571 (Zinc finger C3H1-type profile)
	LmxM.30.0250	Amastigote	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
	LmxM.30.1650	NA	IPR002893 (MYND finger)
	LmxM.30.2810	Procyclic	IPR004088 (KH domain), IPR013899 (DUF), IPR002625 (Smr)
HEL67	LmxM.31.0400	Procyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
NRBD	LmxM.31.0750	Procyclic	IPR000504 (RRM)
	LmxM.31.0950	Procyclic	IPR002999 (TUDOR domain)
RPB7	LmxM.31.1280	Amastigote	IPR005576 (SHS2),
Puf7	LmxM.31.1750	Procyclic	9x IPR033133 (PUM_HD)
	LmxM.31.3390	NA	IPR002893 (MYND finger)
	LmxM.31.3490	Procyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
RGG2	LmxM.32.0260	Procyclic	IPR000504 (RRM)
Puf6	LmxM.32.1150	Procyclic	9x IPR033133 (PUM_HD)
Alba3	LmxM.33.2580	Procyclic	IPR002775 (Alba)
CAF40	LmxM.33.4550	LD Amastigote	2x IPR016024 (ARM repeat)
DRBD7	LmxM.33.4560	Procyclic	2x IPR000504 (RRM)

	LmxM.34.0370	Procyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
DRBD2	LmxM.34.2200	Procyclic	IPR000504(RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
	LmxM.34.2270	Procyclic	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
RBP35	LmxM.34.3200	Procyclic	IPR000504 (RRM)
ZC3H28	LmxM.34.4950	Amastigote + Procyclic	IPR000571 (Zinc finger C3H1-type profile)
Puf1	LmxM.36.0050	Procyclic	6x IPR001313(Pumilio-family RNA binding repeat)
	LmxM.36.0740	Procyclic	3x IPR000571 (Zinc finger C3H1-type profile)
USMBP	LmxM.36.1620	Procyclic	IPR001878(Zinc knuckle CCHC)
PolyZFP2	LmxM.36.1635	Procyclic	IPR001878(Zinc knuckle CCHC)
	LmxM.36.1850	Lesion	IPR011545 (DEAD/DEAH box helicase),IPR001650(Helicase conserved C-terminal domain), IPR014014(Q_MOTIF),IPR014001
	LmxM.36.5820	Metacyclic	NA
	LmxM.36.5845	Amastigote	2x IPR036910 (HMG box)
	LmxM.36.6770	Amastigote	IPR002553 (Adaptin), IPR016342 (APC, beta), IPR016024 (ARM repeat)

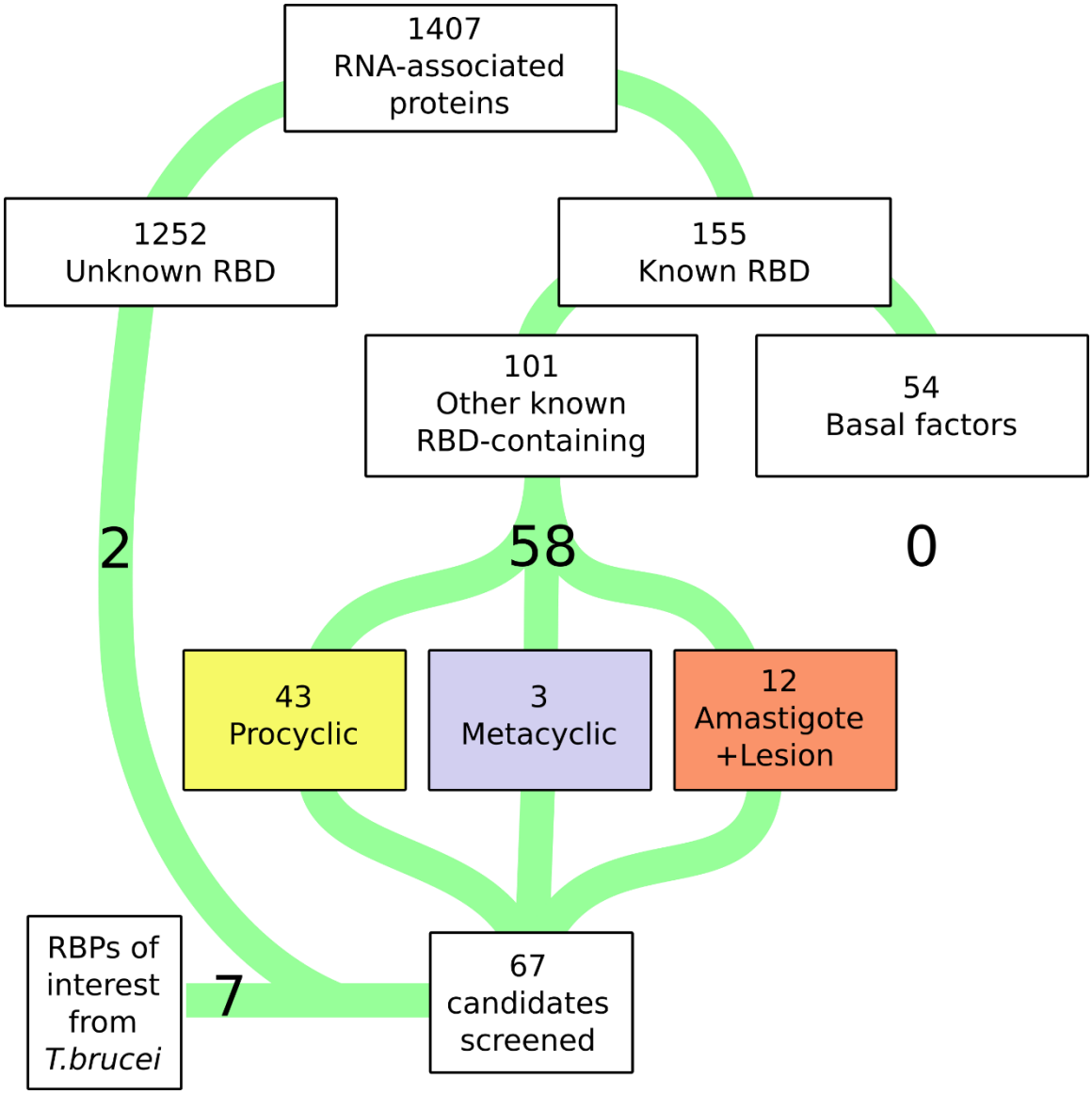


Figure 3.7. Flow-chart summary of candidate selection process. The 1407 mRNA associated proteins from the *L.mexicana* XL-RBPome (De Pablos et al. 2019) were searched for those with well characterised RBDs. Of those containing a known RBD, factors more likely to be conserved basal machinery components were identified and removed. Remaining candidates were sorted by highest protein abundance isolated from each lifecycle stage. RBPs that were most abundant in the human infectious stages, with known RBDs were the highest priority, followed by other candidates with known RBDs. Two factors that were highly stage specific and most abundant in the human infectious stages were included despite having no clear RBD (distant homology to nucleic acid binding proteins). Seven factors were included due to information from previous studies in T.brucei but weren't isolated in the *L.mexicana* XL-RBPome.

3.10 Chapter 3 discussion

3.10.1 Analysis of the RBPome results

The size and scope of the *L. mexicana* RBPome (De Pablos *et al.*, 2019) means that selecting a feasible number of genes for a CRISPR bar-seq screen was a large task that could be approached from many angles. Undoubtedly, the ability to use CRISPR at high throughput, through the large-scale production and transfection of plasmid or small guide RNA libraries would greatly increase the threshold for the number of genes screened. The use of genome wide CRISPR screens is now widespread in many organisms (Koike-Yusa *et al.*, 2014; Shalem *et al.*, 2014; Bassett, Kong and Liu, 2015; Chen *et al.*, 2015; Parnas *et al.*, 2015; Sidik *et al.*, 2016; Joung *et al.*, 2017). In humans, the latest CRISPR screening protocols, as well as other high throughput methods, have been used to characterise the post-transcriptional regulome in great detail (Van Nostrand, Freese, *et al.*, 2020; Wheeler *et al.*, 2020). Undoubtedly, this is the direction that will be taken for knockout screens of any family of *Leishmania* proteins in the future being a highly efficient way of identifying large numbers of phenotype driving genes and linking them to specific functions or pathways in the cell. The availability of the CRISPR system in its current form is hugely valuable in an organism that lacks a functional RNAi pathway, the other main tool of choice for these types of screen (Wurst *et al.*, 2009; Moss *et al.*, 2015; Rico *et al.*, 2017). With the system designed by Beneke *et al.*, 2017 and further optimized in the Gluenz and Mottram labs, a large number of proteins can be both deleted or tagged using the same sets of primers. Combined with DNA barcoding, the existing methods have produced several successful bar-seq screens in kinetoplastid parasites (Beneke *et al.*, 2019; Damianou *et al.*, 2020; Baker *et al.*, 2021). Based on the scale of these studies, a selection of around 100 genes was chosen as a reasonable target for this study.

The criteria used to select RBPs for the bar-seq screen depend entirely on the question being answered. In this case the focus of the screen was placed on two key areas. Firstly, identifying factors that cause fitness phenotypes in the infectious metacyclic promastigotes. Identification of factors that regulate differentiation of non-infectious to infectious parasites could shed light on proteins and pathways necessary for transmission and establishment of infection. The second focus was to assess the phenotype of RBP null mutant cell lines, both in mammalian macrophages and in a full mouse model of infection. Establishing links between specific regulatory factors and potential; infectivity phenotypes lays the foundations for virulence studies, contributing to our understanding of disease progression and providing new targets for intervention. In *Leishmania*, RNA-binding proteins have been relatively poorly studied in these contexts. Consequently, many of the mRNA bound proteins isolated in the XL-RBPome are valid targets for further investigation.

Identification of RBPs that contain characterised RBDs was attempted first to provide further information for these proteins. It also helped to determine what proportion of *Leishmania* RBPs isolated in the XL-RBPome screen contained known RBDs, from which hypotheses about their function can be made and what proportion lacked any known RBD, making them more challenging

to study. There are several indicators that can be used to identify an RBD within predicted protein sequences. One of the most commonly used annotations are Gene Ontology-Terms (GO-Terms)(Ashburner *et al.*, 2000; Carbon *et al.*, 2021). Gene Ontology annotations are a series of terms describing the biological function or predicted biological function of a given gene. GO-Terms are assigned based on evidence that can come in many forms and can even be curated manually. However, when looking at putative *L. mexicana* proteins, the high degree of evolutionary separation between these parasites and eukaryotic model organisms means that many are not annotated with even a single term. Whilst GO-Terms provide information on the function of a protein, for example the term'RNA-binding', they are less helpful for classifying sub families of RBPs based on known domains. One of the pieces of data used to assign GO-terms is the presence of a PFAM domain (Sonnhammer, Eddy and Durbin, 1997; Sonnhammer *et al.*, 1998; Mistry *et al.*, 2021). PFAM annotations are based on multiple sequences alignments and Hidden Markov models (HMMs) that describe different protein domain families. Novel protein sequences were compared to the database of known domains and assigned protein domain annotations based on sequence similarity. PFAM domains were found more commonly as annotations on more *L. mexicana* RBPs than GO-terms, providing a more detailed picture of the types of known RBD that are found. The caveat to this is that the assignment of PFAM domains based on protein sequence doesn't have experimental evidence for RNA-binding to back it up so must be treated as a prediction in proteins without this evidence.

Ultimately, InterPro codes were chosen as a search method because they were often the only annotation given to an uncharacterised putative RBP and they are based on collated information from many different sources, including the GO-term and PFAM databases, to assign information about protein domains in novel sequences (Apweiler *et al.*, 2001; Jones *et al.*, 2014; Finn *et al.*, 2017; Blum *et al.*, 2021). InterPro annotations also reference the number of repeats of a given domain, information that can be valuable when studying RNA-protein interactions. To search for known RBDs by Interpro domain, a list of well characterised RBDs had to be compiled. The aim of this was to search for relevant annotations in as many of the XL-RBPome identities as possible, casting the net wide to include domains that may or may not be RNA-binding depending on the specific protein. Due to the likely high level of false positives and negatives these results should not be interpreted as a precise prediction of the number of RNA-binding proteins in *Leishmania mexicana* but as a means of annotating the proteins that have been isolated from mRNA pulldowns and aiding the selection of RBPs for the bar-seq screen.

InterPro codes for the most studied and frequently occuring RBDs were compiled from studies in model organisms and in humans (Castello *et al.*, 2012; Mitchell *et al.*, 2013; Sysoev *et al.*, 2016; Wessels *et al.*, 2016; Nostrand *et al.*, 2017; Hentze *et al.*, 2018). Studies in *Trypanosoma brucei* also informed the choice of RBD codes (Kramer and Carrington, 2011; Erben *et al.*, 2014; Lueong *et al.*, 2016). The ATtRACT database lists characterised RNA-binding domains and motifs, giving each a unique code (Giudice *et al.*, 2016). This database was also used to identify the commonest

characterised RNA-binding domains which were included in the search. Lastly, the InterPro database itself was searched with the terms "RNA-binding" and "Nucleic acid binding" to return more unusual or specific protein domains of interest. In other eukaryotes, searching for "Nucleic acid binding" domains would return a lot of transcription factors but due to the lack of transcriptional gene regulation in *Leishmania* this becomes worthwhile. The domain codes resulting from these InterPro searches were curated manually and included if the domain was associated with RNA-binding function. This part of the methodology could be dramatically improved using data mining and keyword searching algorithms as demonstrated by Castello *et al.*, (2017). Many of the domain codes found were highly specific to individual proteins and unlikely to identify similar proteins when compared to the *Leishmania mexicana* RBPome but this strategy highlighted some RBPs containing less-studied domains such as TUDOR, KH and Cold-shock that might otherwise have been overlooked.

The results of searching both the whole *L. mexicana* genome and the XL-RBPome for RNA-binding InterPro domains highlight both the diversity of RNA-binding proteins and the large proportion of proteins in these parasites with no homology to characterised domains (De Pablos *et al.*, 2019). The decrease in proteins containing "other" RNA-binding domains when comparing the genome to the XL-RBPome, is likely due to removal of proteins containing nucleic acid binding domains but not bound to mRNA in significant quantities during immunoprecipitation. There are also many more "basal" proteins, those involved in the core machinery of splicing, transcription or translation, in the XL-RBPome than in the genome. The simplest explanation for this is that whilst a genome contains all the potential genes, a proteome contains a subset of these proteins subject to many biases. In the case of the XL-RBPome, during an mRNA pulldown and subsequent mass-spectrometry, the abundance of an RNA-associating protein will introduce a positive bias towards detecting that protein. Due to the extreme abundance of the basal machinery proteins (ribosomal proteins, core elongation factors, t-RNA synthetases etc.) compared to most RNA-binding proteins, it is not surprising that a large proportion of these proteins have been isolated. These proteins are also often constitutively expressed. Many proteins involved in splicing, transcription or translation also bind a large number of targets compared to many regulatory RBPs which are much more selective; this further increases the probability of isolating them during mRNA pulldown.

The common families of RBPs that have been studied elsewhere are well represented in the *L. mexicana* XL-RBPome. Excluding the "other" category for reasons discussed previously, comparing the total known RBD-containing proteins in the genome to those in the XL-RBPome shows that around 50% of the total predicted RBP gene products were isolated experimentally (Figure 3.1). This is most likely because the different timepoints taken are only snapshots of the proteins associated with RNA at any one time. Regulators of specific proteins may only be expressed or may only bind mRNA during specific conditions or lifecycle stages that do not necessarily coincide with the four time points described in De Pablos et al. (2019). Overall, the identification of RBDs using InterPro codes provided more information about the likely structure and function of RBPs in the XL-RBPome

but further bioinformatics would be required to make accurate predictions about the absolute number of RBPs containing specific RBDs in *L. mexicana.* For example, hidden Markov models or tools such as PSI-BLAST may reveal less conventional proteins with structural similarity (Oyama *et al.*, 2008; Li *et al.*, 2012). The total number of RBPs predicted to contain certain RBDs in this study may be an underestimate. For example, an in depth study on the number of CCCH Zinc finger proteins in *Leishmania* concluded that there are 54, 16 of which are specific to *Leishmania* itself, compared to the 44 detected in this study (Kramer, Kimblin and Carrington, 2010). In this case, the disparity in predicted protein numbers is mostly due to their further investigation of non-conventional CCCH motif containing proteins that are not recognized by InterPro.

3.10.2 Divergence in RBP amino acid sequences in kinetoplastids and other organisms.

The relative conservation of RBPs in *Leishmania* was also investigated to inform the selection process. Despite the degree of evolutionary separation between *Leishmania* and other eukaryotes where RBPs have been studied extensively, many proteins involved in core cellular processes have a surprisingly high sequence homology (Beckmann *et al.*, 2015; Matia-González, Laing and Gerber, 2015). As described earlier, when studying *Leishmania,* it is important to investigate the pathways and proteins that are involved with the infectious stages (metacyclic promastigotes and amastigotes) or differentiation to them. Highly conserved, core components of the eukaryotic cellular machinery are less likely to be involved in *Leishmania* specific processes.

It was hypothesised that *trans*-regulators of *Leishmania*-specific processes are more likely to have divergent protein sequences compared to regulatory proteins studied in other eukaryotes. The BLASTp server was used as a straightforward method of comparing *Leishmania* RBPs from the XL-RBPome to whole proteome of several species: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Saccharomyces cerevisiae, Drosophila melanogaster* and *Homo sapiens* (Figure 3.2). RBPs have been well studied in both trypanosome species selected and it is likely that many proteins have orthologs in all three kinetoplastids investigated. The results generally agreed with this hypothesis but despite this, the majority of *T. brucei* top BLAST results (smallest E-value) across all RBPs tested were below 50% identity to the *L. mexicana* sequence. This illustrates how it cannot be taken for granted that proteins from *Leishmania* will have high sequence similarity to other kinetoplastid orthologs. The three model organisms used for comparison, *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*, are all organisms where RBPs have been studied extensively (Scherrer *et al.*, 2010; Tsvetanova *et al.*, 2010; Gerstberger, Hafner and Tuschl, 2014; Beckmann *et al.*, 2015; Matia-González, Laing and Gerber, 2015; Ghosh and Sowdhamini, 2016; Sysoev *et al.*, 2016; Wessels *et al.*, 2016; Hentze *et al.*, 2018; Van Nostrand, Freese, *et al.*, 2020). The results of this analysis identified proteins in kinetoplastids that have low sequence identity to other RNA-binding proteins, despite often containing a known RNA-binding domain. This highlights an important point; just because a protein has a known RBD does not mean it is a direct ortholog of a well characterised protein. Other protein domains in an RBD-containing protein contribute to the

context-specific function and can vary considerably between different cell types and organisms. In fact, some candidate *Leishmania* RBPs had novel domain combinations that have not yet been fully characterised (eg. KH domain and SMR domain in LmxM.30.2810). The results of this analysis could be improved in several ways. Running the BLASTp search manually from the desktop allows for much more control over the search parameters and would make the process more streamlined. Additionally, variations on the basic BLASTp algorithm, such as PSI-BLAST, are optimised to detect homology between more distantly related proteins (Oyama *et al.*, 2008; Li *et al.*, 2012). An investigation using improved parameters may better identify distant orthologs of *Leishmania* RBPs. Further analysis could include the comparison of RBDs specifically, giving a better measure of evolutionary divergence of the RNA-binding mechanisms across trypanosomatids and other eukaryotes. However, major divergence between RBDs characterised in higher eukaryotes and those in trypanosomatids can make identification of the domains a challenging task.

3.10.3 *In silico* prediction of RNA-binding

One clear finding from the bioinformatic screen for previously characterised RBDs within the XL-RBPome was that a large proportion of the *L. mexicana* proteins isolated on mRNA contained no known domains at all. Although these proteins are intriguing and valid candidates for further study, the lack of any known RNA-binding mechanisms make them very difficult to study or even select for a knockout screen. With over 1000 proteins that apparently interact with *L. mexicana* mRNA but contain no known RBD, it is hard to find traits to favour selection of any one protein over another. It is likely that a proportion of these proteins do not bind RNA directly, but are closely associated and strongly bound to proteins that are. In an attempt to find more information about the unknown-RBD containing proteins that make up the majority of the XL-RBPome, bioinformatic predictions of RNA-binding were explored. In general, *in silico* predictions of RNA-binding are not very powerful as our general understanding of binding mechanisms and RNA-protein interaction is still in its infancy. In the last decade, several paradigm shifts in this field have demonstrated the true diversity of RNA-binding proteins and the ability for proteins with unrelated functions to have secondary RNA-binding properties (Scherrer *et al.*, 2010; Gerstberger, Hafner and Tuschl, 2014; Beckmann *et al.*, 2015; Ghosh and Sowdhamini, 2016; Hentze *et al.*, 2018; Van Nostrand, Freese, *et al.*, 2020). Various methods exist to predict RNA binding, most being based on the characteristic structures or sequences of known RNA-protein interacting domains. Large scale bioinformatic studies have been carried out to detect novel RBP encoding genes in the human genome using Hidden Markov Models (HMMs) to classify structural families of experimentally proven RNA-interacting domains (Ghosh and Sowdhamini, 2016). This new database was subsequently used to search the human proteome for novel RBPs. However, this technique does require the annotations of known RBDs to make predictions about domains in novel proteins and is not suitable for *de novo* RBD detection (Marchese *et al.*, 2016).

Several techniques use structural and biochemical properties as predictors of RNA-binding activity. Techniques such as BindN and BindN^{+} use pKa, hydrophobicity and molecular mass to predict residue binding probabilities, the latter integrating PSI-BLAST searches to provide evolutionary context to the results (Wang and Brown, 2006; Wang *et al.*, 2010). However, these and other similar methods do not provide information about binding domains. Instead, they work at the scale of individual amino acids making them unsuitable for the purposes of this study. *Cat*RAPID signature, was selected as the most suitable algorithm to discover RBDs in the uncharacterised XL-RBPome identities (Livi *et al.*, 2015). *Cat*RAPID relies on biochemical and structural characteristics such as hydrophobicity, steric hindrance, and predicted secondary structures so does not require domains to be pre-annotated. But, unlike many other predictive algorithms, by using position sensitive 'signatures' it can predict which region of the amino acid sequence is likely to bind RNA. *Cat*RAPID signature also performed well in tests against other predictive algorithms for tasks similar to those in the current study (Cirillo, Agostini and Tartaglia, 2013). After processing the *L. mexicana* XL-RBPome using the *cat*RAPID signature server, the results were analysed to test several hypotheses. Most importantly, the algorithm did not assign a mean or median score above 0.5 for this dataset suggesting it could not detect RNA-binding domains in many of these identities. One hypothesis to explain this is that being trained on known RBDs from evolutionarily distant organisms, the algorithm is not good at detecting more the divergent RBDs in kinetoplastids. The fact that proteins containing known RBDs performed significantly better than those without supports this. Even so, the fact that many known RBD containing proteins were given low scores (<0.5) means that *cat*RAPID signature in its current form was not reliable enough to identify *Leishmania* RBPs in general. It is possible that *cat*RAPID signature or similar methods will be useful in the future if trained with more kinetoplastid experimental data on RNA-binding. The high scores given to proteins containing certain types of RNA-binding domains (e.g. RRMs) show the potential for detecting novel binding proteins if the method was expanded upon. In recent years, using advanced artificial intelligence and neural networks to accurately predict protein structures from amino-acid sequences has become more powerful and is gaining widespread recognition (Senior *et al.*, 2020; Jumper *et al.*, 2021). Currently, folding predictions made with tools such as AlphaFold remain quite inaccurate for many *Leishmania* proteins. This is also likely due to the evolutionary divergence between *Leishmania* proteins and those that the neural networks have been exposed too from higher eukaryotic model organisms and humans. Exposing these tools to more trypanosomatid proteins has already yielded improved results for protein folding prediction in these parasites and it seems likely that they will become valuable for predicting novel domains that can interact with RNA in the near future (Wheeler, 2021).

3.10.4 Stage specificity of *L. mexicana* RBPs

The huge quantity of proteomic data gathered from the XL-RBPome, and other RNA-bound proteomes like it, can be studied from several angles to improve our general understanding of protein-RNA interactions in *Leishmania*. Because quantitative mass spectrometry was performed in triplicate

on the four distinct *L. mexicana* lifecycle stages (procyclic promastigotes, metacyclic promastigotes, macrophage-derived amastigotes and lesion-derived amastigotes) simultaneously, the relative quantities of each protein isolated in the different stages can be compared. In the case of an mRNA associated proteome, the quantity of each peptide eluted can be affected by the expression level of that protein as well as its binding affinity for RNA in a given stage. However, in the XL-RBPome the use of UV-crosslinking minimises the effects of binding affinity by securing proteins with weaker binding affinities as long as they are in close proximity to the RNA. The effects of UV-crosslinking can be seen when comparing the XL- and NonXL-RBPome, which are nearly identical in terms of protein identities but have differences in relative protein abundance. This is because for the NonXL-RBPome, the strength of protein-RNA interactions determines the amount of protein bound to and eluted from the mRNA. Conversely, all UV-crosslinks are strong covalent bonds but can only act over a short distance so the factor governing XL-RBPome pulldowns is likely to be the number of protein-RNA contacts rather than just their strength. Whilst CCCH zinc-finger proteins were enriched in the XL-RBPome compared to the non-XL-RBPome, DEAD-box helicases are known to bind RNA strongly and a large quantity of these were pulled down in both the crosslinked and non-crosslinked RBPome (Bono *et al.*, 2006; Schütz *et al.*, 2010; Marchat *et al.*, 2015; De Pablos *et al.*, 2019).

The XL-RBPome was sorted by 'highest mean condition', the stage in which protein quantity was highest, and the number of proteins in each of the four stages was counted. It has been shown previously that many cellular processes are upregulated during the highly metabolically active, rapidly dividing procyclic promastigotes. Factors belonging to core processes such as transcription, translation, splicing and DNA-replication appear to be upregulated on both the RNA and protein levels (El-Sayed *et al.*, 2005; Walker *et al.*, 2006; Saxena *et al.*, 2007; Dillon *et al.*, 2015; Inbar *et al.*, 2017; De Pablos *et al.*, 2019; Vigneron *et al.*, 2020). This appears to be the case in the XL-RBPome (Figure 3.1). The lack of proteins most associated with mRNA in the metacyclic promastigote stage also agrees with previous work, both in *Leishmania* and *T. brucei*, showing metacyclic promastigotes to be relatively quiescent in comparison to the much more translationally active procyclic stage (Shapiro *et al.*, 1984; El Fakhry, Ouellette and Papadopoulou, 2002; Bentel *et al.*, 2003; Zinoviev and Shapira, 2012; Inbar *et al.*, 2017; Vigneron *et al.*, 2020). RBPs that were most enriched in the metacyclic promastigotes as well as those most enriched in the amastigotes were prioritised as knockout candidates as they are more likely to be involved in processes that are crucial for infectivity, metacyclogenesis, amastigogenesis or survival in the mammalian host. Despite this, some of these proteins were not included in the bar-seq screen as priority for other reasons, for example the lack of a known RNA-binding domain or evidence that they were highly conserved from *Leishmania* to humans. It is worth noting that proteins may be expressed in multiple stages and only function as an RNA-binding protein in certain stages. Also, many RNA-binding proteins act as translational repressors or to destabilise mRNA (e.g. PUF proteins (Hoek, Zanders and Cross, 2002; Luu *et al.*, 2006; Droll *et al.*, 2010; Folgueira, Martínez-Bonet and Requena, 2010; Müller, Matuschewski and Silvie, 2011; Schumann Burkard *et al.*, 2013; Jha *et al.*, 2014; Azizi, Dumas and

Papadopoulou, 2017)), in which case expression in the procyclic promastigotes may indicate regulation of genes that are crucial for the infectious stages alone. Because of this, RBPs isolated predominantly from the procyclic stage were included in the screen if they were of interest for other reasons.

3.10.5 Predicting RBP localisation

Signalling sequences were detected in in the predicted protein sequences for all 67 knockout candidate RBPs. The TargetP2.0 server that was used integrates several methods, including the widely used SignalP, to provide an overview of targeting peptides for the provided amino acid sequence (Nielsen and Engelbrecht, 1997; Almagro Armenteros *et al.*, 2019; Armenteros *et al.*, 2019). The basic method uses neural networks to detect both cleavage sites and differentiate between signal peptides and non-signal peptides. The accuracy and scope of these tools has improved dramatically with the integration of modern deep learning and artificial intelligence. The server was tested against many eukaryotic genomes and the proteins used to train the neural networks were also from a diverse range of eukaryotes. However, no kinetoplastids were involved at these early stages so it is likely that kinetoplastid specific signalling sequences may not be well detected.

Considering the majority of the 67 proteins tested had orthologs with a cytoplasmic distribution on *T. brucei*, it is not surprising that the majority do not appear to contain detectable targeting peptides (Figure 3.5). Mitochondrial transit peptides (mTPs), targeting proteins to the mitochondrial matrix, were detected with a positive score (>0.5) in three proteins from the *Leishmania* RBP knockout screen. As well as the three positive scored proteins, a further six had a score above 0.1. Because the TargetP2.0 server is unfamiliar with kinetoplastid proteins it may be that these proteins contain divergent mitochondrial transit peptides. Of the three RBPs with a high mTP score, one has been experimentally validated. Subsequent investigation of LmxM.28.0825 (RBP16) revealed mitochondrial localisation when tagged at the C-terminus with a triple HA epitope tag (Field *et al.*, 1988). This is in contrast to N-terminal tagging which showed a cytoplasmic distribution in both *L. mexicana* (Ferreira *et al.*, 2020) and in *T. brucei* (Figure 3.5)(Dean, Sunter and Wheeler, 2017). This result indicates that the TargetP high confidence targets have real biological relevance and also a reminder that tagging the N-terminus of a protein with a localisation signal, can have drastic effects on its subcellular localisation and consequent functions.

The presence of signal peptides was also reported by the TargetP server. Signal peptides target proteins to the endoplasmic reticulum where they can enter the secretory pathway. However, proteins with signal peptides can be involved in other areas of the secretory pathway without being secreted themselves; for example the lysosome (Armenteros *et al.*, 2019). Interestingly, the only protein to be given a positive signal peptide score was the universal minicircle binding protein (UMSBP) which has been characterised in *Leishmania* as a mitochondrial protein (Singh *et al.*, 2016). While UMSBP has been characterised as a single stranded DNA-binding protein, its association with mRNA in the XL-RBPome and the ability of many zinc-finger proteins to bind both DNA and RNA led to its

inclusion in this study (Hall, 2005). It is possible that while TargetP2.0 has identified the localisation signal, the evolutionary divergence has caused it to be misclassified. Alternatively, this protein may have functions in the secretory pathway that are yet to be investigated.

3.10.6 Inclusion of RBPs absent from the *L. mexicana* RBPome

In addition to proteins that were chosen for inclusion in the bar-seq screen based on the criteria described in 3.2 and the results of Chapter 3, several proteins were chosen that were not isolated in the XL-RBPome. As with any proteome, the proteins detected in each lifecycle stage represent only a snapshot of expression at one specific timepoint. As a result of this there are a small number of RBPs that were chosen for knockout but are not found in the XL-RBPome (LmxM.23.0730, LmxM.25.0290, LmxM.27.0130, LmxM.29.2200, LmxM.30.1650, LmxM.31.3390). Most of these are proteins that have been described in *T. brucei* as bound to RNA, many of which were isolated in the tethering screen produced by Erben et al. (2014). Some such as RBP10 (LmxM.23.0730) have already been linked to differentiation in *T. brucei* (De Pablos *et al.*, 2017; Mugo and Clayton, 2017). RBP10 acts as a regulator of differentiation in trypanosomes, stalling differentiation from procyclic to bloodstream forms when depleted and inducing it when expressed. RBP10 can be considered a master regulator of differentiation, affecting many different pathways from surface protein production, kinase signalling cascades and even regulating stage specific RBPs. RNA-binding protein 43 (RBP43: LmxM.25.0290) has been isolated several times in *T. brucei* and is investigated in *L. mexicana* in experiments in Chapter 4.10 and is discussed in 6.4.2.

The CCCH zinc-finger protein 3 (ZFP3) has been investigated extensively in *T. brucei* but not in *Leishmania.* In trypanosomes, ZFP3 interacts with both the translational machinery and with other *trans*-regulators of differentiation such as ZFP1 and ZFP2 (Paterou *et al.*, 2009). ZFP3 was found to be constitutively expressed but associated with the polysomes in a stage specific manner. Specifically, ZFP3 interacts with GPEET and EP1 procyclin transcripts via a cis regulatory element where elevated ZFP3 levels increase EP1 transcript levels but reduce GPEET (Walrad *et al.*, 2009). Transcripts upregulated in the infectious, stumpy form are regulated by ZFP3 via interaction at the 3’ UTR (Walrad *et al.*, 2012). Like ZFP3, ZFP1 (LmxM.29.22000) is a small CCCH zinc finger protein implicated in the control of differentiation (Hendriks *et al.*, 2001). ZFP1 expression was upregulated during the differentiation from bloodstream to procyclic forms in *T. brucei.* Ablation of ZFP1 disrupts the repositioning of the kinetoplast, a key step in the differentiation process (Hendriks and Matthews, 2005).The last two proteins in the screen that were not isolated in the XL-RBPome (LmxM.30.1650 and LmxM.31.3390) both contain MYND domains. While MYND domains themselves have been characterised as domains facilitating protein-protein interactions, these *Leishmania* MYND proteins are orthologs of proteins isolated from mRNA pulldowns in *T. brucei* (Casas-Sanchez *et al.* in preparation).

All RBPs of interest were compiled in a list of over 100 proteins that were further reduced to 67 that were included in the bar-seq screen (Table 3.1

Table 3.1.*Leishmania mexicana* RBPs selected for knockout. All RBPs selected for knockout are listed along with the lifecycle stage they were isolated in at highest abundance in the XL-RBPome (mRNA-associated and UV-crosslinked). Names for RBPs are provided where they have been used in previous studies in *Leishmania* or where the syntenic ortholog of a *Leishmania* RBP has been named in either *T.brucei* or *T.cruzi.* Interpro codes for predicted domains within each predicted protein sequence is provided in the final column.

). These included those isolated in the XL-RBPome and those added as described above. Some proteins in this list were also found to be methylated in a previous study (Ferreira *et al.*, 2020). Post-translational modifications of RBPs, such as methylation, have been proposed to act as means of regulating the RBPs which regulate gene expression. PRMT7 is a methyltransferase that specifically catalyses mono-methylation of Arginine residues in *Leishmania major*. RBPs were highly enriched among the target proteins of PRMT7 with PRMT7 deletion resulting in the hypo- or hyper methylation of many different RBPs. The RBPs selected for the *L. mexicana* bar-seq screen that were either hypo-methylated, hyper-methylated or methylated but unaffected after PRMT7 deletion are indicated in supplementary Table 7.6 (Supplementary).

4 Screening *Leishmania mexicana* RBPs for involvement in differentiation or infectivity

4.1 Introduction

As discussed previously (1.3.3), RNA-binding proteins have not been studied extensively in *Leishmania*. Several large scale studies in kinetoplastids have determined which proteins bind RNA, and with the addition of the *Leishmania mexicana* RBPome, it is possible to link RNA-binding (Erben *et al.*, 2014; Lueong *et al.*, 2016; Nandan *et al.*, 2017) and transcript target selection (De Pablos *et al.*, 2019) to specific lifecycle stages. This study also compares the protein and RNA expression of these RBPs across the lifecycle stages. To further understand the roles of these proteins and identify *trans*-regulators of key processes in *Leishmania*, a bar-seq screen was conducted using the high throughput CRISPR/Cas9 system that has been designed and recently optimised for use in kinetoplastids (Beneke *et al.*, 2017). Drug resistance donor DNA cassettes were amplified by PCR from plasmid templates, with 30bp homology flanks included in the primers. DNA cassettes encoding small guide RNAs (sgRNA) targeting upstream and downstream of the gene of interest were amplified by PCR using a standard reverse primer as template. Both of these DNA constructs were transfected into *Leishmania mexicana* parasites constitutively expressing T7 RNA-polymerase for production of sgRNAs and Cas9 for targeted introduction of double stranded breaks. In a successful knockout, the donor DNA replaced the gene of interest, introducing drug resistance. Once produced, knockouts were screened for loss of fitness phenotypes that may be linked to RBP involvement in differentiation or infectivity. The methodology for the bar-seq screen presented in this chapter was derived from similar studies investigating different classes of proteins in *Leishmania mexicana* and other related organisms (Costa *et al.*, 2018; Damianou *et al.*, 2020; Baker *et al.*, 2021).

A schematic representation of the experiment is shown in Figure 4.1. Unique, twelve nucleotide DNA "barcodes" were introduced to each knockout cell line through primer sequences that were incorporated into the transfected donor DNA. This was utilised to conduct pooled assays investigating differentiation from procyclic promastigotes to metacyclic promastigotes as well as infectivity in a macrophage and a mouse model of infection. DNA sequencing at key progressive timepoints revealed the relative fitness of the RBP null mutant cell lines under different conditions.

The aims of these experiments were to:

- Test which RBP encoding genes can be deleted from the *Leishmania* genome and still produce viable promastigote parasite clones.
- Produce a library of barcoded RBP null mutant *L. mexicana* cell lines for use in future projects.
- Test the RBP null mutant cell lines that are viable in promastigotes in response to:
    - Differentiation from procyclic promastigotes to metacyclic promastigotes.
    - Infection of bone marrow derived macrophages.
    - Infection of mouse footpads.

.

4.2 CRISPR/Cas9 system optimisation

The method described by Beneke *et al.* (2017) improved the feasibility of large scale knockout screening in *Leishmania* because constitutively expressing T7 RNA-polymerase and Cas9 means that only PCR products need to be transfected (Beneke *et al.*, 2017; Figure 4.2). Crucially, the plasmid-based system for amplification of drug resistance cassettes has been designed so that the same primers can be used to produce KO or tagging cassettes with different resistances or tags (Dean *et al.*, 2015). This method reduces time designing and producing constructs and is easily adapted to many different situations. For this study, the JM6571 cell line was selected as it has reasonable CRISPR/Cas9 efficiency but also been recently passaged through mice giving it a high infectivity both in footpad and macrophage models of infection (Baker *et al.*, 2021). An initial knockout experiment was performed in this cell line using the DEAD-box helicase LmxM.15.0130, which associated most with mRNA in the metacyclic promastigote stage in the XL-RBPome. Primers were designed using the LeishGedit server (Beneke *et al.*, 2017) for tagging and knockout cassettes including 30bp of homology upstream and downstream of *LmxM.15.0130*. Primers were also generated through the same server to produce 3' and 5' sgRNA encoding DNA constructs. Both sgRNA and drug resistance encoding PCR products were nucleofected into JM6571 cells in mid log phase which were recovered for several hours before drug was added and the culture was plated out on 96 well plates and diluted to select for clonal populations. After two weeks, five clones were transferred from plate to flask and DNA was isolated for analysis. PCR analysis showed that three

out of five clones had at least one gene copy remaining, whereas the remaining two had no detectable copies of the *LmxM.15.0130* CDS (Figure 4.3a).

To investigate the growth phenotypes in promastigote parasites, growth curves were produced comparing the two knockout cell lines to the three that retained at least one copy of the *LmxM.15.0130* CDS and to the JM6571 (Figure 4.3b). A 50% decrease in cell number appeared during the transition from late log phase to the stationary phase for both knockout clones. The incomplete knockouts showed normal growth, similar to JM6571 except for a slight decrease in cell number after day seven. These results suggested this initial experiment had identified an RBP which, while not essential for promastigote growth, had a severe effect on fitness when removed.

4.3 Producing RBP knockout lines

After successfully removing a *Leishmania* RBP, a screen was designed to scale this process up and assess the fitness of multiple knockout cell lines simultaneously. Unique 12 nucleotide barcodes were added to the upstream forward primer for each RBP knockout attempt as described in previous studies (Beneke *et al.*, 2017; Baker *et al.*, 2021). PCR was used to amplify both a *PUR* and a *BSD* containing donor DNA fragment for each gene. Both donor cassettes were transfected simultaneously with the sgRNA cassettes to ensure removal of at least two copies of the gene of interest after double drug selection.

After drug selection, parasite cultures were diluted to several concentrations (1:5,1:10,1:50,1:100) or serially diluted on 96 well plates to produce a culture with greatly reduced genetic diversity. These are referred to as clones throughout this text to differentiate them from the heterogenous populations produced prior to dilution (discussed in 6.3.2). After clones were isolated they were passaged at low concentration to reduce the chance of DNA from dead cells interfering with diagnostic PCRs. PCRs to check integration and CDS removal were carried out on DNA extracted from the passaged culture of each transfected cell line (Figure 4.4a). Schematic representations show the orientations of the diagnostic PCRs (Figure 4.4b). The results for most of these PCRs were inconclusive and hard to interpret. Several DNA extraction methods were tested to find the optimal balance of speed, cost and DNA quality. The column-based tests (DNeasy and E.Z.N.A blood and tissue extraction kit) performed consistently better than quicker DNA extraction methods such as PCRBio (Figure 4.6). The E.Z.N.A kit was significantly cheaper per column and produced more DNA of similarly high quality to the DNeasy kit so was used for the rest of the experiment. PCR conditions were also optimised to produce consistent results across different experimental stages (2.5.4).

Knockout attempts selecting with one drug only produced many incomplete knockout cell lines that may be heterozygotes (Figure 4.6). Perhaps surprisingly, double drug selection of the transfected populations without subsequent dilution also produced parasite cultures where copies of the target CDS remained. This is likely due to the high rate of aneuploidy in *Leishmania* parasites which is discussed further in sections 1.2.3 and 4.10.3. Because of this, single drug selection and heterogenous parasite populations were avoided for this experiment. Overall 28 out of 67 attempted knockouts

clearly showed successful integration of donor DNA at the correct locus and removal of the gene of interest (Figure 4.7). A further 16 showed successful integration of resistance cassettes but incomplete removal of the gene of interest. Some of these, for example Δ*LmxM.27.0130*, had a pattern of multiple bands including one of the expected size for the CDS of interest and were rejected as partial knockouts to preserve the quality of the screen (Figure 4.4a). Knockout of *LmxM.30.1650* was carried out by Natalia Teles and is currently under investigation. The remaining knockout attempts yielded no viable parasites after transfection and drug selection (overview: Figure 4.8). The majority of these were attempted multiple times (supplementary Table 7.4), often producing no viable clones again, suggesting that these RBPs may be essential in *L. mexicana* procyclic promastigotes. Essentiality of RBPs would be consistent with the theory that post-transcriptional regulation is the primary mechanism of gene regulation in *Leishmania* parasites (De Pablos, Ferreira and Walrad, 2016). The 28 cell lines with PCR results indicating complete RBP gene deletion were taken forward as candidates for the pooled bar-seq screen.

4.4 Testing differentiation to metacyclic promastigotes

To analyse the differentiation from procyclic to metacyclic promastigotes the screen was initially designed to include a metacyclic purification, before infecting both macrophages and mice to characterise infectivity and amastigote survival. However, some RBP knockout cell lines were seen to have possible morphological phenotypes. The most striking being Δ*LmxM.05.0850* which lacks a visible flagellum and is shorter and wider than wild type or JM6571 *Leishmania mexicana* (Figure 5.13). Metacyclic purification involves the centrifugation of parasite culture through a ficoll gradient, which separates the smaller, lighter metacyclic promastigotes from the heavier, rounder procyclic promastigotes, leptomonads or nectomonads (if present). Δ*LmxM.05.0850* cells were tested by ficoll gradient purification at day 8 of a standard growth in Grace's medium alongside JM6571 cells and were largely found in the discarded pellet with very few cells remaining in the supernatant. This would mean that for cells with morphological defects, the purification would result in removal from the screen before infectivity was assessed. Consequently, to avoid exclusion due to properties outside the intended screen, metacyclic purification was performed in the pooled screen and sequenced but stationary phase culture was used for inoculation of mouse footpads and bone marrow-derived macrophage culture.

4.5 Testing macrophage infection

Bone marrow macrophages were extracted from BALB/c mouse femurs for testing pooled, stationary phase RBP knockout cell lines for infectivity. Harvesting of bone-marrow derived macrophages was carried out with the help and advice of Rachel Neish (see acknowledgements). The same strain of mice was used for macrophage extraction and footpad infection to make the results comparable. Uptake of JM6571 *Leishmania mexicana* amastigotes was tested on bone marrow derived macrophages that had been frozen, defrosted and re-activated in differentiation medium (0). Infection ratios of 1:1, 2:1 and 8:1 were tested with a six-hour incubation before being washed to remove

extracellular parasites. After 72 hours, DNA was extracted from all macrophage cultures and used as the template for PCRs to test amplification of the barcode region. PCR using DNA extracted from a 1:1 infection ratio as template showed a weak amplification when either stationary phase or purified metacyclic promastigotes were used for inoculation (Figure 4.9a). At 2:1, DNA from the same number of macrophages infected with purified metacyclic promastigotes consistently produced a stronger PCR amplification, likely indicating a more successful infection. With an 8:1 infection ratio, the band strength was equal to that of a PCR from *Leishmania* cell culture derived DNA template (Figure 4.9b). However, the 8:1 heavily infected macrophages, unlike those at lower infection ratios, had many extracellular amastigotes. It was important to balance between a detectable level of infection and intact cells as well as reducing the number of attached promastigotes that require washing off the macrophage surface. A ratio of 6:1 was used for the main screen followed by thorough washes to reflect this. The success of these washes during the final experiment at 24h and 48h post infection can be evaluated in Figure 4.10.

4.6 Pooling RBP knockout cell lines

The introduction of unique 12-nucleotide barcodes for each RBP knockout cell line allowed for the pooling of all 28 successful knockouts and three controls. The three control cell lines each contain a unique barcode that was integrated into the SAT resistance region in the ribosomal locus, originally added as a selection marker to maintain stable expression of T7 RNA polymerase (Beneke *et al.*, 2017; Baker *et al.*, 2021). This should mean they would be identifiable but aphenotypic and could be used as wild type controls. Subsequent exposure to differentiation conditions was followed up with extraction of DNA, PCR of the barcoded region and sequencing allowing quantitative identification of each cell line in the population. After all cell lines were grown to mid log phase (~$5x10^6$/ml) they were pooled to a final concentration of $1x10^6$/ml and DNA was extracted for PCR amplification and sequencing. The sequencing results show that pooling in equal proportion was largely successful with a few exceptions (Figure 4.11). Cell lines Δ*LmxM.33.2580*, Δ*LmxM.05.0850*, Δ*LmxM.22.1500* and Δ*LmxM.36.1635* all had a mean read count above one standard deviation around the group mean. Δ*LmxM.30.1650* and Δ*LmxM.24.1570* also showed more moderately raised day 0 read counts than group mean. Whole genome sequencing results later showed that these variations are likely due to multiple integrations of the barcoded DNA at the same locus in individual clonal lines (see 5.2). The pools were produced six times independently to account for pooling error, and each pool progressed independently through the whole experiment (for example: pool one inoculated mouse one, pool 2 inoculated mouse 2). During the first seven days of the pooled experiment the six replicate pools were each counted to check that total parasite growth was comparable between replicates (Figure 4.12). Overall, these six replicates showed a standard growth curve for JM6571 *Leishmania mexicana* in Grace's medium. Inoculation, monitoring and harvesting of mouse footpads was carried out by Rachel Neish and Jayanthi Anand.

4.7 Barcode sequencing preparation

At each time-point indicated on the timeline (Figure 4.1) DNA was extracted and snap-frozen at -80ºC until the final samples had been processed. This included post-pooling timepoints (0h,24h,48h,168h), after metacyclic promastigote purification (at 168h post-pooling), post-inoculation of bone marrow-derived macrophages (24h,72h) and post inoculation of BALB/c mouse footpads (three and six weeks). Extracted DNA was used as a template for PCR amplification of the barcode region from the pool of cell lines. Initial PCR tests on the axenic culture-derived parasite stages were conducted to check the consistency of the DNA samples and optimise the annealing temperature (Figure 4.13a). Because the binding site of the reverse primer was present in all transfected donor DNA constructs in the bar-seq screen, it was hypothesized that off target sequencing of contaminants introduced during PCR could potentially occur (Figure 4.13b). Contaminating bands were seen in early PCR test controls containing no template DNA. To reduce the chance of this happening an alternative primer was designed which bound some of the original binding sequence and some non-primer specific sequence (1768). A second alternative reverse primer (1767) was designed to bind entirely in the non-primer region. However, the caveat to using either alternative reverse primer is that they only amplify barcodes from the integrated *BSD* containing donor DNA, reducing the total template available for sequencing. The second alternative reverse primer (1767) was rejected for this experiment due to poorer results when using BLASTn to check for off target binding in the genome (Altschul *et al.*, 1990).

Reverse primer 1768 was used for the main experiment in combination with the standard forward primer 1766. The annealing temperature of 60°C was used as it is the most stringent but did not show a reduction in PCR yield. This primer pair was tested in PCR of DNA from an infected and an uninfected mouse footpad as well as from the parental T7/Cas9 *L. mexicana* strain (JM6571) (Figure 4.15). This result showed that using the altered reverse primer (1768) and restricting PCR preparation to a decontaminated alternative location, no off-target amplification was seen. A weak, diffuse band below 100bp most likely represents excess primer which decreased in intensity as it was incorporated into the amplified region in lanes six and seven. After DNA-extraction of all samples from the main screen, more PCR tests were carried out to optimise the number of cycles used to amplify the barcode region (Figure 4.14a). In these initial tests, an equal amount of DNA template (2µl) was added to every reaction and 25 cycles of PCR were used throughout. While the flask grown stages gave consistent results, there was a lot of variability in the amastigote stages. This is because even if total DNA concentration was normalised, the excess mouse DNA interferes with quantification. Based on the results of these initial tests the template concentrations were normalised and the cycles of macrophage derived samples were increased to 30 to improve PCR yield (Figure 4.14b). The repeated PCRs were much more evenly amplified and had a greater yield. In an optimal situation, the number of cycles used should be less than the plateau reached when any one of the PCR components is used up, i.e. the logarithmic phase of the amplification. For this reason, all barcode amplifications were compared to PCRs with the same conditions but over amplified (40 cycle

positive control). Amplifications from stationary phase with reduced template concentration and purified metacyclic promastigote DNA were repeated at 28 and 25 cycles respectively to reduce the chance of over amplification. Once consistent, PCRs were repeated once more and pooled to produce enough DNA template for sequencing.

4.8 Analysis of results

Raw sequencing reads, representing the relative proportion of barcodes from each cell line, were normalised by total read count so that the relative proportion of each cell line in the total pool can be compared over time. Assuming that the number of barcode reads is proportional to the number of parasites in the pool, changes in the normalised reads between timepoints indicate changes in fitness of the RBP knockout cell line. As the cell lines progressed through the life cycle, reductions in cell number were cumulative, a factor that must be considered when analysing the results. There is also an element of competition with the other knockout parasites and the control cell lines that must be considered. Both caveats are discussed further in section 4.10.7

The sequencing was largely successful with only one replicate of the stationary phase that failed sequencing (S168_5). For stationary phase samples, mean reads were calculated using five replicates instead of six with a resultant small increase in error. One barcode for the cell line Δ*LmxM.30.0250* was missing (CGCGTTTCTTAA) from the high-level reads. Whole genome sequencing showed that Δ*LmxM.30.0250* was a true null mutant at the correct locus but analysis of reads through the barcode region revealed the 12nt barcode was replaced with the sequence GTAG. The common regions flanking the barcode were intact. The primers used to introduce the 12nt barcode were correct suggesting an error introduced between PCR of the donor DNA cassette and sequencing of the null mutant clone. Because the barcode could not be detected, this line has been excluded from the data but the whole genome sequencing is presented in section 5.2.

Viewing the data as a collection of individual growth curves produced a visually complicated result that failed to present patterns clearly but individual trajectories are presented in supplementary **Figure 7.2**. Other studies have used clustering methods to identify growth curves with similar trajectories (Ruy *et al.*, 2019; Baker *et al.*, 2021) but given the smaller data set in this study, this technique is not suitable. Viewing normalised read counts as a bar graph can be informative for looking at changes in the composition of the pool over time (Figure 4.16). The total number of reads for each pool at each experimental time-point was used to normalise the data and the relative number of reads at day zero was normalised to account for barcode copy number evidenced by whole genome sequencing (see section 5.2). A bar chart was plotted with each cell line shown as a proportion of the total pool at each stage of the experiment. The three barcoded control cell lines successfully differentiated through all lifecycle stages tested, including infection of the mouse and the bone marrow derived macrophages. The three control cell lines (sand fly lines 1,2 and 3, Baker *et al.*, 2021) were included as aphenotypic controls so were predicted to be successful at every experimental stage and to have a similar trajectory to each other. Figure 4.16 shows that this is the case with

consistently similar proportions of the pool representing each control even after six weeks post infection in the mouse model. Individual tracks in this figure can be highlighted using the filter function (e.g. Figure 4.16b). Over time, the number of cell lines constituting a major part of the barcode pool decreased as the impact of the RBP knockouts negatively affected cell numbers. By the final samples, six weeks post footpad infection in the mouse, only eight knockouts and three control cell lines made up 90% of the total sequencing reads (SF1, SF2, SF3, Δ*LmxM.29.2200*, Δ*LmxM.26.1530*, Δ*LmxM.36.5820,* Δ*LmxM.19.0190,* Δ*LmxM.18.0590,* Δ*LmxM.14.1140,* Δ*LmxM.31.0950* and Δ*LmxM.17.0550*). As well as the reduction in the complexity of the pool over time, there is a general trend of increasing variation of the log2 fold change normalised barcode reads between cell lines as the experiment progresses (Figure 4.17). There is also an overall trend towards negative changes in fitness compared to the controls rather than positive ones, as expected in a bar-seq screen where fitness is usually negatively impacted by CDS replacement.An alternative bar chart can be used to look at the relative fitness of knockouts in different lifecycle stages (Figure 4.18). The x-axis was ordered from highest to lowest barcode read count at six-weeks post-infection in the mouse model. As expected the three control cell lines are placed at the left with high read counts indicating fitness at every stage of the experiment. Only three cell lines make up a larger proportion of the final pool (FP6W) than the controls, Δ*LmxM.17.0550*, Δ*LmxM.31.0950* and Δ*LmxM.36.5820*, suggesting an increased fitness in the mouse model of infection. Five knockout cell lines have a similar profile to the control cell lines: Δ*LmxM.19.0190*, Δ*LmxM.14.1140*, Δ*LmxM.29.2200*, Δ*LmxM.18.0590,* Δ*LmxM.26.1530*, suggesting that knockout has little effect on cell numbers during the course of the experiment.

Many cell lines have dramatically reduced cell counts in the human infectious stages but at low read counts the differences can be hard to compare on a bar chart. To clarify the results further, a heatmap was produced with the $\log_2$ fold change in normalised read count for each cell line compared to the mean of the three control cell lines at each stage (Figure 4.19). Statistics reported on the heatmap were derived from a one-way ANOVA (Kruskal-Wallace) in each experimental stage between the six replicates of each knockout cell line (five for stationary phase) and six replicates of the mean of control cell lines 1-3. The assumption made here is that the control cell lines have no significant phenotype. To adjust for the level of type one error from repeated tests, the two-stage step-up method of Benjamini, Krieger and Yekutieli was used as a correction (Benjamini, Krieger and Yekutieli, 2006). Time-point specific plots presenting the individual replicates for each barcoded line with the mean of the control cell lines are shown in Figure 4.20. Cell lines where the comparison of the medians shows a significant difference from the controls (green) are highlighted in blue. Overall, this clearly shows that knockout of RBPs generally has a negative effect on cell fitness as the majority of cell lines have a negative fold change. Many cell lines showed a significant negative fold change in reads during the promastigote stages, the most pronounced being: Δ*LmxM.25.0290*, Δ*LmxM.15.0130,* Δ*LmxM.31.3390,* Δ*LmxM.30.1650,* Δ*LmxM.22.1500,* Δ*LmxM.19.0295,* Δ*LmxM.05.0850,* Δ*LmxM.18.1420*. A negative fold change was also seen between stationary phase and the purified metacyclic promastigote populations. As expected, LmxM.05.0850 read count was

dramatically reduced. Other cell lines with a similar dramatic reduction at this stage despite having no obvious morphological phenotype include: Δ*LmxM.18.1420,* Δ*LmxM.19.0295,* Δ*LmxM.31.3390,* Δ*LmxM.32.1150,* Δ*LmxM.33.4550,* Δ*LmxM.36.1635*. These RBPs could represent those that fail to differentiate to metacyclic promastigotes and warrant further investigation. Another RBP null mutant line that is severely reduced in the metacyclic population is Δ*LmxM.36.6770* but this is likely because of the dramatic reduction seen in the stationary phase compared to 48 hours after pooling.

Comparing the first 24 hours of macrophage infection to the stationary culture (day seven) used as the inoculum, several cell lines have reduced further in fitness: Δ*LmxM.18.1420,* Δ*LmxM.19.0295,* Δ*LmxM.23.0730,* Δ*LmxM.24.1570* and once again Δ*LmxM.36.6770*. Because of the timing, it is possible that these represent RBPs where knockout affects parasite uptake. By 72 hours post infection several cell lines appear to have lost fitness in comparison to those at the 24-hour time point: Δ*LmxM.36.6770* which is extremely reduced at this point, Δ*LmxM.34.2270,* Δ*LmxM.33.4550,* Δ*LmxM.24.1570,* Δ*LmxM.23.0730,* Δ*LmxM.19.0295,* Δ*LmxM.18.1420,* Δ*LmxM.15.0130 and to a lesser extent* Δ*LmxM.13.0450*. For those that were largely unaffected during uptake (24 hours post infection) but reduced dramatically after 72 hours, it is likely that survival or replication of amastigotes has been perturbed by RBP knockout.

The strongest negative fold changes were observed in the mouse model of infection with Δ*LmxM.36.6770* nearly being eliminated altogether at three weeks post infection. At this time point strong negative fold change was seen compared to stationary phase culture for: Δ*LmxM.36.1635,* Δ*LmxM.33.4550,* Δ*LmxM.33.2580,* Δ*LmxM.27.0130,* Δ*LmxM.25.0290,* Δ*LmxM.24.1570,* Δ*LmxM.23.0730,* Δ*LmxM.22.1500,* Δ*LmxM.19.0295,* Δ*LmxM.19.0190,* Δ*LmxM.05.0850*. Of these, all had a further negative fold change when comparing six weeks post infection to three weeks post infection in the mouse. Cell lines that did not appear to have a fitness defect in the first three weeks of infection did not dramatically reduce fitness further at six weeks post infection.

One anomaly that was observed clearly was the increase in sequencing reads for Δ*LmxM.24.1570* in the stationary phase seven days post pooling. When checking the raw reads this relative increase was seen across all six repeats ruling out single sample anomalies that distort the mean read count. Subsequent experiments found no evidence of a dramatic increase in cell count for this cell count when grown to stationary phase alone (Figure 5.4).

4.9 Conclusions

In summary, a bar-seq screen of *Leishmania mexicana* RNA-binding protein null mutant lines revealed stage specific fitness defects as consequences of gene deletion. Significant fitness defects were seen at all stages of the experiment with the most severe occurring in the human infectious amastigote stage. The bar-seq screen results can be used to select RNA-binding proteins for further study into the regulation of differentiation, infectivity and virulence. For RBPs where deletion has induced a strong phenotype in the amastigotes, further study will be pursued as the basis for other projects. Four candidate RBPs are discussed further in Chapter 4.10.

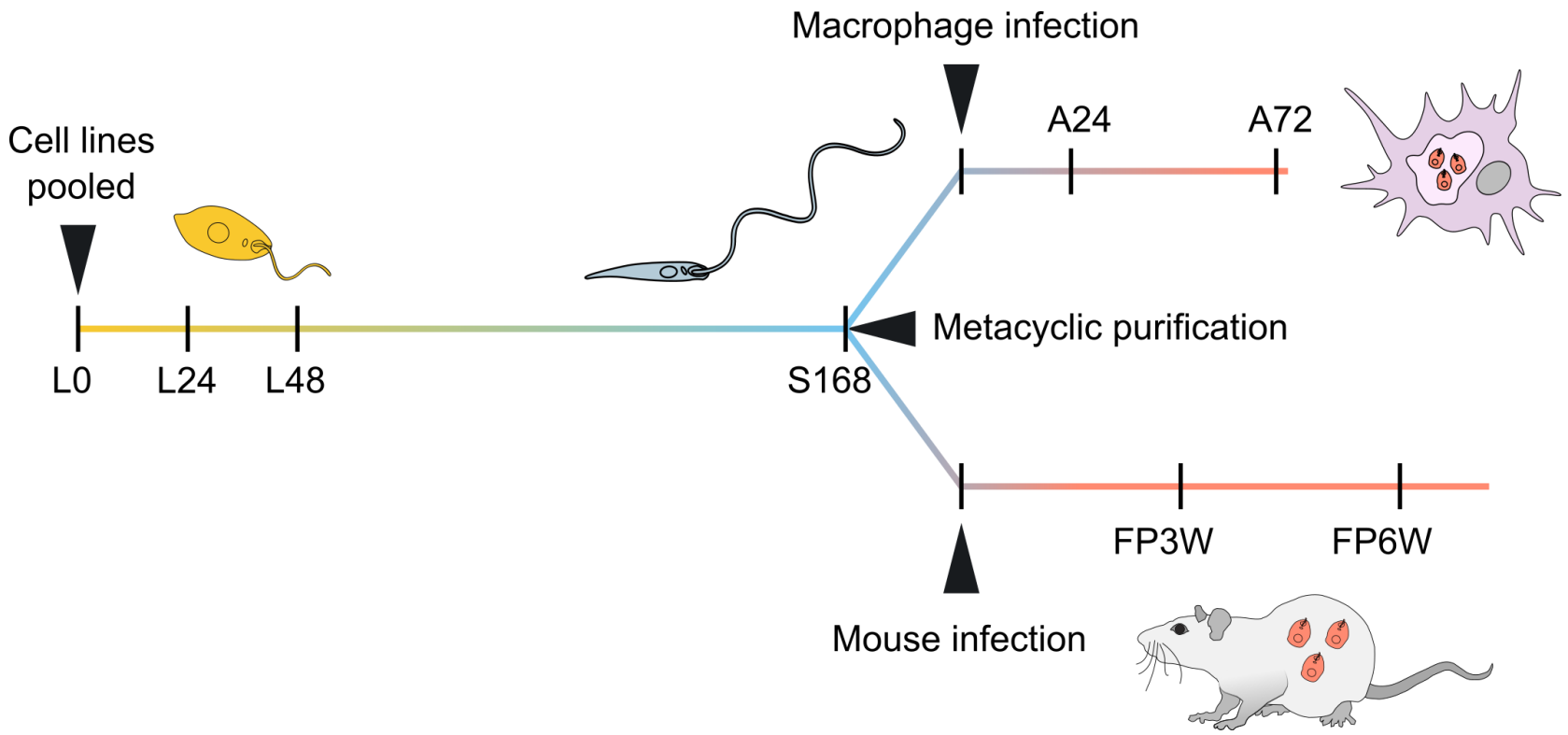


Figure 4.1. Bar-seq screen overview. All 31 cell lines were pooled in equal proportion in six replicate flasks and were differentiated from procyclic to metacyclic promastigotes. DNA samples were taken immediately after pooling as well as one- and two-days post inoculation. Stationary culture was sampled at day seven before and after a metacyclic purification of the culture. Stationary culture was used to infect both mice and bone marrow derived macrophages on day seven. Macrophages were harvested for DNA samples one- and three-days post-infection. Footpad lesions were taken from six mice at three weeks and six mice at six weeks post infection and frozen. DNA from infected mouse tissue and all other samples was extracted after the final samples were taken. L0 = log phase 0h post-pooling, L24 = log phase 24h post-pooling, L48 = log phase 48h post-pooling, S168 = stationary phase 168h (seven days) post-pooling, M168 = metacyclic purification 168h (7days) post pooling, A24 = amastigotes from bone marrow derived macrophages 24h post-infection, A72 = amastigotes from bone marrow derived macrophages 72h post-infection, FP3W = amastigotes from mouse footpads three weeks post-infection, FP6W = amastigotes from mouse footpads six weeks post-infection.

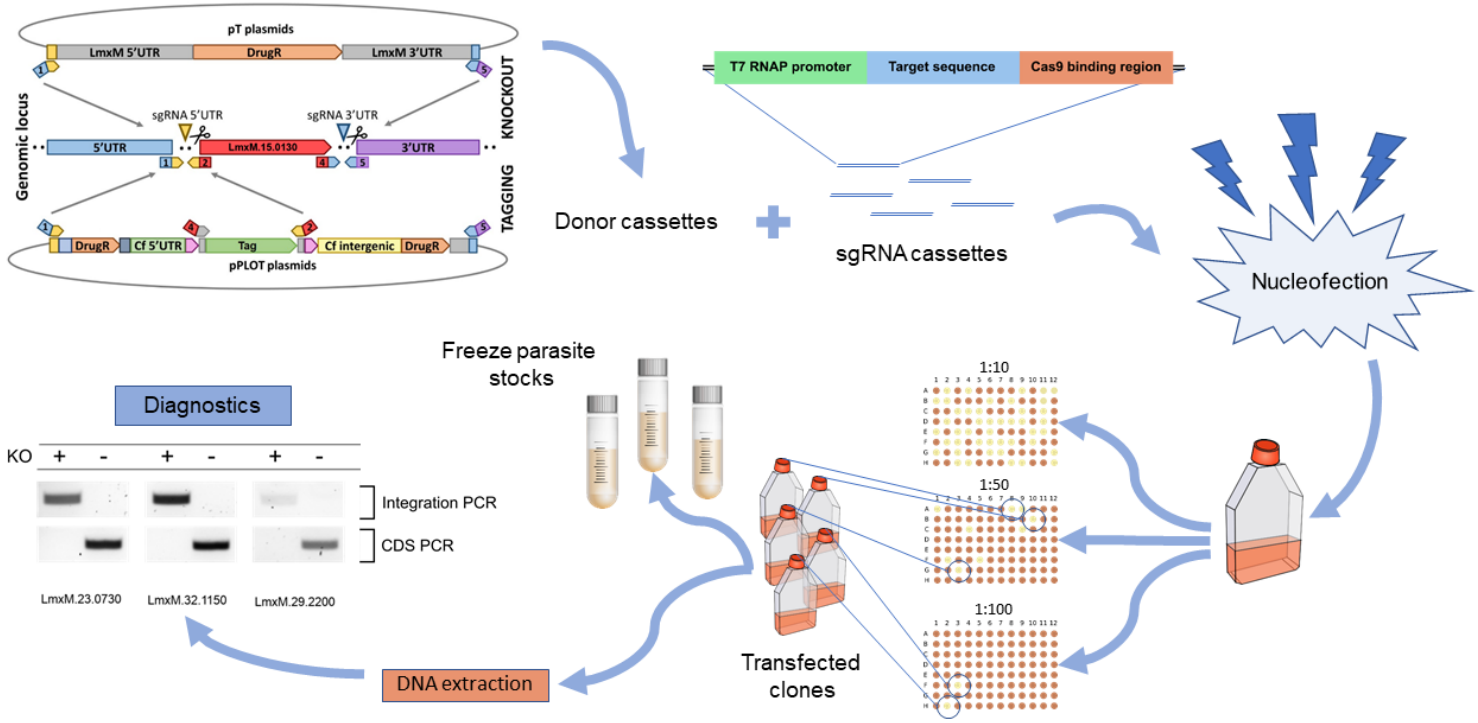


Figure 4.2. CRISPR/Cas9 knockout workflow. Donor DNA was amplified from two different pT plasmid templates, one containing a *BSD* gene (Blasticidin resistance) and one containing a *PUR* gene (Puromycin resistance) (Beneke *et al.*, 2017). Thirty base-pairs of homology upstream and downstream of the gene of interest were included in the primers for targeted replacement of the correct locus. A unique 12 nucleotide barcode was included in the upstream forward primer. DNA cassettes for production of both a 5' and 3' sgRNA *in-vivo* were amplified using a gene specific forward primer and a common reverse primer. All four DNA cassettes were nucleofected into T7-RNA polymerase and Cas9 expressing *L.mexicana* (JM6571). After a short recovery, both drugs are added to the culture and the parasites are diluted onto 96 well plates for clone selection. After two weeks clones were selected and stored as well as sampled for DNA extraction. Extracted DNA was screened by PCR to detect integration of donor DNA and removal of the CDS. Diagram adapted from Beneke *et al.* 2017.

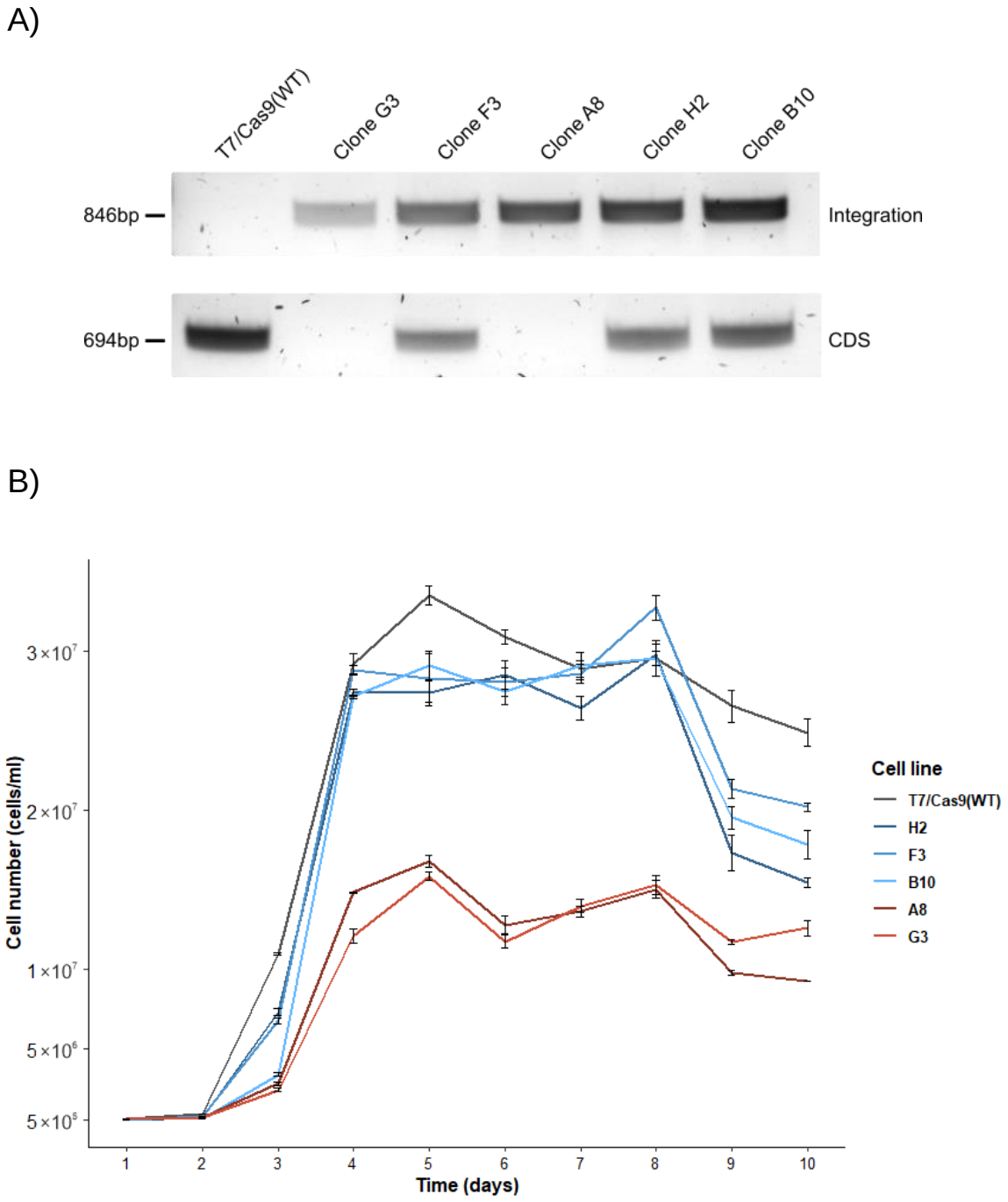


Figure 4.3. Initial CRISPR knockout tests. An initial attempt to replace the CDS of an RBP (*LmxM.15.-130)* was carried out using the CRISPR/Cas9 system. Single drug selection with Blasticidin was used to clone parasites in liquid media on 96-well plates. A) Diagnostic PCRs were carried out on DNA extracted from each clone to check integration of donor DNA and removal of the CDS (see Figure 4.3.1b for maps). PCRs were run on 1% agarose gel with SYBRsafe (Thermo Fisher). The image was inverted in post processing. B) Growth curves of cell lines from A were conducted in triplicate with an initial concentration of 5×10^5. Null mutants A8 and G3 (red) and partial deletions, H2, F3 and B10 (blue) are compared to the T7/Cas9 expressing *L.mexicana* (JM6571) shown in grey.

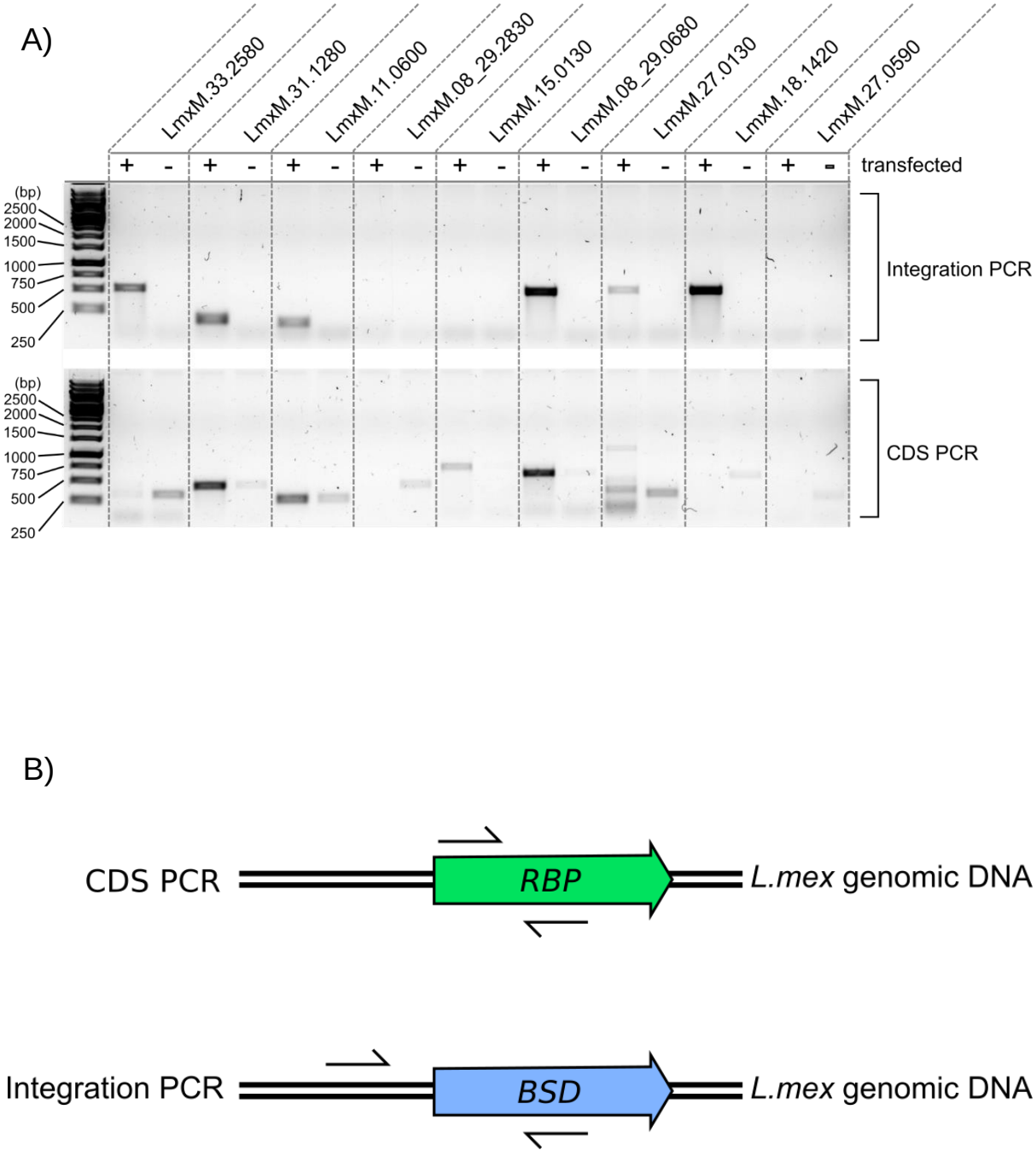


Figure 4.4. Initial diagnostic PCRs. A) PCRs using genomic DNA from cloned, nucleofected parasites as a template were imaged on a 1% agarose gel using SYBRsafe (transfected = +). For each PCR primer pair (RBP specific) PCRs were also carried out using T7/Cas9 *L.mexicana* (JM6571) for comparison (transfected = -). B) Maps describe the PCRs using template DNA from both T7/Cas9 *L.mexicana* (JM6571) and knockouts generated from this cell line. If no integration is present the integration PCR will fail because there is no reverse primer binding site but the CDS PCR will amplify the gene of interest. If integration is successful and the CDS has been replaced, the converse will be true. Amplicon lengths ranged from 100bp to 700bp depending on CDS length and optimal primer binding sites.

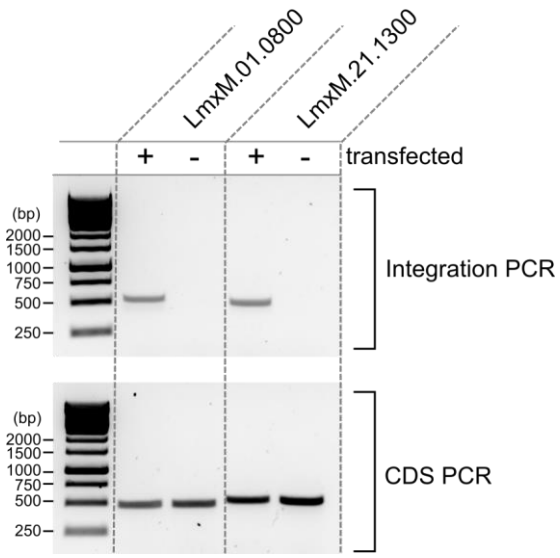


Figure 4.6. Diagnostic PCRs for extra CDS copies. Examples of diagnostic PCRs for clonal populations of *L.mexicana* that were selected with both Blasticidin and Puromycin but show at least one remaining copy of the CDS. This is despite successful integration at the correct locus. Parasites nucleofected with donor DNA = '+', untransfected JM6571 cells = '-'. PCRs are presented on 1% agarose gels using SYBRsafe.

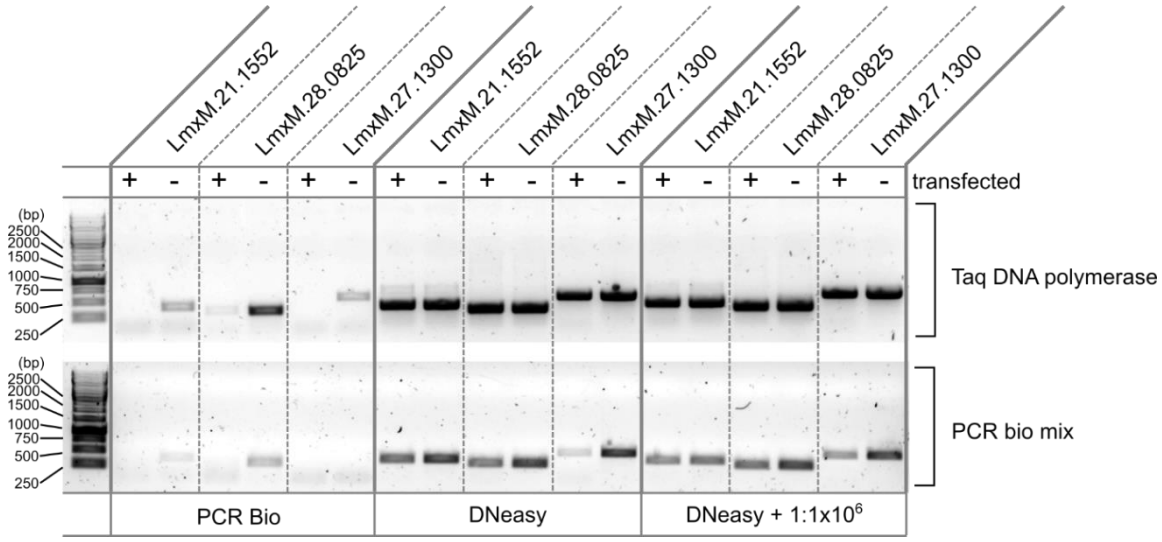


Figure 4.5. Diagnostic PCR optimisation. CDS PCRs as presented in Figure 4.3.1 were repeated using DNA templates extracted with two different kits (PCRbiosystems and DNeasy [Qiagen]). PCRs are presented on 1% agarose gels using SYBRsafe (Invitrogen) for visualisation. The final group was also extracted with a DNeasy kit but parasite cultures were passaged 1:1×10^6 before extraction to remove DNA from dead cells. Two different PCR mixes were compared, the first being a standard Taq polymerase reaction mixed as needed and second a premade Taq mix from PCRbiosystems.

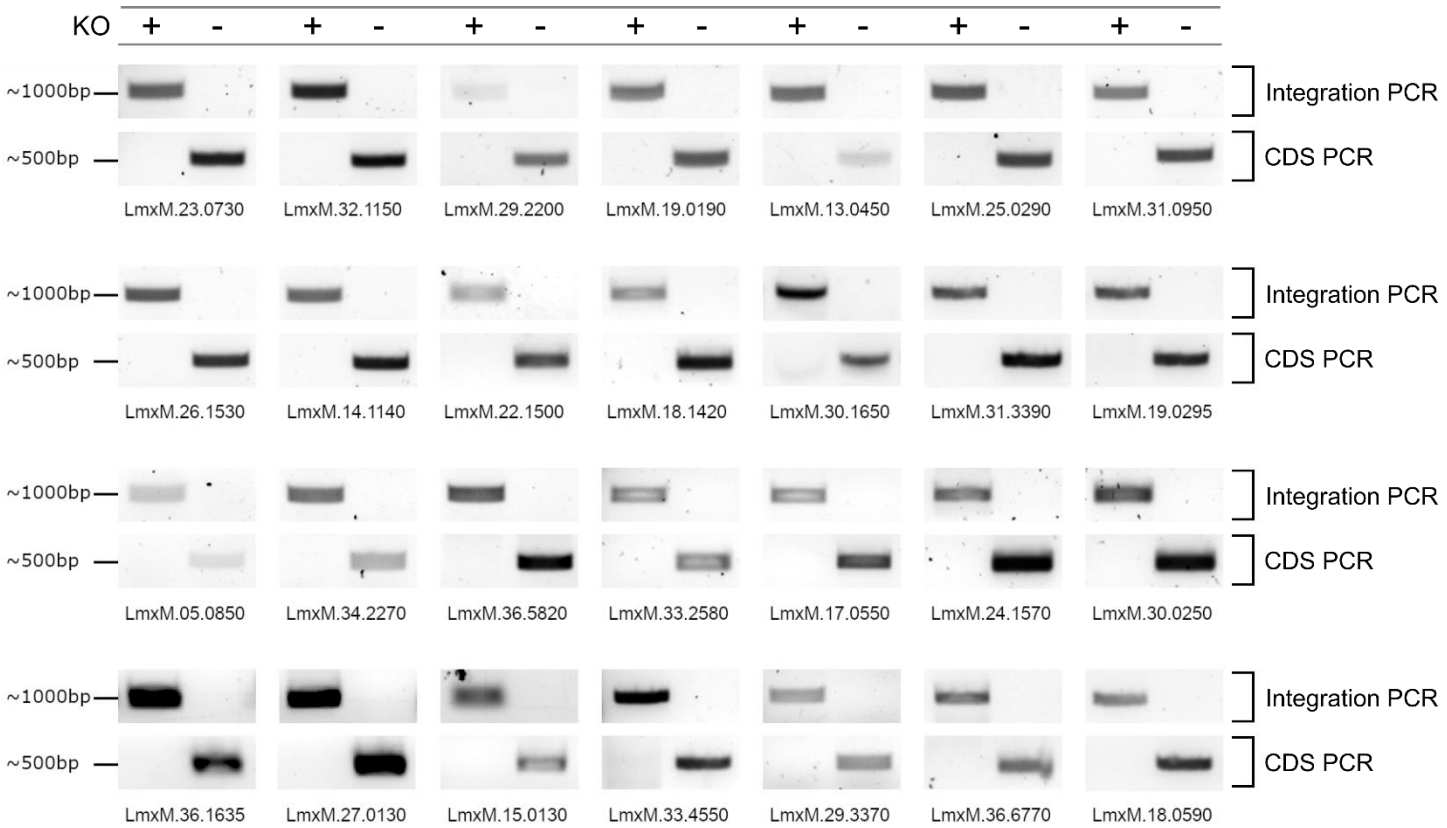


Figure 4.7. Confirmation of 28 complete RBP deletions. Diagnostic PCRs were carried out as explained in Figure 4.4 for clones of all cell lines that were produced during 67 attempted RBP deletions. Of these, 28 (pictured) clearly showed integration of the donor DNA at each RBP locus and no PCR evidence of remaining CDS copies. All PCRs for drug selected clones (KO = '+') were compared with the same PCR from T7/Cas9 expressing *L.mexicana* (JM6571) DNA template (KO = '-'). Images have been inverted in post processing. All integration and CDS PCRs for the same cell line were imaged on the same 1% agarose gel using SYBRsafe (Invitrogen). While the marker ladders are not included here, the band sizes for each PCR product were confirmed on individual separations and a given in supplementary Table 7.5.

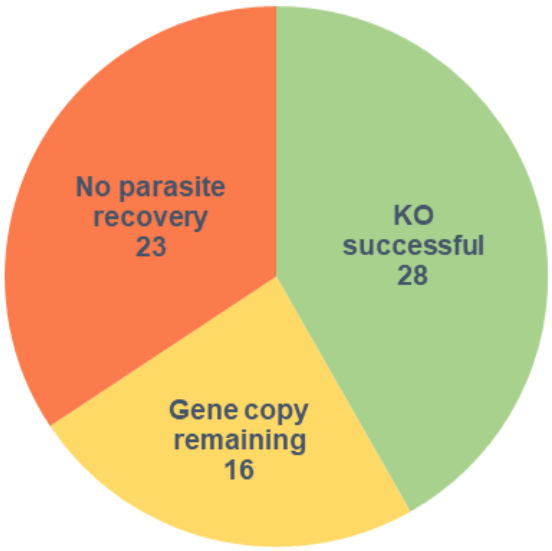


Figure 4.8. Overview of knockout attempts. In total 67 RBP knockouts were attempted for this project. Of these 28 were successful in completely replacing all CDS copies of the RBP of interest. Of the 67 attempts, 16 RBP knockout attempts either had a copy of the gene remaining in all clones tested or had indeterminate PCR results. Finally, 23 attempted knockouts resulted in no viable parasites after drug selection.

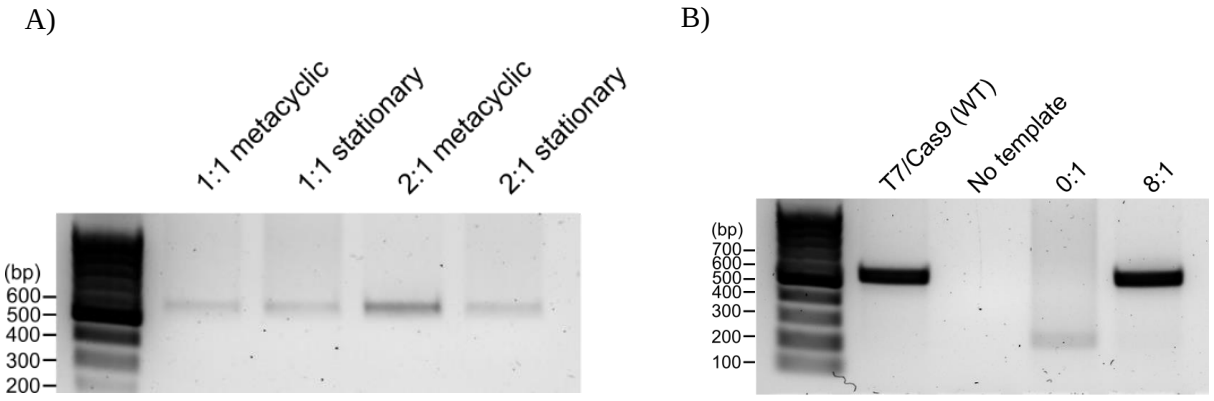


Figure 4.9. Testing PCR amplification from amastigotes. BALB/c bone marrow derived macrophages were infected at 1:1, 2:1 and 8:1 infection ratios of T7/Cas9 expressing *L.mexicana* (JM6571). DNA was extract using the DNeasy kit and used as template for PCRs amplifying a 694bp region from the LmxM.15.0130 locus (un-edited region in JM6571). (A) For the 1:1 and 2:1 infection ratios, the infectivity of stationary phase promastigote culture (seven days post inoculation) was compared to metacyclic promastigotes purified as described in methods 2.2.7. (B) For comparison, the same PCR was carried out on DNA from T7/Cas9 *L.mexicana* cultured in flask as well as uninfected macrophages (0:1).

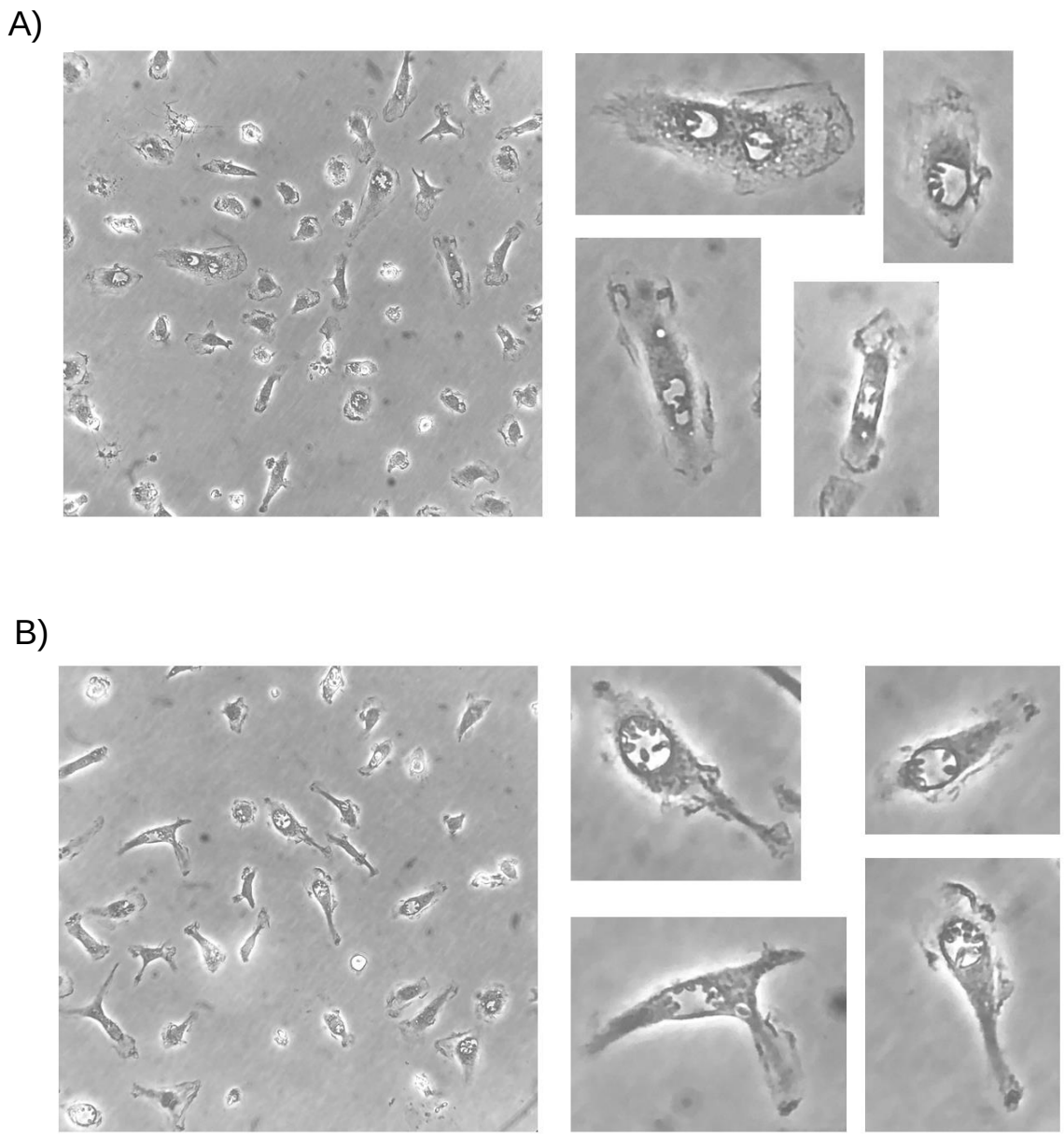


Figure 4.10. Macrophages infected with pooled RBP null mutants. Macrophages infected with the stationary phase of pooled barcoded RBP null mutants were imaged one- and two-days post infection. Images were taken to check attached promastigotes had been washed off the cell surface, to verify macrophage infection and to show that amastigotes had not lysed macrophages or replicated in the culture medium. Exact scale unavailable as plastic culture plates not compatible with accurate imaging systems. *L.mexicana* amastigotes of many different cell lines can be seen inhabiting parasitophorus vacuoles at one day post infection (A). After two days post-infection many amastigotes can be seen in the majority of vacuoles (B).

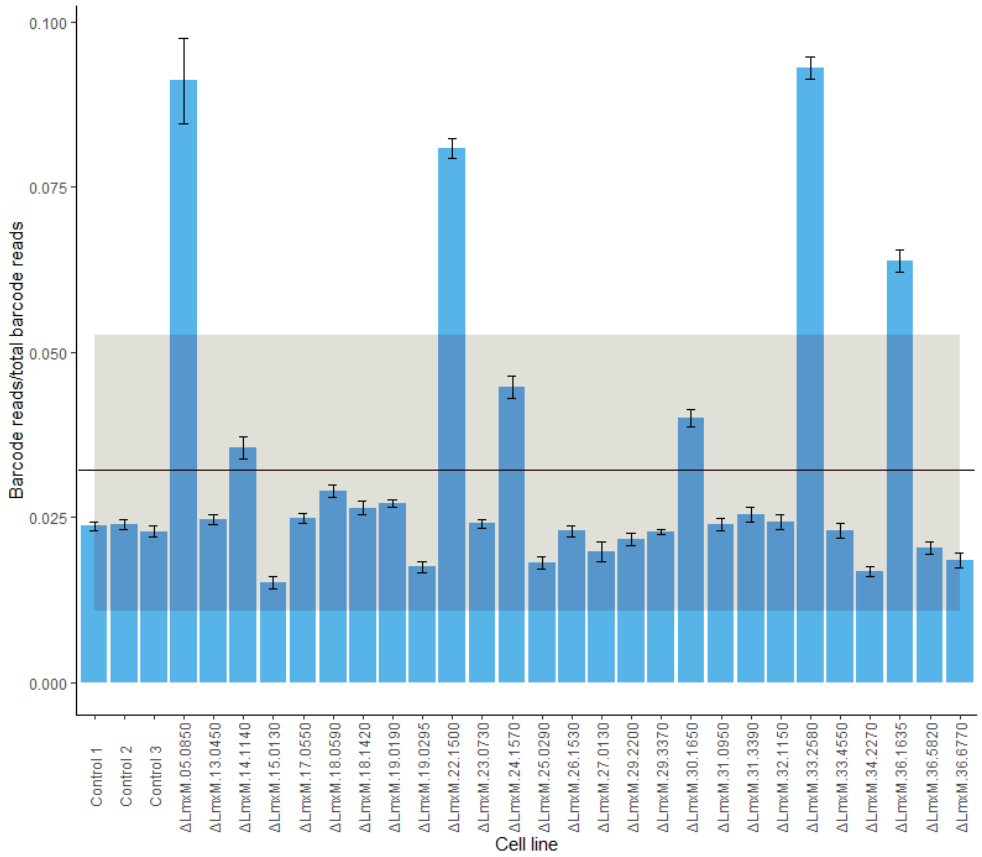


Figure 4.11. Comparison of initial barcode reads. Barcode reads sequenced from DNA samples taken immediately after pooling of RBP null mutant cell lines are presented as a bar graph. Bars represent mean barcodes for each cell line in the six replicate pools. Error bars represent standard deviation. A horizontal line (black) shows the mean barcode reads of all cell lines with one standard deviation either side shaded red. Controls 1-3 represent three independently barcoded control cell lines with no knockout phenotype (Baker *et al* 2021).

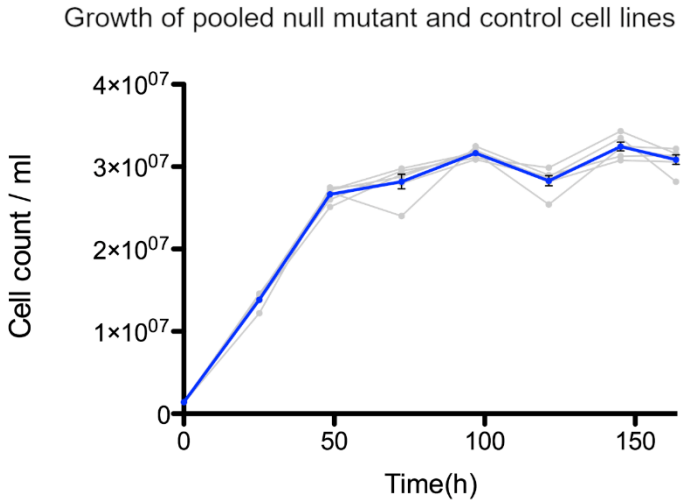
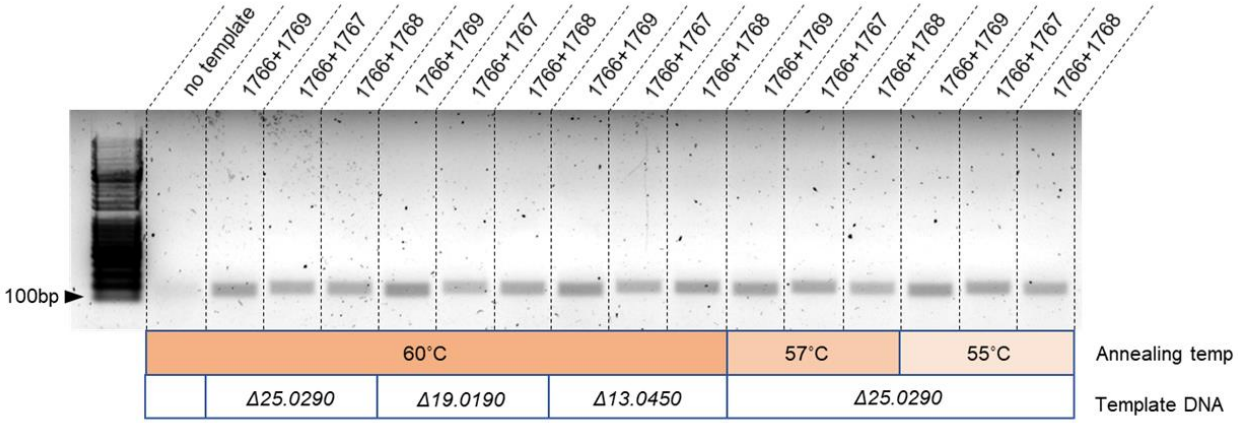


Figure 4.12. Promastigote growth curves of pooled null mutant lines. Growth curves were produced for the six replicates of the RBP null mutant pools to check the reliability of the data. At pooling, the concentration of all cultures was measured to be ~1x10^6 cells/ml. The blue line represents the mean cell count across all six replicate pools with bars showing standard error around the mean.

A)



B)

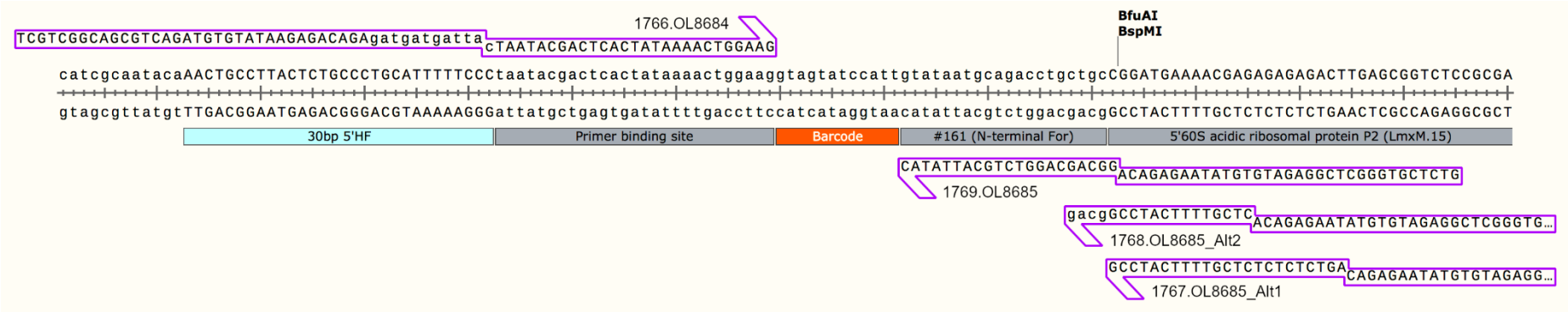


Figure 4.13. Initial tests of barcode amplification by PCR. A) PCR was used to amplify the barcoded region integrated into the genome of RBP null mutant lines when the CDS was replaced. Barcode amplification primers were tested on several different null mutant lines to check for consistency of amplification. Three annealing temperatures 55, 57, and 60°C, were also tested for one of these lines *(ΔLmxM.25.0290*). An amount of water equal to template DNA was added to a 'No template' control which was subject to PCR with primers 1766 and 1769 to detect contaminating template. Three different reverse primers (1767,1768 and 1769) were tested in combination with the same forward primer (1766). B) Alternative primers (1767 and 1768) are shown with slightly shifted binding sites. Reverse primer 1768 was used in combination with 1766 for the samples used in the final sequencing. Predicted sizes for PCR amplified fragments are: 1766+1769 = 139bp, 1766+1768 = 152bp, 1766+1767 = 160bp.

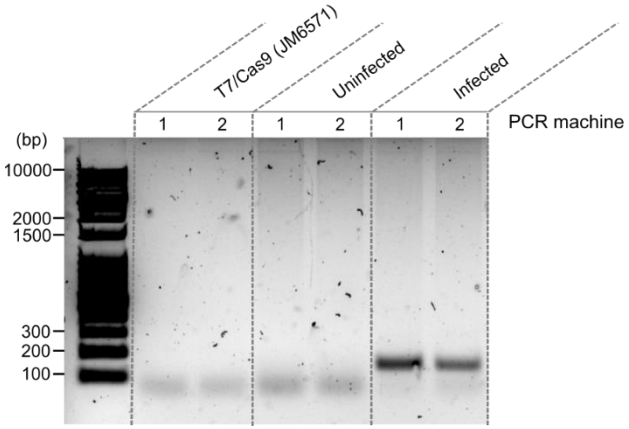


Figure 4.15. Testing barcode amplification from infected mouse footpads. DNA was extracted from T7/Cas9 *L.mexicana* (JM6571) cell culture, uninfected mouse footpads and footpads infected with barcoded *L.mexicana* parasites. PCR was used to amplify the barcode region using the standard forward primer (1766) and alternative reverse primer 2 (1768) using each of these DNA templates. The total product including adaptor binding areas introduced in the primers was 179bp. Two different PCR machines were tested, a decontaminated machine that had been used previously (1) and a machine from a separate lab (2).

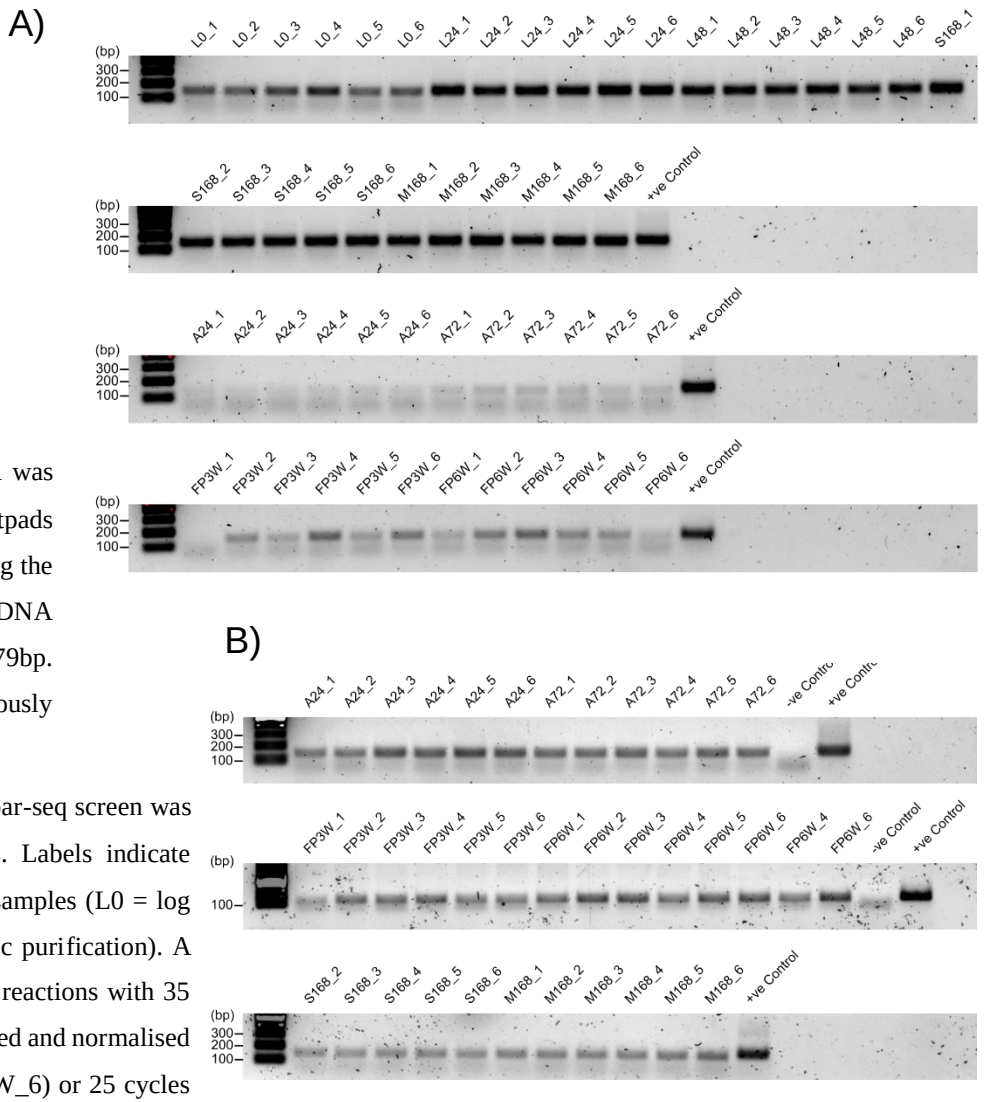
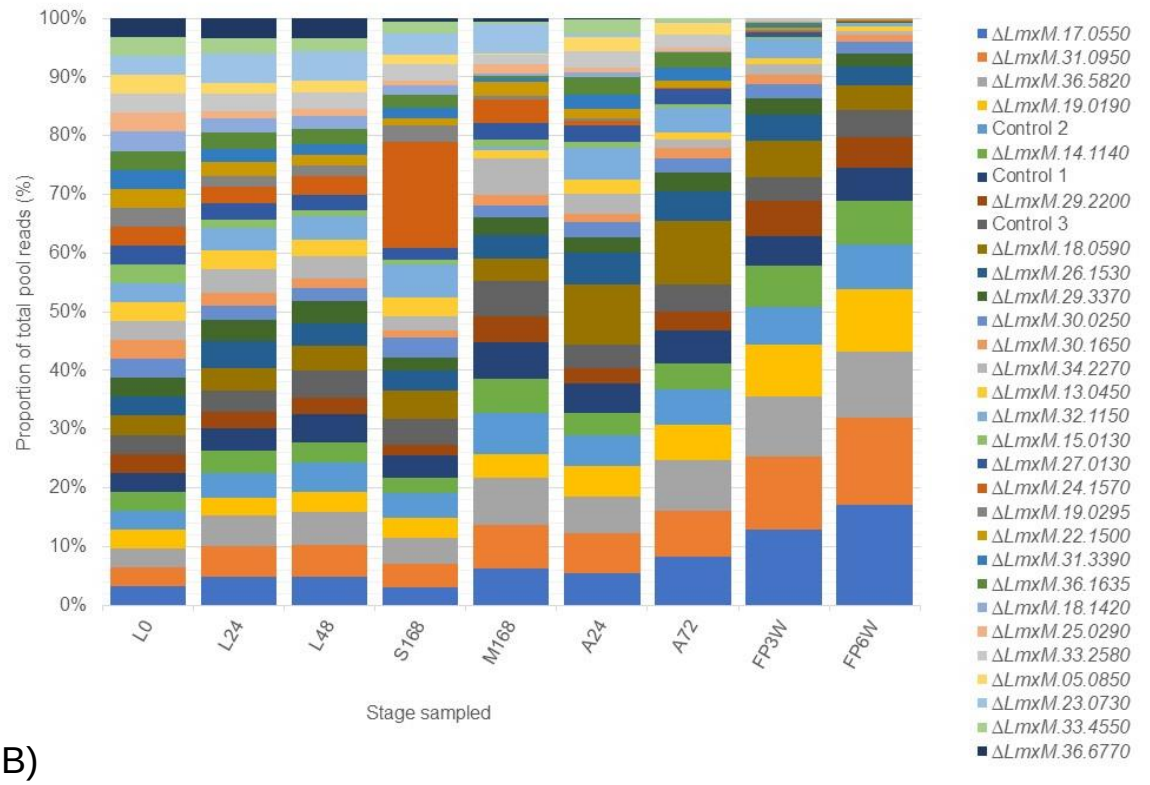


Figure 4.14. Barcode amplification by PCR. A) DNA extracted from all stages of the bar-seq screen was used as the template for PCRs to amplify the barcode region from all null mutant lines. Labels indicate experimental stages from Figure 4.1. Initially 25 PCR cycles were used for all experimental samples (L0 = log phase 0h post inoculation, S168 = stationary phase 168h post inoculation, M168 = metacyclic purification). A control sample (+ve) containing the same template quantity was included in each set of PCR reactions with 35 cycles to show the result of the PCR running to completion. B) PCRs were repeated with increased and normalised template concentrations and run for 30 cycles (A24_1 to A72_6), 28 cycles (FP3W_1 to FP6W_6) or 25 cycles S168_1 to M168_6) with a 40-cycle positive control.

A)



B)

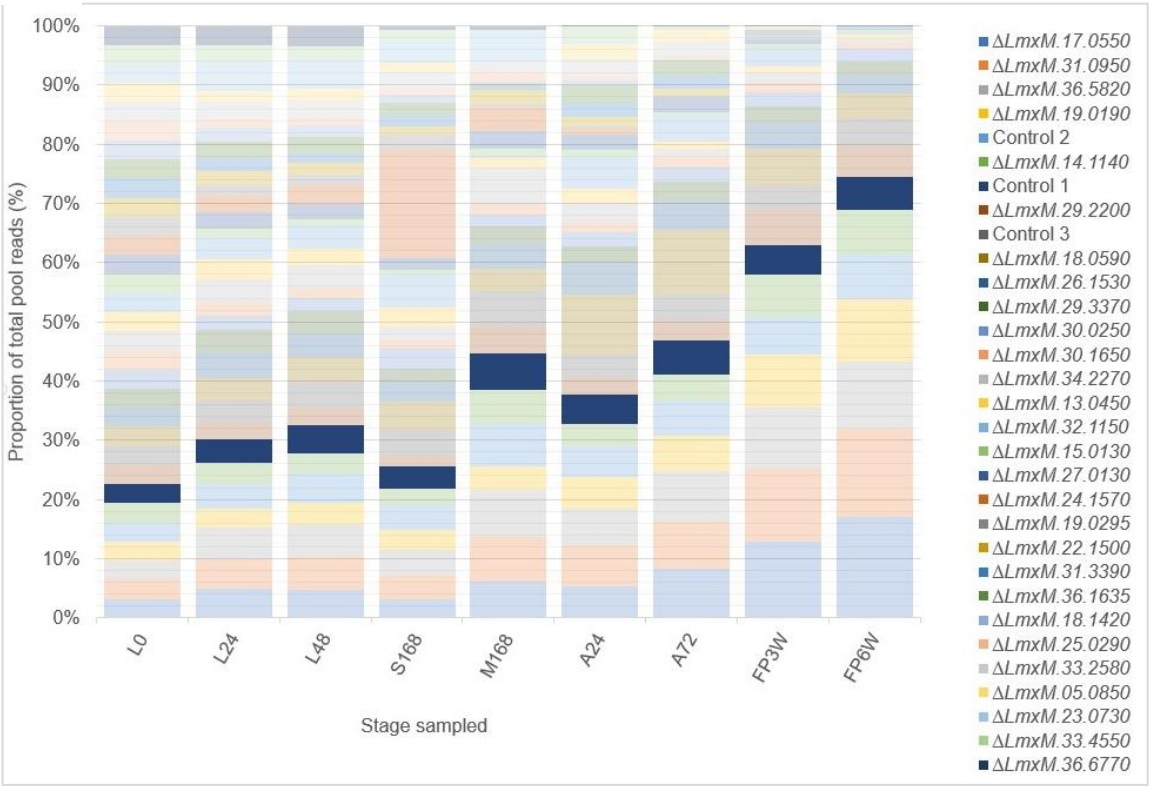


Figure 4.16. Bar charts showing pool composition over time. A) The raw sequencing reads were normalised based on the day 0 counts of all cell lines and then expressed as a percentage of the total reads for each stage sampled. Cell lines are coloured individually and arranged vertically from the least represented (top) to most represented (bottom) in the last stage of the screen. B) Individual tracks can be highlighted as shown with the filter function. For example, control cell line one is highlighted in dark blue and, like the other two controls, maintains its fitness throughout the screen.

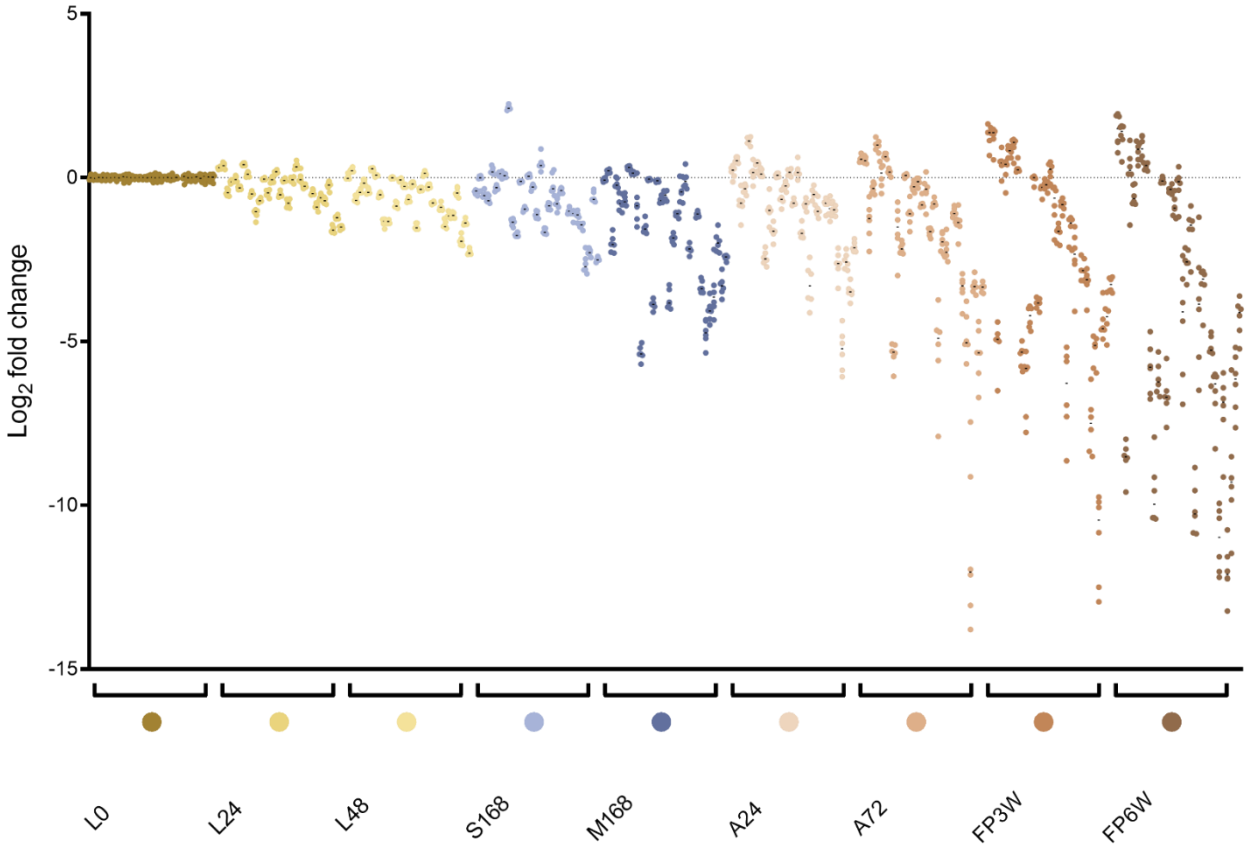


Figure 4.17. An overview of sample variation in the RBP knockout screen. The log_2 fold change in normalised barcode reads compared to the mean of the three control cell lines was plotted for each cell line at each stage of the barcoded knockout screen. Dots represent reads from each of the six replicate pools (5 replicates for Stat 7 dpi). Yellow/brown stages represent log phase promastigotes, blue represent stationary phase promastigotes and pink/red represent amastigotes. L0 = log phase 0h post-pooling, L24 = log phase 24h post-pooling, L48 = log phase 48h post-pooling, S168 = stationary phase 168h (7days) post-pooling, M168 = metacyclic purification 168h (7days) post pooling, A24 = amastigotes from bone marrow derived macrophages 24h post-infection, A72 = amastigotes from bone marrow derived macrophages 72h post-infection, FP3W = amastigotes from mouse footpads three weeks post-infection, FP6W = amastigotes from mouse footpads six weeks post-infection.

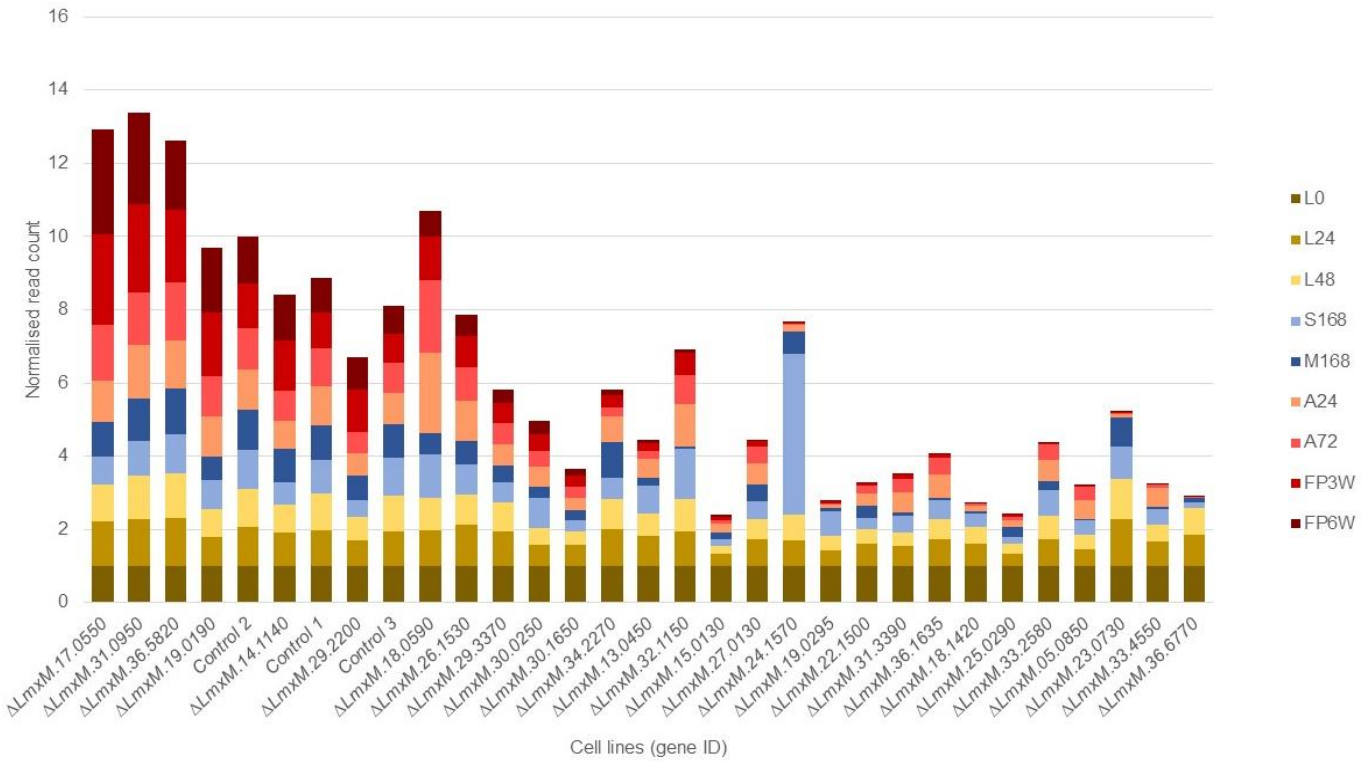


Figure 4.18. Alternative bar chart of relative null mutant fitness. A different presentation of the data from the RBP bar-seq screen is useful for assessing individual cell lines. Normalised read counts (by day 0 for each cell line and to the total sequenced reads in each stage) were compared for each of the 31 cell lines included in the knockout screen. Yellow/brown stages represent log phase promastigotes, blue represent stationary phase promastigotes and purified metacyclic promastigotes, and pink/red represent amastigotes. Null mutant lines are ordered on the X-axis by highest to lowest read count in the final stage of the experiment (FP6W). As expected, the three control cell lines are found to the left of the graph with high read counts after six week of mouse footpad infection. L0 = log phase 0h post-pooling, L24 = log phase 24h post-pooling, L48 = log phase 48h post-pooling, S168 = stationary phase 168h (7days) post-pooling, M168 = metacyclic purification 168h (7days) post pooling, A24 = amastigotes from bone marrow derived macrophages 24h post-infection, A72 = amastigotes from bone marrow derived macrophages 72h post-infection, FP3W = amastigotes from mouse footpads three weeks post-infection, FP6W = amastigotes from mouse footpads six weeks post-infection.

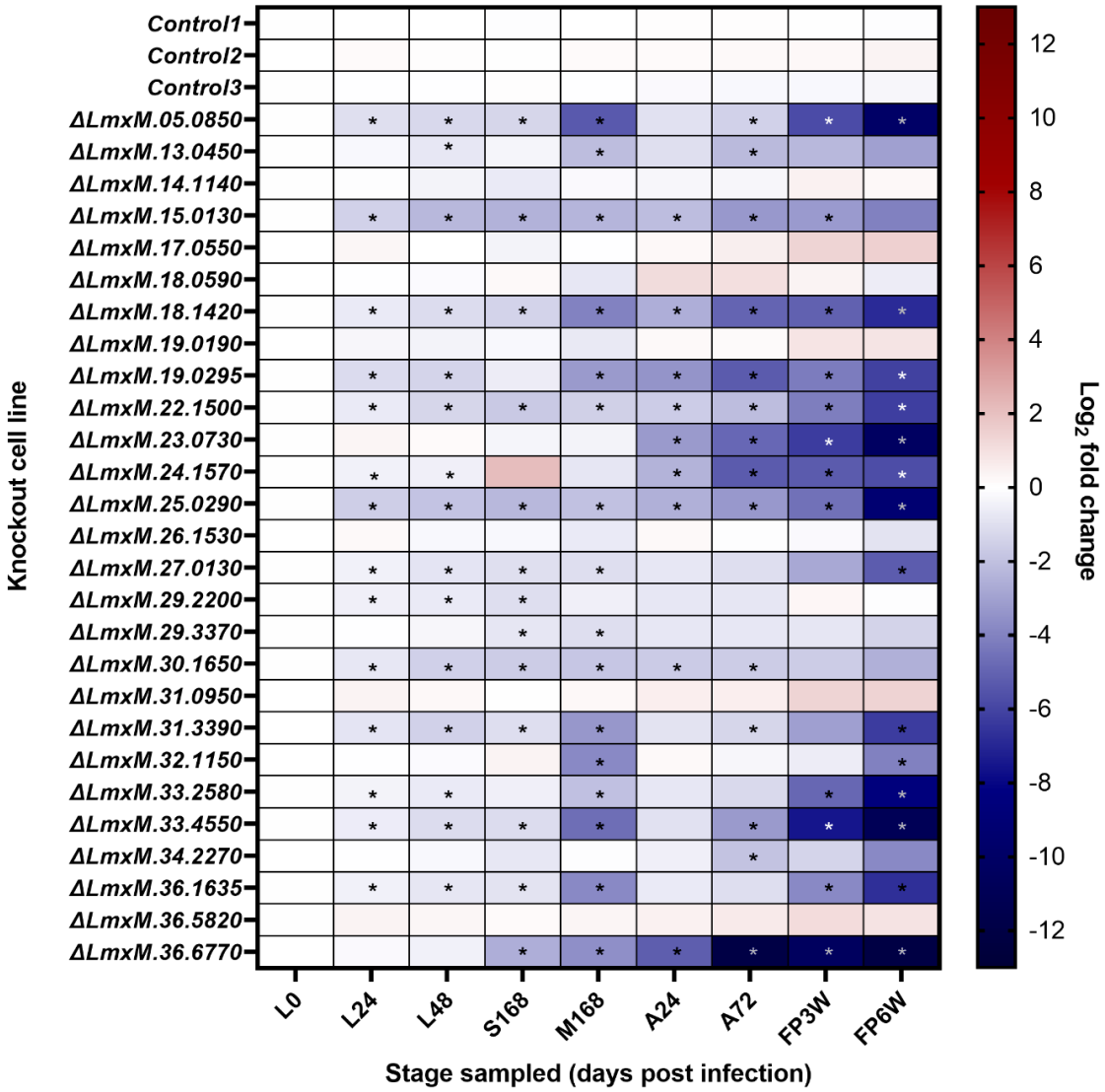
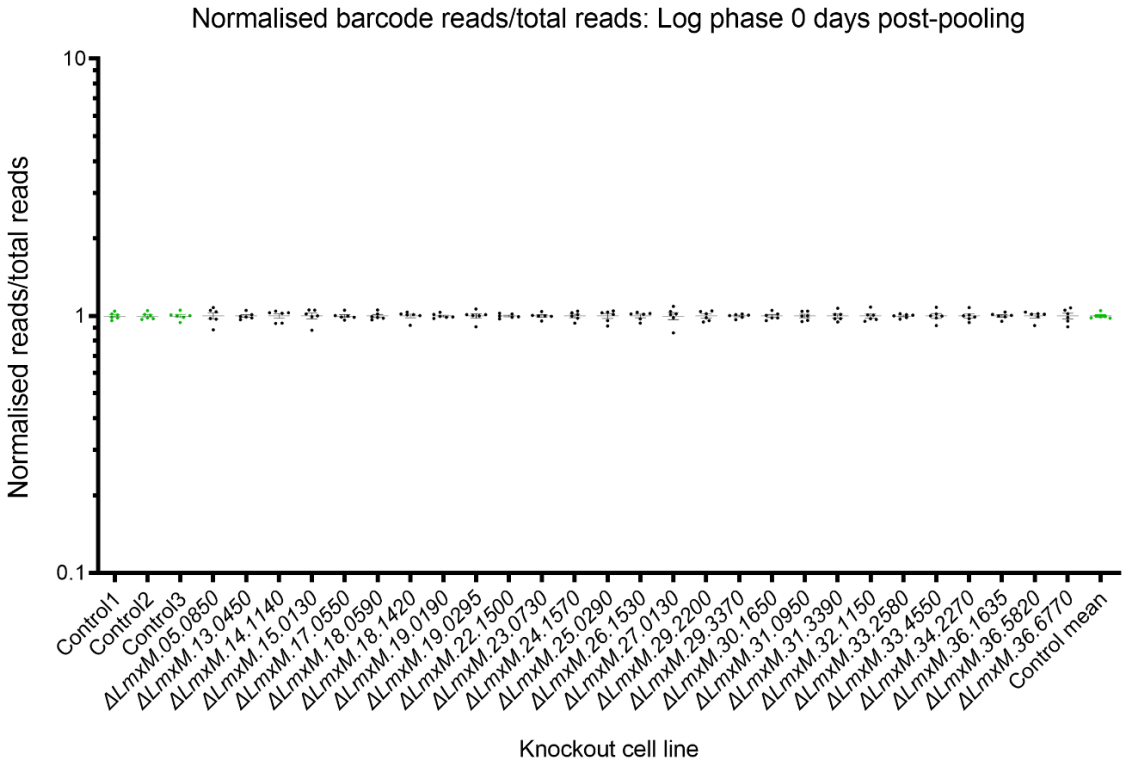
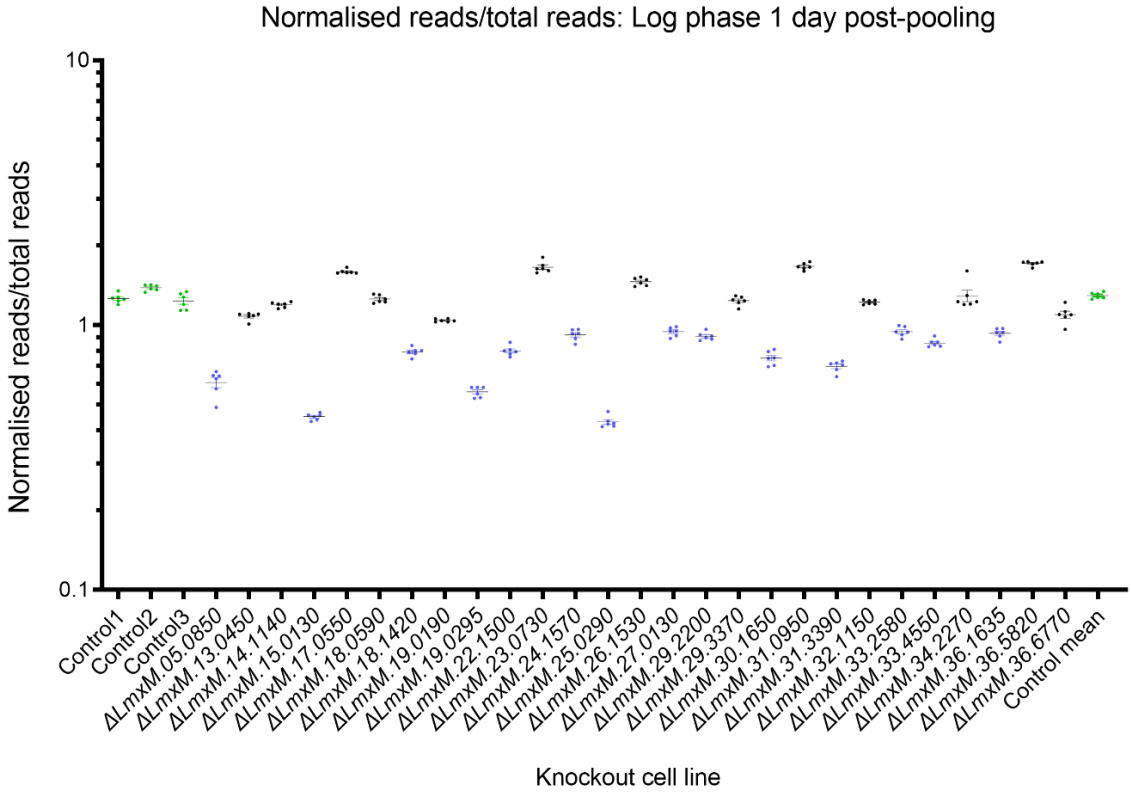
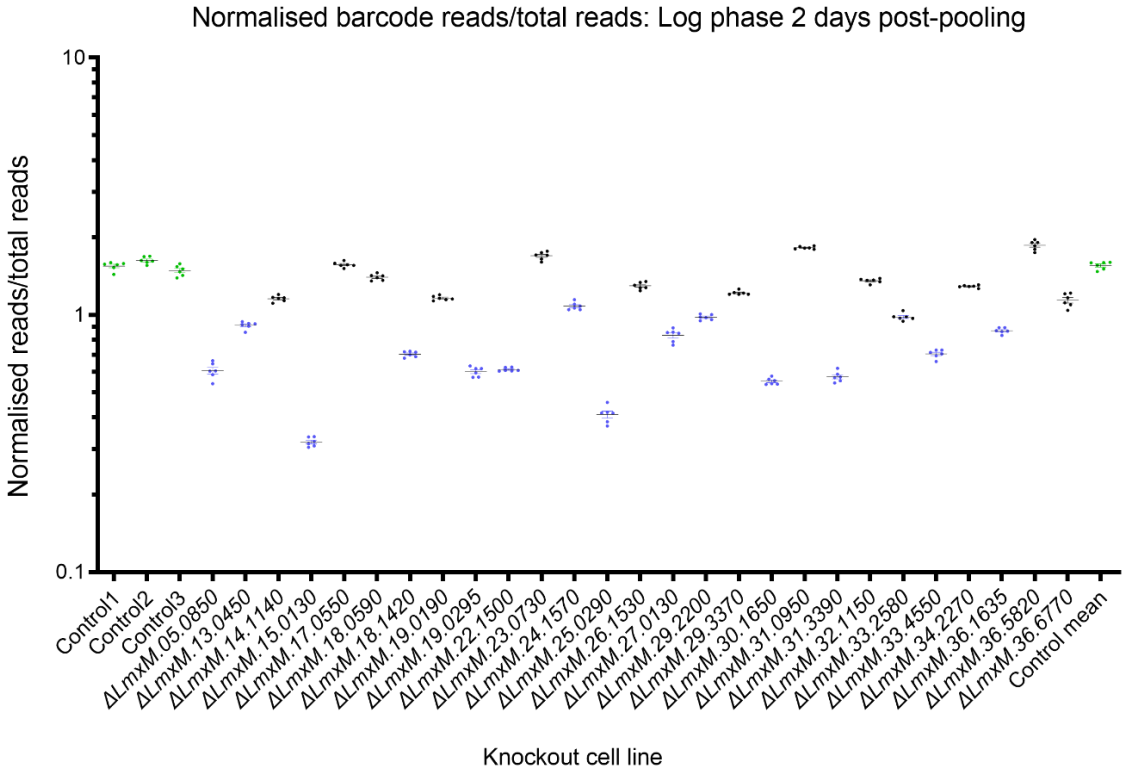
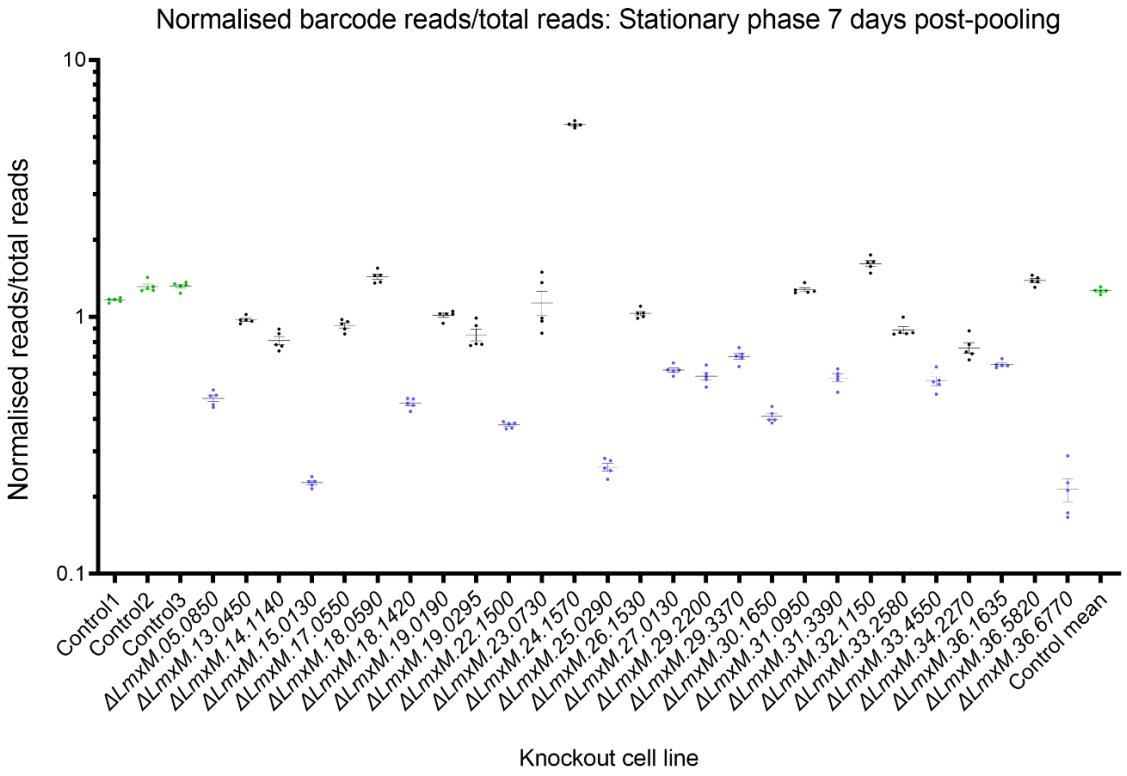


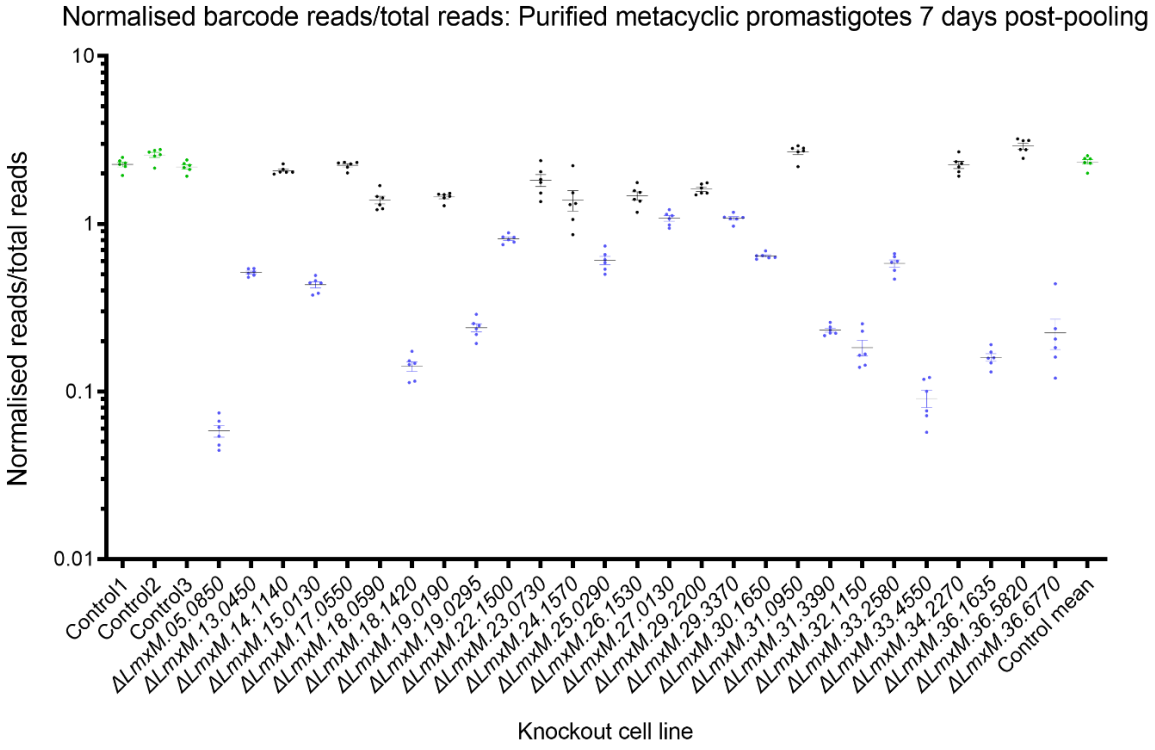
Figure 4.19. Heatmap showing comparisons of RBP null mutant growth and infectivity to aphenotypic controls. The log_2 fold change in normalised barcode reads (by day 0 for each cell line and to the total sequenced reads in each stage) were plotted as a heatmap using Prizm 9.0 and R-studio. Blue and red represent negative and positive log_2 fold change respectively when comparing normalised barcode reads to the mean of the three control cell lines Asterisks mark cell lines that had significantly different medians when normalised read counts were compared to the control cell lines using a Kruskal-Wallace test followed by the two-stage step-up method of Benjamini, Krieger and Yekutieli to correct for the false discovery rate. Statistics carried out using Prizm 9.0 were comparisons between normalised read counts of all cell lines in a given stage with the mean of all three control cell lines (n = 6). A p-value of <0.05 was considered statistically significant.

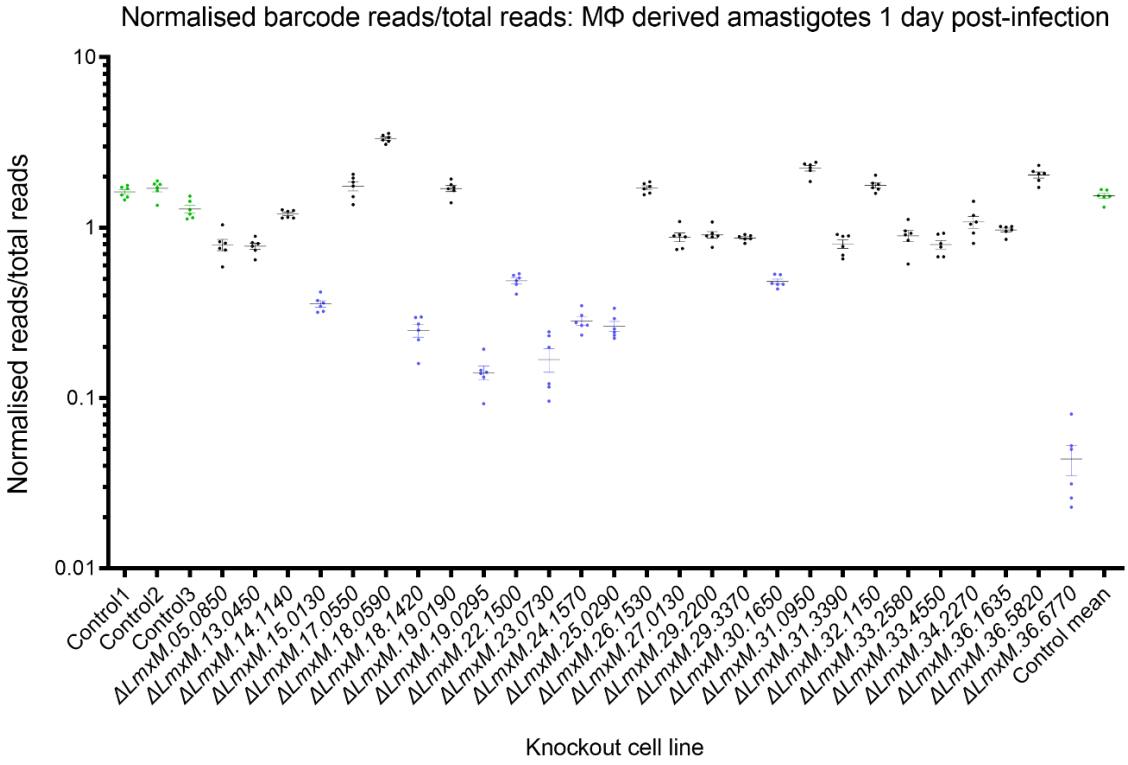

Normalised barcode reads/total reads: Log phase 0 days post-pooling
Normalised reads/total reads
10
1
0.1
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line

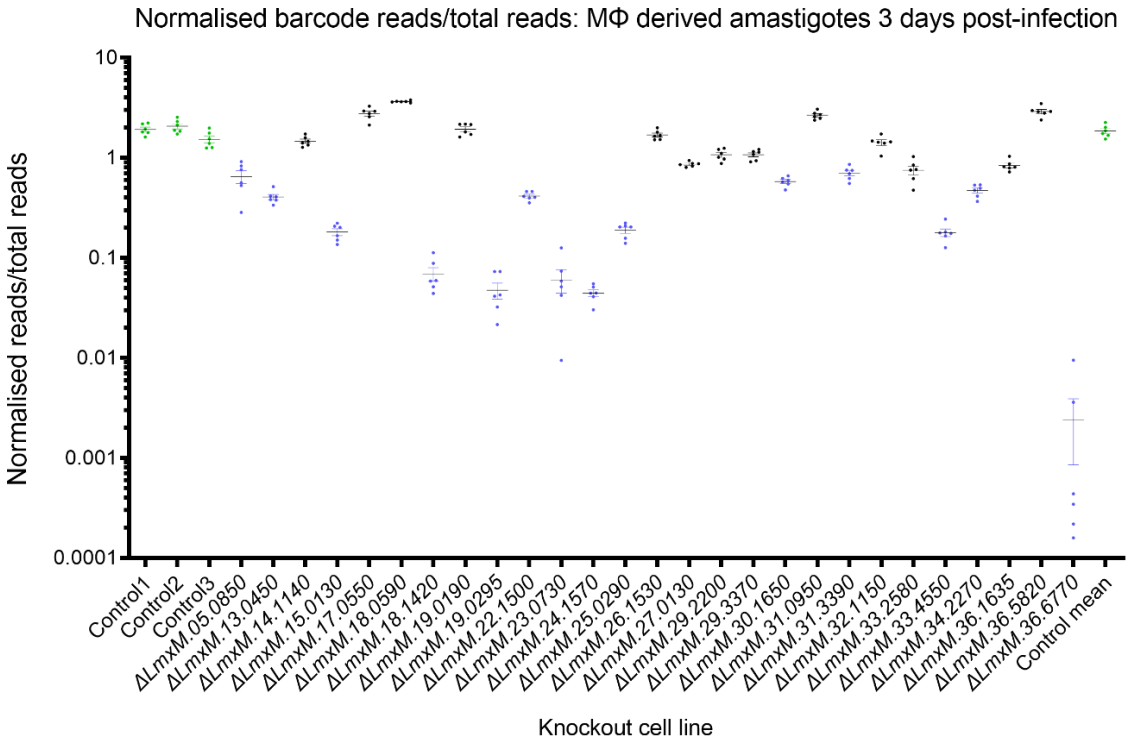

Normalised reads/total reads: Log phase 1 day post-pooling
Normalised reads/total reads
10
1
0.1
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line

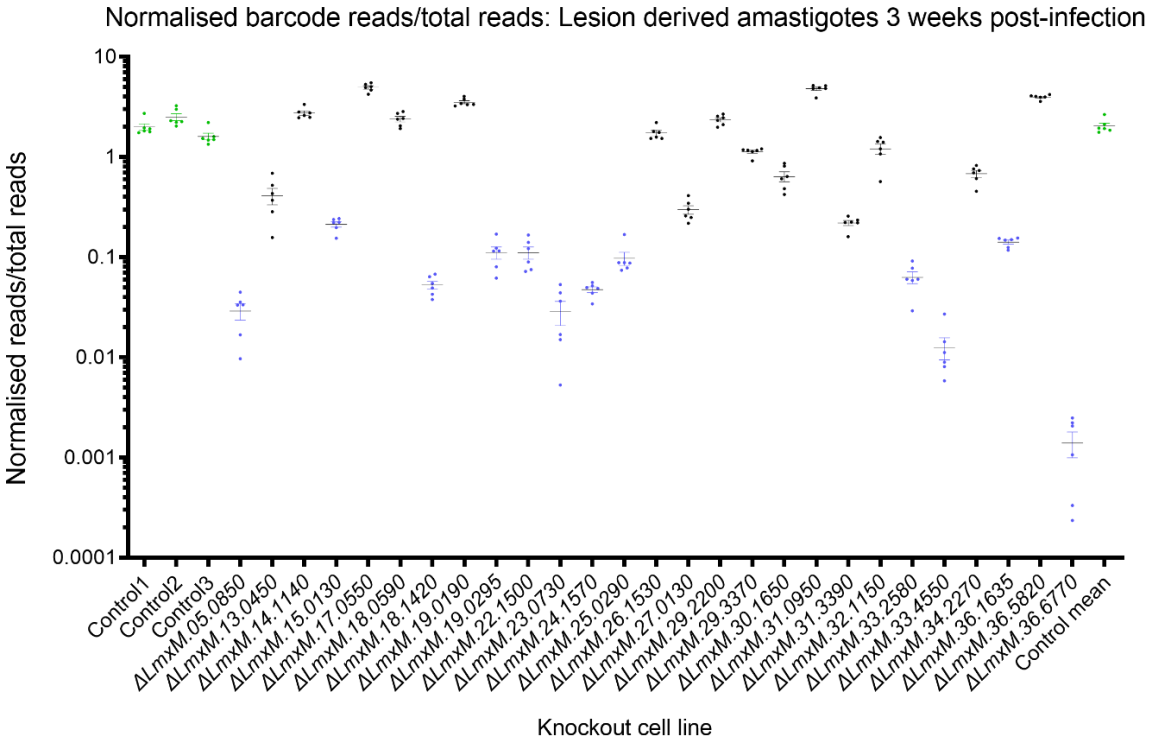

Normalised barcode reads/total reads: Log phase 2 days post-pooling
Normalised reads/total reads
10
1
0.1
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line


Normalised barcode reads/total reads: Stationary phase 7 days post-pooling
Normalised reads/total reads
10
1
0.1
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line


Normalised barcode reads/total reads: Purified metacyclic promastigotes 7 days post-pooling
Normalised reads/total reads
10
1
0.1
0.01
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line


Normalised barcode reads/total reads: MΦ derived amastigotes 1 day post-infection
Normalised reads/total reads
10
1
0.1
0.01
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line


Normalised barcode reads/total reads: MΦ derived amastigotes 3 days post-infection
Normalised reads/total reads
10
1
0.1
0.01
0.001
0.0001
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line


Normalised barcode reads/total reads: Lesion derived amastigotes 3 weeks post-infection
Normalised reads/total reads
10
1
0.1
0.01
0.001
0.0001
Control1
Control2
Control3
ΔLmxM.05.0850
ΔLmxM.13.0450
ΔLmxM.14.1140
ΔLmxM.15.0130
ΔLmxM.17.0550
ΔLmxM.18.0590
ΔLmxM.18.1420
ΔLmxM.19.0190
ΔLmxM.19.0295
ΔLmxM.22.1500
ΔLmxM.23.0730
ΔLmxM.24.1570
ΔLmxM.25.0290
ΔLmxM.26.1530
ΔLmxM.27.0130
ΔLmxM.29.2200
ΔLmxM.29.3370
ΔLmxM.30.1650
ΔLmxM.31.0950
ΔLmxM.31.3390
ΔLmxM.32.1150
ΔLmxM.33.2580
ΔLmxM.33.4550
ΔLmxM.34.2270
ΔLmxM.36.1635
ΔLmxM.36.5820
ΔLmxM.36.6770
Control mean
Knockout cell line

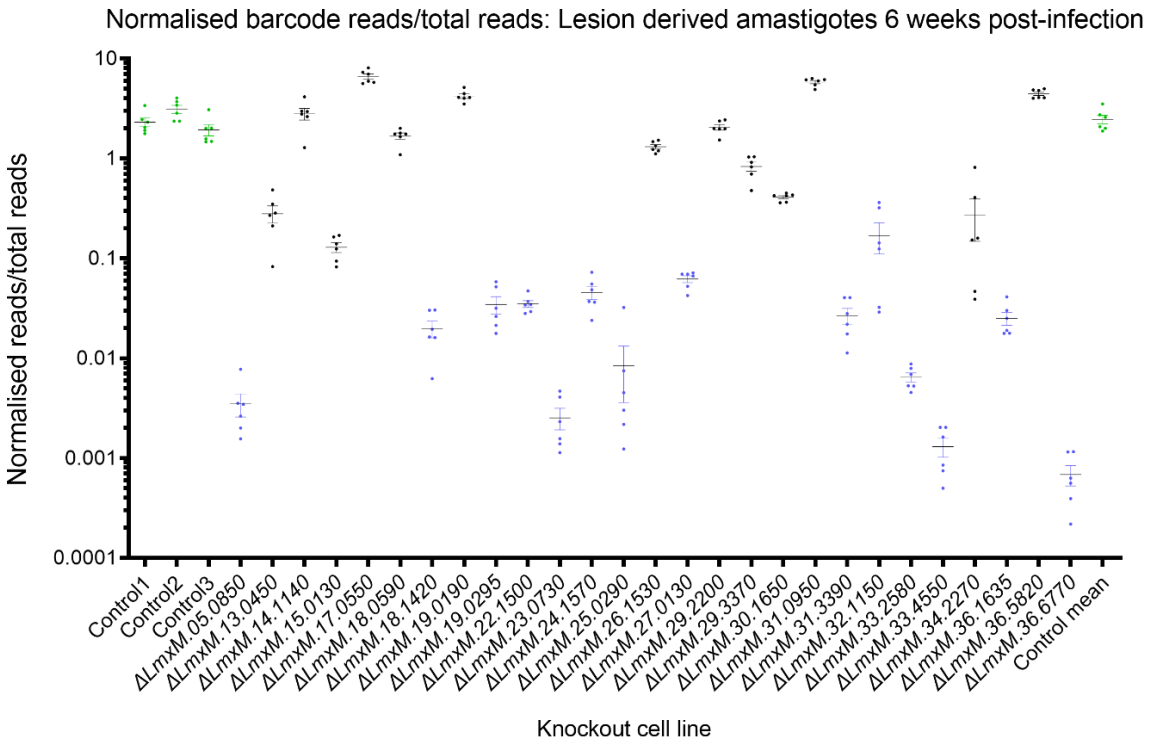


Figure 4.20. Stage specific comparisons of normalised barcode reads. A visual representation of the statistical comparisons made in Figure 4.6.4 here showing the individual data points. The three control cell lines are highlighted in green as well as the control mean at the far right. Note the control mean has six replicates, one mean of control cell lines one, two and three for each of the six experimental replicate pools. Cell lines with significantly different median scores from the Kruskal-Wallace test and two-stage step-up correction are highlighted in blue. Data were plotted in Prizm 9.0 using a $\log_{10}$ scale to see the full extent of the data.

4.10 Chapter 4 discussion

4.10.1 Bar-seq screens in *Leishmania*

The introduction of the CRISPR/Cas9 system to *Leishmania* greatly expanded the possibilities for knockout screens. Before the introduction of CRISPR/Cas9, gene deletion mutants required the cloning of bespoke knockout vectors flanked by at least 500bp of homology to the region of interest. Knockout donor DNA was amplified from the vector template and then nucleofected into the parasites where homologous recombination would replace the gene of interest. The cells where this relatively rare event had taken place could then be selected for using drugs. While many successful null mutants have been produced this way, scaling this up to produce enough knockouts to make a pooled screen worthwhile was overly time consuming. Between 1990 and 2017, around 65 gene knockouts had been attempted (Jones *et al.*, 2018). By contrast, the CRISPR method designed by Beneke *et al.*, (2017) facilitates the rapid production of null mutants by increasing the efficiency of genome editing and streamlining the workflow for producing donor DNA. Because of the accuracy and reliability of the Cas9 enzymes, gene specific homology regions were reduced to 30bp, short enough to be incorporated into PCR primers which were used to generate a suitable donor repair cassette. This allows for a single plasmid template to be used for the amplification of all knockout donor DNA, with the only gene specific regions included in the primers, thus bypassing time consuming plasmid cloning steps. Additionally, DNA barcodes can be included in the upstream forward primer and integrated into a sequencing workflow that allows for pooled screens of many individual null mutants.

The CRISPR technology is especially useful in *L. mexicana* as this species lacks the RNAi pathway that has been used extensively in *T. brucei* and other organisms to knock down proteins of interest (Lye *et al.*, 2010). Recently, the large scale, barcoded CRISPR screen approach has been applied to several groups of genes in *Leishmania* including protein kinases, flagellar proteins and de-ubiquitinases (Beneke *et al.*, 2019; Damianou *et al.*, 2020; Baker *et al.*, 2021). As yet, RBPs have been mostly studied individually in *Leishmania* or as interactors in regulatory complexes. Using the CRISPR/Cas9 system, the aim of this study was to further the functional and phenotypic understanding of *L. mexicana* RBPs using a barcoded knockout screen.

Knockout cell lines were cloned by dilution in 96 well plates prior to being pooled for the bar-seq screen. An alternative method, the selection of mixed populations in flasks has some advantages. Because it represents multiple integration events of the donor DNA, the mixed population method is more representative and may produce more reliable phenotypes. The effects of compensatory mutations and aberrant chromosomal copy number on barcode reads are dramatically reduced when using a mixed population as the resulting phenotype is an average of many cells. All three screens of this type published in *Leishmania* to date used mixed populations. However, one of the problems with using a mixed population is that parasites with a fitness advantage resulting from remaining CDS copies, if present in the population, can outcompete those with a true gene deletion. When

screening for knockout clones, the simultaneous integration of two different drug resistance cassettes guarantees a minimum of two gene copies being removed. This is important because the screening for homozygous clones increases the workload when working on many knockout cell lines simultaneously. There is a cost to efficiency when using double drug selection as parasites with multiple integrations of the same drug resistance cassette will be selected against with the alternative drug. Despite using double drug selection, *Leishmania* parasites have a remarkable ability to retain a functional copy of the gene. This is likely due to their genomic plasticity, with fluctuations between a base state and aneuploidy being common (Ubeda *et al.*, 2008; Rogers *et al.*, 2011; Sterkers *et al.*, 2011; Barja *et al.*, 2017; Iantorno *et al.*, 2017; Reis-Cunha, Valdivia and Bartholomeu, 2017; Bussotti *et al.*, 2018; Damasceno *et al.*, 2020; Negreira *et al.*, 2021). Unfortunately, retention of a functional copy of the gene is often likely to confer a fitness advantage over knockout cells, explaining the ability of mixed populations to lose their knockout phenotype over time. For these reasons, only clonal populations where removal of the CDS had been confirmed by PCR were included in the pooled screen. Due to the seven-week duration of the bar-seq screen, it was crucial that parasites with a fitness advantage were not inadvertently added to the pool as this would have reduced the proportion of other cell lines in the pool and distorted the data. However, this approach is not currently practical for studies with larger library sizes such as those that have already been published due to the time-consuming nature of screening for positive clones.

4.10.2 PCR confirmation of RBP null mutant lines

Initial tests using single drug selection to produce null mutants of the DEAD-box helicase *LmxM.15.0130* were successful in identifying an RBP with a clear knockout phenotype; a severe promastigote fitness defect. Two independent independent null mutants with a similar phenotype suggests that is a result of *LmxM.15.0130* removal. Throughout this thesis ‘clonal’ has been used to differentiate cell lines selected after dilution in liquid on 96-well plates from the heterogeneous populations derived directly from transfection. However, because of the dilution method and selection on liquid plates, there is a possibility that more than one cell was placed in a single well. Also, due to a short recovery time, it is possible that multiple cell lines are derived from a single integration event but are genetically identical. For the purposes of reducing the chance of cell lines retaining gene copies, both of the above are improvements on a mixed population representing many integration events. With more resources and time, an optimal solution would be to produce clones on solid agar/M199 plates, which appears to more reliably produce clonal populations. Multiple clones for each cell line could then be pooled to make the data more representative. However, this method would increase the workload by several times and has not yet been attempted in *Leishmania* or *T. brucei* for a large-scale screen.

During the production of the RBP knockout library, PCRs were used exclusively to determine which clones were to be included in the screen. Overall this has been successful with all of the 13 clones later sent for whole genome sequencing showing expected genome engineering with removal of the

CDS. However, it should be noted that the PCRs described (Figure 4.4) could be improved if larger scale screens were attempted. The lack of a PCR band for the CDS was used to indicate knockout, which could in rare cases be due to PCR failure rather than CDS removal and lack of a binding site. To eliminate this problem and avoid false negative results, band-shift PCRs with primers that bind either side of the integrated region could be used to detect the difference between the wild type and integrated donor DNA, giving positive results for both outcomes. In practice, the larger genes often failed to amplify when using this approach and several genes produced bands of a similar size to the integrated DNA making results hard to interpret. Returning to the method used in this study, the PCR to confirm integration of donor DNA confirms the quality of the template is sufficient and the positive band amplified from the JM6571 (T7/Cas9) control DNA template, confirms that both the PCR conditions and primers are functional. This, combined with the DNA sequencing results was considered sufficient evidence that the RBPs had been successfully deleted.

The knockout of *LmxM.15.0130*, using only Blasticidin selection, demonstrated that single drug selection, in combination with the CRISPR system, is a viable method for producing null mutants if multiple clones can be subsequently screened by PCR to confirm CDS removal. Single drug selection also allows other drugs such as Puromycin to be used in downstream genomic manipulations.

4.10.3 RBPs as essential regulatory proteins

The production and screening of knockout clones represented a major part of the experimental portion of this thesis. RBPs were selected for knockout in order of priority based on the criteria and data described in Chapter 3. The initial PCR results were not clear enough to determine which clones were null mutants. The subsequent analysis showed that this was most likely due to the quality of the extracted DNA. Cheaply available commercial Taq-polymerase mix was tested, and, while time efficient, reduced the reliability of the diagnostic PCR results.

Once diagnostic PCRs had been optimised, the proportion of successful and unsuccessful knockouts could be determined. The large number of unsuccessful knockouts can be split into two groups: those that returned no clones and those that returned clones still containing a CDS copy. While neither of these groups could be included in the barcoded knockout screen, they represent a striking result with much potential for further study. The inability to produce viable parasites suggests that many of these genes are essential in the promastigote stages. Unfortunately, the time constraints of the study did not allow for three or more repeat transfections per cell line which would be required to clarify that the lack of surviving clones was a repeatable result. However, of the transfections that were repeated, those that produced no viable clones or only produced clones with a remaining CDS copy often produced the same result in subsequent transfection attempts. Further investigation of these RBPs that could not be deleted should be carried out to identify truly essential genes. Essentially, multiple different guide RNAs should be designed to confirm that poor CRISPR targeting is not the cause of knockout failures. An inducible CRISPR system would also be a useful tool for differentiating

essential RBPs from methodological failures. Other methods currently available are discussed further in section 6.

The concept of RNA-binding proteins as essential regulators of gene expression has been well documented; for example, in human cancer cell lines, human embryonic stem cells, *C. elegans* and several times in kinetoplastids (Ammerman *et al.*, 2010; Serpeloni *et al.*, 2011; Kwon *et al.*, 2013; Fei *et al.*, 2017; Norris, Gracida and Calarco, 2017; Wang *et al.*, 2019; Mishra *et al.*, 2021). In recent years, the idea of post-transcriptional *trans*-regulators as being a secondary mode of control or fine tuning has been largely abandoned in favour of viewing them as equal and complimentary to transcriptional regulators (Beckmann *et al.*, 2015; Hentze *et al.*, 2018; Van Nostrand, Freese, *et al.*, 2020). Many RBPs have been described as the primary regulators of genes involved in essential processes. In *Leishmania* the essentiality of RBPs is likely accentuated due to the lack of transcriptional regulation. Polycistronic gene expression means that gene specific transcriptional control is not viable. Consequently, it is not surprising that many RBPs could be essential in *Leishmania* as many transcription factors are in other eukaryotic organisms where transcription factors are universally involved in the regulation of essential cell processes (Bertomeu *et al.*, 2018; Yilmaz *et al.*, 2018; Mair *et al.*, 2019). Moreover, it is possible for some factors to display essentiality in a stage specific manner as RBPs in other organisms are often tightly regulated. Considering the majority of RBPs are most associated with mRNA in the procyclic stage (Figure 3.4), essentiality in procyclic promastigotes is the most likely explanation for the majority of unsuccessful knockouts. Several of the RBPs that produced no viable null mutants in this screen had been studied previously in kinetoplastids and found to be essential: DRBD3 (LmxM.04.1170), DRBD7 (LmxM.33.4560), RBP28 (LmxM.25.1080), RBP3 (LmxM.25.0520), TRRM1 (LmxM.27.2100), PUF7 (LmxM.31.1750), RGG2 (LmxM.32.0260) (Estévez, 2008; Fisk *et al.*, 2008; Wurst *et al.*, 2009; Droll *et al.*, 2010; Levy *et al.*, 2015). Two proteins that had been previously described as essential were found amongst RBP lines that only ever presented clones with an extra CDS copy: RBP16 and the helicase SUB2 (Pelletier and Read, 2003; Serpeloni *et al.*, 2011; McDonald *et al.*, 2018). Essentiality in procyclic promastigotes is not necessarily mutually exclusive with essentiality in other stages. The proportion of essential RBPs in *Leishmania* based off this screen could be considered between 34-58% considering either just the failed knockout attempts or all knockout attempts that only produced clones with a remaining CDS in addition to those that failed. This is a higher rate of knockout faliures than seen in either *L. mexicana* de-ubiquitinases (20%) or kinases (21%) possibly indicating higher essentiality amongst RBPs (Damianou *et al.*, 2020; Baker *et al.*, 2021). Comparing more broadly to CRISPR knockout screens in other organisms is challenging due to different methodologies and screening conditions. A genome wide screen in *Toxoplasma gondii* described ~40% of protein coding genes targeted had a significant contribution to fitness (Sidik *et al.*, 2016). An even larger knockout screen in a range of human cell types covered multiple splice variants and predicted genes as well as characterised genes making comparison even more difficult. However, by considering only genes characterised in the RefSeq database (NCBI) that were essential across all human cell types they tested, 2.75% were essential. It is not surprising that a greater proportion of

genes are essential in single celled parasites than higher eukaryotes, where many genes are tissue or cell type specific and may not be required in most cells.

The high occurrence of cell lines with a remaining CDS but surviving selection in both drugs may be surprising, but appears to be a common phenomenon in *Leishmania.* These parasites are known to have a highly plastic genome, capable of tolerating high levels of mosaic aneuploidy and even expressing some genes on episomes (White *et al.*, 1988; Sterkers *et al.*, 2011; Barja *et al.*, 2017; Reis-Cunha, Valdivia and Bartholomeu, 2017; Damasceno *et al.*, 2020; Negreira *et al.*, 2021). Under extreme pressure provided by drug selection, any cells that can retain a copy of the gene in addition to integrating the resistance cassettes will be greatly favoured, even more so if gene removal is deleterious. The whole genome sequencing of knockout clones shed more light on the number of copies of integrated donor DNA cassettes (section 5.2). One explanation that can be ruled out is that the drug concentration used was too low for selection, as un-transfected controls were entirely killed by the drug combination at the time clones were selected.

4.10.4 A bottleneck for morphological mutants in the bar-seq screen

Having detected the abnormal morphology of cell line Δ*LmxM.05.0850* during the production of knockout cell lines, it was important to test if the metacyclic promastigote purification presented a bottleneck that excludes cell lines with morphological phenotypes. Centrifugation of parasites through a gradient of ficoll or percoll has been designed to purify metacyclic (infective) promastigotes based on their smaller, lighter cell bodies (Späth and Beverley, 2001). This is an effective method and removes the need for bespoke antibodies or lectins that were previously required to isolate this lifecycle stage. However, when parasites have abnormal morphologies, this centrifugation step could result in their removal from the screen. Although this would detect the abnormalities, removal of the parasites at this stage means that they are not properly challenged with macrophage infection in either of the mammalian models of infection used. To allow all cell lines to be screened in the human infectious stages, stationary phase culture (day seven) was used for both the infection of bone marrow-derived macrophages and mouse footpads. Metacyclic promastigote purifications of day seven culture were still sequenced to obtain data on potential metacyclic stage specific defects resulting from RBP deletion. As expected, Δ*LmxM.05.0850* were largely removed during the metacyclic stage purification but were not severely unfit during the macrophage infections. It has been reported that, despite metacyclic being the infectious stage, heterogeneous populations are transmitted when sand-flies bite the mammalian host (Giraud *et al.*, 2019) suggesting there may be biological relevance to using stationary phase culture for infections as opposed to purified metacyclic promastigotes.

4.10.5 Choice of macrophage lines and pooling methodology

Due to the progressive nature of the RBP bar-seq screen, the pooling of promastigotes was critical for the success of downstream experimental stages. Knockout cell lines were thawed and passaged

sequentially depending upon their growth characteristics. Some cell lines (e.g. Δ*LmxM.15.0130*) take several days longer to reach a suitable number of promastigotes for pooling. Other cell lines would have differentiated to metacyclic promastigotes if thawed at the same time. When all cell lines were still in the procyclic stage and had reached a high enough density, they were pooled independently six times. Pooling them independently means that the error involved in pipetting, counting and pooling cells is accounted for in the replicates and so in the statistical comparisons with the control cell lines. Each replicate pool was matched with an individual mouse and an individual macrophage culture at day seven when stationary culture from the pools was used to inoculate mice and macrophage culture. This allowed each replicate of each cell line to be tracked as a fitness trajectory across the duration of the experiment.

In the future, it is likely that pooled, barcoded knockout experiments will be performed at much higher throughput than the current screen. However, the time limiting steps of screening for positive clones and of assembling the pools in a short space of time, will still need to be tackled. Pooling of parasites is particularly difficult with large numbers of cell lines because if the process takes many hours, some cell lines will reach the stationary phase. Because of the number of cells required, a relatively high concentration (>2×10^6/ml and less than1×10^7/ml) is needed when setting up the pools. Stationary phase parasites are much less replicative than those in log phase which would disadvantage some cell lines compared to those pooled earlier.

Macrophage infections were also optimised to ensure protocols used were suitable for the cell lines and conditions of the knockout screen. The amplification of *L. mexicana* DNA from infected macrophages is considerably harder than from *Leishmania* cell culture. This is mostly because of the smaller number of parasites available and the large quantities of macrophage DNA that is extracted along with the parasite DNA. Because of the small quantities of available DNA, optimisation of macrophage infection was chosen as the first step. At low infection ratios, purified metacylic stages appeared to be easier to amplify, perhaps suggesting a more established infection. However, amplification from stationary culture infections was also detectable. The decision to use a 6:1 ratio of stationary culture in the bar-seq screen was based on a trade-off between having enough parasites to be easily amplified by PCR but not so many that there are extracellular amastigotes. The large quantities of extracellular amastigotes seen after an 8:1 infection could be due to the differentiation of extracellular promastigotes that were not washed off the surface or due to release of amastigotes from heavily infected macrophages. It was also important to keep the infection ratio and number of macrophages high enough that extracted parasites were representative of the pool. Below a certain threshold, the error in the experiment would increase due to few parasites of each cell line being exposed to the macrophages. At very high infection ratios many promastigotes remain attached to macrophages externally, even after extensive media washes, potentially distorting the results of the infectivity testing. Cells that are external to macrophages may not even be viable amastigotes but would be sequenced when macrophages are harvested for DNA extraction. Extensive washes and an infection ratio of 6:1 resolved this problem as seen in Figure 4.10.

The type of macrophages used was also important to consider. Bone marrow macrophages were chosen because of their higher availability for infection, despite being more variable and difficult to prepare and culture. Being primary cells from the same mouse species also means that the results are more comparable than immortalised cell lines such as J774 (Ralph and Nakoinz, 1975).

4.10.6 Barcode amplification and sequencing

The amplification of the barcoded regions from the RBP knockout pool DNA is a crucial step in this experiment. The absolute amount of DNA produced by each amplification does not have to be identical as the barcode reads can be normalised as a proportion of the total reads for each stage. However dramatically different amounts can affect the read depth so, where possible, template quantities were normalised before amplification. As with qPCR it may be important to measure amplified barcodes only to the point that the PCR reaction is in the linear stage. It has been suggested that the quantitative relationship between samples may be distorted as the plateau phase of amplification is reached. However, several studies have also shown that this is not necessarily the case and that it is possible to quantify differences between reactions that have progressed into the plateau phase (Morrison and Gannon, 1994; Jansson and Hedman, 2019). In the interest of caution, samples for this screen were not amplified past a maximally amplified control.

Contamination with non-experimental barcodes represents a much greater problem. Due to the small quantities of DNA being extracted, especially for the amastigote stages, any contaminant amplicons could distort the results significantly. Measures taken to eliminate this included the decontamination of all equipment and working in a separate lab to avoid the constructs and primers that had been used for this and other similar experiments. Additionally, an alternative reverse primer was used to reduce the chance of binding to contaminating constructs or primers. However, this primer only binds to the *BSD* donor cassette which must be considered in the final results. When using polyclonal populations, amplifying barcodes from the *BSD* cassette alone would give a representative result as long as multiple integrations aren't common. When using clonal populations, the number of integrated copies has a large impact on the number of barcodes amplified. Most cell lines appear to have one copy of the *BSD* containing cassette and one copy of the *PUR* containing cassette integrated in place of the two extant alleles (sequencing results: Figure 5.2). However, several cell lines had extra copies of *BSD* when observing the *PUR* to *BSD* ratio of sequencing reads during whole genome sequencing. These cell lines correspond with the cell lines that have dramatically high read counts in barcode reads amplified directly after pooling (Figure 4.11). Because the six replicates were independently pooled, individual pipetting errors can be clearly differentiated from these large increases in read count across all six replicates. By normalising to the day zero read counts these differences in barcode integration have been accounted for in the downstream results making it easier to compare relative fitness of RBP knockout cell lines.

4.10.7 RBP null mutants with fitness defects in the bar-seq screen

The results of the RBP bar-seq screen successfully demonstrated which RBP null mutants had fitness defects at different stages of the *L. mexicana* lifecycle. Over 70% of the RBPs screened (20/28) showed a fitness defect at some point during the screen. Together with the 39 RBPs that could not be recovered as knockout clones, 80% of the RBPs screened (59/67) potentially cause some level of disruption to the parasite's life cycle progression when removed. These results pave the way for further investigation of RBPs as essential regulators of differentiation, infectivity or virulence in *Leishmania*. Overall, there were not enough RBP candidates to be able to group phenotypic results by the type of RBD. Instead, each RBP is best treated separately as they are a diverse group of proteins where presence of a certain type of RBD does not necessarily preclude involvement in particular cellular pathways. For example, CCCH zinc finger proteins are thought to be involved in a wide array of different cellular processes in trypanosomatids (Mörking *et al.*, 2004, 2012; Bhandari and Saha, 2007; Paterou *et al.*, 2009; Walrad *et al.*, 2009; Kramer, Kimblin and Carrington, 2010; Benz *et al.*, 2011; Bhandari *et al.*, 2011; Ling, Trotter and Hendriks, 2011). Over the course of the screen, the increasing variance in barcode read counts meant that the standard error generally increased (Figure 4.17). However, when individual replicates for the barcode read count of each cell line are examined, the six replicates are closely grouped for the majority of the data even after six weeks in the mouse. The similar values of the three control cell lines, each with their own unique DNA barcode, provides a useful baseline for comparison with all other cell lines. For many of these RBPs, these data represent the first information about their function in kinetoplastids.

Several phenotypes that have been observed in null mutants or RNAi screens in *T. brucei* have been replicated in this screen in *L. mexicana* and are discussed below. Although the life cycles of these two distantly related parasites differ considerably, similarities can be found between procyclic and metacyclic trypomastigotes in *T. brucei* and procyclic and metacyclic promastigotes in *Leishmania.* In both species the procyclic stage is more proliferative and non-infective compared to the metacyclic stage which is non-proliferative and infective. Similarly, parallels can be drawn between bloodstream trypomastigotes in *T. brucei* and amastigotes in *Leishmania* in that they have both differentiated from metacyclic stages to respond to the changes in host conditions. Several of the candidate RBPs screened had been previously investigated in *T. brucei* in the context of differentiation, infectivity or virulence and are discussed here.

As explored in section 0, ZFP3 (LmxM.27.0130) has been studied as a *trans*-regulator of the cell surface procyclin mRNAs (Paterou *et al.*, 2009; Walrad *et al.*, 2009). Whilst its expression is constitutive, it associates with the polysomes in a stage specific manner (Paterou *et al.*, 2009). RNAi was used to reduce cellular levels of ZFP3 in trypanosomes, resulting in a general reduction in procyclin levels (Walrad *et al.*, 2009). Specific interactions with the EP1 procyclin, indicative of late procyclic promastigotes, explain the increased levels of EP1 when ZFP3 was ectopically expressed (Paterou *et al.*, 2009). In this bar-seq screen, although a slight reduction in ZFP3 barcodes was seen in the promastigote stages in *Leishmania mexicana*, the more pronounced fitness defect occurred

after six weeks in the mouse. Whether *Leishmania* ZFP3 is involved in surface protein regulation has yet to be investigated but is possible that reductions in surface coat protein levels (e.g. amastins, HASPB, SHERP or glycolipid-anchored proteins such as GP46 and GP63) would impact the fitness of either infective metacyclic promastigotes or amastigotes in establishing footpad infection. Defects in metacyclic promastigote fitness could explain the relative success of the ZFP3 null mutants in macrophage infections where there is no other immune exposure and parasites are directly incubated with their mammalian host cells. In *T. brucei*, a cytoskeletal nozzle phenotype was seen following both ZFP2 and ZFP3 ectopic expression (Paterou *et al.*, 2009). A morphological phenotype was not immediately obvious in *L. mexicana* ZFP2 null mutants, but more detailed imaging would be required to confirm this. All three of these proteins appear to be involved in the same mRNP complex, with the similar ZFP1 protein able to directly homodimerize with either ZFP2 or ZFP3 in *T. brucei* (Paterou *et al.*, 2009). To date these interactions have not been confirmed in *Leishmania*.

The closely related CCCH zinc-finger protein ZFP1 (LmxM.29.2200) has also been included in the RBP bar-seq screen. In *T. brucei*, ZFP1 is expressed transiently during differentiation to procyclic forms (Hendriks *et al.*, 2001), and in *T. cruzi*, during metacyclogenesis (Mörking *et al.*, 2004). Both are involved with the same messenger ribonucleoprotein complex in *T. brucei*. Attempts to knock out this protein were unsuccessful in *T. brucei* unless an ectopic copy was added back, suggesting essentiality in procyclics. In this screen, ZFP1 was not found to be essential; null mutants were successfully produced. However, it has a small but significant fitness defect during the promastigote stages of the bar-seq screen. Further investigation is required to determine if ZFP1 deletion interferes with the repositioning of the mitochondrial genome as in *T. brucei* (Hendriks and Matthews, 2005).

Another protein linked to differentiation in *T. brucei* is the RRM containing protein RBP10 (LmxM.23.0730). Expression of RBP10 was found to be greater in trypanosome bloodstream forms than in procyclic forms (De Pablos *et al.*, 2017). Overexpression of RBP10 in the procyclic form altered the transcriptome to more closely resemble that of the bloodstream form. When RBP10 was depleted in *T. brucei* using RNAi, differentiation from procyclic to bloodstream form was halted whereas ectopic expression of RBP10 in procyclic form converted them to bloodstream form (Mugo and Clayton, 2017). In the *L. mexicana* screen, RBP10 null mutants showed no significant defect during promastigote stages but had a strong decline in barcode reads in both macrophage- and lesion-derived amastigotes. This is consistent with previous studies in *T. brucei* and suggests that RBP10 is required for differentiation to amastigote forms in *Leishmania* (De Pablos *et al.*, 2017). One of several proximal proteins to RBP10 in *T. brucei*, RBP8, was included in the *L. mexicana* bar-seq screen but showed no significant defect at any stage.

Other RRM containing RBPs that were included in the RBP bar-seq screen and were previously studied in *T. brucei* are RBP43 (LmxM.25.0290) and RBP23 (LmxM.17.0550). As RBP43 was characterised further in Chapter 4.10 it is discussed in more detail in section 5.8.1. In *L. mexicana*, RBP43 null mutants showed significant fitness defects from day one post pooling to the six-week post infection samples where parasite numbers appear to have greatly reduced. Compared to RBP43,

RBP23 has been less well studied in *T. brucei*. Previous attempts at ablating RBP23 levels resulted in a slow growth phenotype before tetracycline induction of RNAi which recovered after induction (Wurst *et al.*, 2009). In a tethering screen, RBP23 was confirmed as upregulating mRNA in *T. brucei* (Lueong *et al.*, 2016). The only data on RBP23 function so far demonstrated a unique interaction with PABP1, potentially being involved in a small complex with eIF4E4 and eIF4G3 that does not localise to starvation stress granules (Zoltner *et al.*, 2018). In the *L. mexicana* bar-seq screen there was no significant effect due to RBP23 deletion at any stage of the knockout screen.

Several Pumilio and FBF (PUF) domain containing RBPs were chosen for knockout attempts and two, PUF6 (LmxM.32.1150) and PUF2 (LmxM.18.1420), produced viable null mutant lines that were included in the screen. Characterisation of PUF6 has been published in both *T. cruzi* and *L. infantum.* In *T. cruzi*, PUF6 is a constitutively expressed cytoplasmic RBP that is able to bind the *Drosophila* nanos reponse element (NRE) (Dallagiovanna *et al.*, 2005). Overexpression in *T. cruzi* caused a decrease in epimastigote specific transcripts (Dallagiovanna *et al.*, 2008). In *L. infantum*, PUF6 (along with PUF3,-7, -9 and -10) is a low reactivity B-cell antigen during the humoral response to visceral leishmaniasis in a hamster model (Folgueira, Martínez-Bonet and Requena, 2010). This is in contrast to the stronger reactivity described for PUF1, PUF2, PUF4, PUF5 and PUF8. Further molecular characterisation of PUF6 in *L. infantum* revealed that, like many PUF proteins, it causes destabilisation when bound to mRNA (Azizi, Dumas and Papadopoulou, 2017). Specifically, PUF6 binds to short interspersed degenerate retroposon (SIDER) regions in mRNA 3'UTRs as a mechanism of negative post-transcriptional control. Deletion of PUF6 was tolerated in *L. infantum* promastigotes but caused stabilisation of SIDER containing mRNAs. In *L. mexicana* promastigotes, PUF6 null mutants were also tolerated but showed significant reductions in barcode reads during the metacyclic purification and six weeks post infection in a mouse model of cutaneous infection. Similar to PUF6, PUF2 was found to be cytosolic and to have a role in destabilisation of mRNAs in a tethering assay in *T. brucei* (Jha *et al.*, 2014). Depletion of PUF2 by RNAi in trypanosomes inhibited growth in the bloodstream form due to the selective loss of mRNAs with long open reading frames. The PUF2 null mutant cell line in *L. mexicana* had strong fitness defects during the metacyclic promastigote and amastigote portions of the screen which is consistent with these previous findings.

Alba proteins have been extensively studied in kinetoplastids as well as in model organisms such as yeast. Both Alba1 (LmxM.13.0450) and Alba3 (LmxM.33.2580) null mutants were included in the RBP knockout screen. In trypanosomes, RNAi knockdown of Alba3 downregulated Alba1 (Subota *et al.*, 2011) despite their highly divergent sequences (16% identity). In *Leishmania* it has been demonstrated that Alba1 and Alba3 are functionally and physically linked through heterodimerization (da Costa *et al.*, 2017). Cytoplasmic Alba3 expression was ablated using RNAi in *T. brucei* leading to morphological defects and a disruption of the mesocyclic-epimastigote differentiation in the tsetse fly vector. RNAi targeting Alba3 in procyclic parasites caused similar effects including cell cycle arrest, posterior elongation and nucleus migration at the posterior end. Alba3 appears to interact with the translational machinery and has been pulled down as an interactor

with GPEET, EP Procylin and HSP70 3'UTRs (Mani *et al.*, 2011; Diana and Inchaustegui, 2015). Methylation of Alba3 by PRMT7 is thought to positively regulate its stabilisation of key virulence factor mRNAs such as δ-amastin (Ferreira *et al.*, 2020). Alba1 null mutant lines showed significant but moderate fitness defects, prominently after metacyclic purification and during the amastigote stages. A similar pattern was seen for Alba3 but with a stronger defect in lesion derived amastigotes. This is likely due to the destabilisation of crucial cell surface protein and virulence factors mRNAs such as δ-amastin but this requires confirmation in future studies.

Several other proteins included have been isolated in various screens, proteomes and interactomes that have previously been carried out in *T. brucei*, *T. cruzi* and *Leishmania* spp. (Nett *et al.*, 2009; Erben *et al.*, 2014; Klein *et al.*, 2015; Lueong *et al.*, 2016; Amorim *et al.*, 2017; Ferreira *et al.*, 2020; Nocua, Requena and Puerta, 2021). Strong correlations between previously reported phenotypes in other kinetoplastid species and the results of the *L. mexicana* knockout screens suggest that conserved RBP functional roles are being perturbed in similar ways after CDS deletion and RNAi depletion. However, many RBP deletion phenotypes seen in the bar-seq screen are entirely novel. Data presented here on the fitness cost of RBP deletion across all major lifecycle stages is a resource for further understanding the regulatory functions of these proteins in kinetoplastids and their involvement in differentiation and infectivity. Many proteins that have not been discussed here have never been studied previously in kinetoplastids and now can be studied in more detail in the relevant lifecycle stages. Four candidate proteins are investigated further in Chapter 4.10 for the purposes of validating the screen.

5 Functional analysis of *Leishmania mexicana RBPs*

5.1 Introduction

The screening of barcoded, RBP knockout cell lines in Chapter 3.10 identified those with reduced barcode read counts at specific stages of the *Leishmania mexicana* lifecycle. Fitness defects detected in amastigote and metacyclic promastigote parasites are of the greatest relevance to disease. Uncovering the regulation of pathways involved in metacyclogenesis, amastigogenesis or any kinetoplastid specific functions in theses stages may facilitate the discovery of novel drug targets. However, due to the progressive, nature of the pooled screen, there are several other explanations for changes in barcode read counts. One difference identified in section 4.6 was the higher barcode read count in some cell lines immediately after pooling. Due to the consistency of these results between replicates and the low levels of error in other cell lines it was hypothesised that those with increased counts had more copies of the integrated barcode than others. Although CDS replacement was confirmed for each RBP null mutant cell line using PCR, potential off-target integrations had not been investigated. *Leishmania* are also exhibit naturally high levels of aneuploidy and chromosomal mosaicism which could impact the results of the screen (Ubeda *et al.*, 2008; Rogers *et al.*, 2011; Sterkers *et al.*, 2011; Bussotti *et al.*, 2018; Negreira *et al.*, 2021). To clarify the above, whole genome

sequencing was carried out for 13/28 RBP null mutant lines, a representative proportion but restricted by the costs involved.

The bar-seq screen was both competitive and progressive. The phenotypes of the null mutant lines in the screen included an element of competition with the other null mutant lines and the control lines. Because the null mutant pools progressed directly into each subsequent lifecycle stage, phenotypes were also cumulative, with a drop in fitness in one stage leading to low numbers beginning the next. To validate some of the phenotypes seen in the bar-seq screen in isolation, several cell lines with promastigote fitness defects were selected and characterised individually. Phenotypes identified during the production of null mutant lines were also investigated further. Due to the time limitations of the PhD, further investigation of RBPs involved in amastigote specific defects was not possible. These candidates are currently being investigated by other lab members in more detail than would have been possible in this project (Teles *et al.* in preparation). Cell lines with phenotypes in the promastigote stages were chosen to validate the results of the screen and to investigate whether the phenotypes seen were occurring due procyclic or metacyclic promastigote defects. These genes were selected due to their clear phenotypes in the bar-seq screen and their potential importance in stationary phase culture.

Tagging these genes was a priority since it would allow several different lines of investigation. The small, versatile HA-epitope tag was chosen as it could be used for immunofluorescence, western immunoblotting and for RNA- or protein-immunoprecipitation. Immunofluorescence would allow the visualisation of the subcellular localisation of the selected RBPs. Western blotting of tagged RBPs allowed visualisation on PVDF membranes which revealed the molecular masses and expression pattern in log and stationary phase promastigote cultures as well as axenic amastigote culture. Expression patterns could then be compared to the timing of phenotypes observed in the bar-seq screen. The same tagged proteins will also be useful tools for identifying interacting proteins or RNAs in future immunoprecipitation experiments. The aims for this chapter were to:

- Sequence the genomes of the selected null mutant cell lines to confirm CDS replacement.
- Use whole genome sequencing data to look for evidence of off-target donor DNA integration.
- Repeat procyclic to metacyclic promastigote differentiation with these selected RBP knockout cell lines to validate the results of the screen.
- Edit and utilise plasmid vectors to tag candidate RBPs and facilitate further characterisation.
- Use tagged RBP cell lines to assess size, expression and localisation of the candidates.

5.2 Whole genome sequencing of RBP null mutants

The use of clones in the RBP bar-seq screen eliminated the possibility of heterozygous parasites or those with a remaining CDS dominating the pool. However, it was crucial to validate that the knockout phenotypes were not due to off target effects of the CRISPR system and that the RBPs of interest were deleted specifically. Regarding this, the few cell lines that had dramatically different

read counts immediately after pooling also had to be investigated for evidence of multiple integrations. To do this, whole genome sequencing was carried out for around half of the cell lines included in the screen (13/28) (methods 2.4.7). DNA extracted from each cell line (grown individually) was sequenced by Novogene using the Illumina HiSeq platform. All thirteen cell lines were sequenced successfully and sequencing data were mapped to the *Leishmania mexicana* genome (raw data processing, file conversion and alignment carried out by Katherine Newling and Eva Kyriacou using the Oxford Nanopore genome sequence generated by the Gluenz lab). Every RBP gene locus assessed had been replaced, as expected, with a high degree of precision (Figure 5.1). By dividing the read coverage of each of the 13 RBP loci by the median coverage of their parent chromosomes, a heatmap of coverage ratio was produced showing the successful replacement of each gene of interest (Figure 5.2). No evidence off-target disruptions of the other RBP genes were seen in these data. Some natural variation in chromosomal read coverage was observed consistent with similar studies and the wider literature that describe high levels of variation in *Leishmania* even in populations derived from a single clone (Ubeda *et al.*, 2008; Sterkers *et al.*, 2011; Damasceno *et al.*, 2020; Damianou *et al.*, 2020; Baker *et al.*, 2021; Negreira *et al.*, 2021).

The Illumina sequencing carried out by Novogene generated paired reads. Because the reads were paired, locating matched partner reads was used to investigate off target effects. The UTR's flanking the drug resistance cassettes belong to or have homologous sequence to regions in the *Leishmania mexicana* genome. Many reads mapped to the regions flanking RBP CDS deletion have matched partners that are present in a different area of the genome. These distant read partners were located at the genomic regions with high sequence homology to the UTR sequences in the integrated *BSD* and *PUR* donor cassettes. The *BSD* donor cassette contains a 60S acidic ribosomal protein P2 5'UTR and a glutamine synthase 3'UTR with some read partners mapping upstream of *LmxM.15.1203* and downstream of *LmxM.06.0370* respectively. The *PUR* donor cassette contains a malate dehydrogenase 5'UTR and a 60S ribosomal protein L5 3'UTR with some read partners mapping upstream of *LmxM.33.0140* and downstream of *LmxM.34.1880* respectively. All of these regions were checked for evidence of off-target donor DNA integration due to their homology. No evidence of reduced read coverage or disruption was found at these loci in any of the 13 null mutant lines sequenced. Furthermore, for each RBP null mutant line sequenced, checking the matched partners of other reads in the UTRs described above should reveal sequence similarity only at the deleted gene of interest. If any off-target integrations were present, some read partners should be found in other areas of the genome. Additionally, no reads mapped to the donor DNA UTRs were found with mates mapped to locations other than the genomic locus of those UTRs or the intended integration site (specific to each targeted RBP).

Sequencing reads were also mapped to the *PUR* and *BSD* genes. For each RBP knockout cell line, the coverage of both drug resistance coding regions was divided by the median read coverage across the chromosome they were integrated into (Figure 5.2). In the case where there were two alleles of the RBP candidate pre-editing, it can be assumed that successful knockouts selected under double

selection would have a *BSD* to *PUR* ratio of 1:1 at a coverage ratio of 0.5 for each construct (where 1 = 100% coverage, integrated donor cassette copies = ploidy of the chromosome). This was the case for: Δ*LmxM.36.6770,* Δ*LmxM.31.0950,* Δ*LmxM.27.0130,* Δ*LmxM.25.0290,* Δ*LmxM.18.0590,* Δ*LmxM.17.0550 and* Δ*LmxM.15.0130*. Other cell lines likely had more copies of *BSD* than *PUR*: Δ*LmxM.36.1635,* Δ*LmxM.33.2580,* Δ*LmxM.30.1650,* Δ*LmxM.24.1570,* Δ*LmxM.05.0850*. Due to the lack of evidence for off-target donor cassette integrations, the most likely explanation for these uneven ratios is the integration of multiple copies of resistance cassettes at the same locus. These cell lines also correspond with those mentioned in section 4.6, explaining those with higher day zero barcode read counts. As the number of parasites from each cell line were equal on day zero, this suggests that either integrated barcoded regions were replicated or that more than two copies were integrated. The sum of the coverage ratios for *BSD* and *PUR* resistance is around 1 (Figure 5.2) suggesting that the number of integrated copies is consistent with the copy number of the chromosome. This, along with the checking of distant sequencing read partners, means it is highly unlikely that additional off-target integrations of the barcoded donor DNA occurred. The plasticity of the *Leishmania* genome is well documented to display variations in copy number seen at different stages of the lifecycle for many different genes and gene clusters (Ivens *et al.*, 2005; Damasceno *et al.*, 2020; Negreira *et al.*, 2021). Some of this natural variation can be seen in the non-target gene coverage. Consistent with the literature, the median coverage of chromosome 30 was double that of the average (tetraploid) for Δ*LmxM.30.0250* but appeared to be triploid for Δ*LmxM.30.1650*. Based on median chromosomal coverage, chromosomes 24 and 25 also appeared to be (at least partially) tetraploid in Δ*LmxM.24.1570* and Δ*LmxM.25.0290* respectively.

5.3 Growth curves of RBP null mutants

RBP knockout growth phenotypes were investigated following the discovery of the Δ*LmxM.15.0130* stationary phase growth defect (section 4.2). Some of the first successful barcoded knockout cell lines produced were grown in M199 medium in triplicate, where cell counts were measured over the course of eight days and compared to the T7/Cas9 and WT controls (Figure 5.3). While several cell lines had similar curves to the controls, some had a lagging phenotype and took longer to reach stationary phase (Δ*LmxM.36.1635* [ZFP2] and Δ*LmxM.33.2580* [Alba3]). Δ*LmxM.27.0130* had a rapid reduction in cell count after peaking at day four. Δ*LmxM.15.0130* once again showed a strong reduction in cell count, reaching $2x10^7$/ml less than the controls in stationary phase culture.

Growth curves were produced for all four null mutant lines chosen to validate the results of the barcoded knockout screen (Δ*LmxM.05.0850,* Δ*LmxM.15.0130,* Δ*LmxM.24.1570* and Δ*LmxM.25.0290*). Statistical analysis was carried out in RStudio using an unpaired t-test to compare the mean of the control cell lines to the mean of the null mutant lines at each time point. As multiple comparisons can increase the chance of false positive results, a Bonferroni correction for false discovery was included. While significant differences were found for all four knockout cell lines in at least a single time point, it is also important to remember that some cell lines are represented by a single clone. Δ*LmxM.15.0130* was compared to the T7/Cas9 cell line for a final time showing the

same phenotype as in previous repeats (Figure 5.4a). Δ*LmxM.05.08*50 had a pronounced lag in growth, reaching stationary phase at day seven rather than day five (Figure 5.4b). This was true across three independent knockout clones, each of which was grown in triplicate. Δ*LmxM.25.0290* appeared to have a reduced cell count during stationary phase from day four through to day seven (Figure 5.4c). However, this reduction was not as drastic as in Δ*LmxM.15.0130* and by day seven the cell count had recovered to near control levels. Δ*LmxM.24.1570* did not show the increased fitness compared to the control cell lines, during stationary phase, that is suggested by the results of the bar-seq screen (Figure 5.4d). The growth slightly lagged behind the controls until around day seven where levels were similar. This result suggests that rapid stationary phase growth is not a likely explanation for the dramatic increase in barcode reads for this cell line in stationary phase of the barcoded knockout screen. The growth of this cell line is actually slightly impaired by the deletion of *LmxM.25.0290*. Excluding this last cell line, the phenotypes seen in the screen were also observed in isolation, without the element of competition complicating the data. This demonstrates the screen as an effective resource for screening RBP null mutants with a range of phenotypes that may be related to processes such as replication, differentiation or infectivity.

5.4 Tagging selected RBPs

HA epitope tags were chosen for manipulation and visualisation of RBP candidates for several reasons. Firstly, commercial antibodies to the HA-tag have low background in *Leishmania mexicana* when viewed on a western blot. Secondly the HA-tag is small and less likely to disrupt protein function (in particular protein-RNA interactions) than many larger tags such as GFP. Lastly, commercially available anti-HA magnetic beads provide strong binding and clear results for RNA immunoprecipitations (RIPs). The following sections describe the production of tagging vectors, their application to tagging candidate RBPs and confirmation of tagged clones. Detailed methodology for each process described can be found in the methods section (Chapter 2).

Despite being able to use the same primers for tagging as for knockout, a HA-tagging plasmid was not available as a template for PCR amplification of donor DNA. New plasmids were produced to enable this as follows. Firstly, for N-terminal HA-tagging the pGL2666 and pGL2737 plasmids (Beneke *et al.*, 2017; Baker *et al.*, 2021) were edited so that a triple HA-tagging sequence replaced the triple Myc-tag region upstream of mNeonGreen or mCherry respectively. The N-terminal tagging region of the original plasmids was amplified using PCR to isolate a region that contained no repeated sequences. A second PCR used the amplified N-terminal fragment as template and a forward primer containing the triple-HA sequence and binding downstream of the last Myc repeat, to replace triple Myc with triple HA. HindIII and NcoI sites were also introduced at the 5' and 3' ends of this second amplification respectively. The resulting fragment was ligated into a temporary vector and digested with HindIII and NcoI, then ligated into HindIII and NcoI digested pGL backbones to complete cloning. Both complete N-terminal HA-tagging vectors were cloned and sent for sequencing where they were found to match the predicted sequence with no significant differences. The unique linker region between the HA-tag sequences and the fluorescent protein encoding gene can be used as the

reverse primer binding site when amplifying a HA-only tagging cassette. By moving primer binding sites and using either plasmid it is also possible to tag with 3HA-mNeonGreen, 3HA-mCherry or either fluorescent protein alone (Figure 5.5a,b).

For C-terminal tagging, the obvious approach of editing the C-terminal tagging region of the previously described plasmids from triple Myc to triple HA was not straightforward due to the repeated sections and lack of restriction sites. An equally effective method was used to modify another available plasmid. This modified HA-tagging plasmid with a pPOTv2 backbone (Dean *et al.*, 2015; De Pablos *et al.*, 2019) was edited to replace the paraflagellar rod 3'UTR with that of N-Myristoyltransferase which is known to have stable expression across lifecycle stages in *Leishmania* (Price *et al.*, 2003). The triple HA was also replaced with three HA-tags that differ in nucleotide sequence but not in protein sequence, reducing primer binding issues and allowing selection of one, two or three tags depending on the application. Drug resistance cassettes were replaced to allow for Blasticidin and Puromycin selection in the T7/Cas9 cell line (JM6571). All the modifications to pPOTv2_3HA were carried out using standard restriction cloning, introducing restriction sites in the insert-amplification primers. The resulting plasmids and PCR amplified tagging constructs are shown in Figure 5.5c,d.

N-terminal tagging was attempted using double drug selection of whole populations of transfected parasites in flasks. As described previously (section 4.2), the CRISPR/Cas9 system developed by Beneke et al., (2017) allows for the same primers to be used for knockout and tagging. Primers were designed for gene deletion using the LeishGedit server followed by manual checks and optimisation where necessary. Tagging cassettes for four selected RBP genes (*LmxM.15.0130, LmxM.05.0850, LmxM.25.0290* and *LmxM.24,1570*) were amplified from the pGL26663HA and pGL27373HA plasmids with the upstream forward primer (previously used for knockout) containing 30bp of homology upstream of the gene of interest. The cell line Δ*LmxM.15.0130* was selected due to the promastigote growth defect found in initial CRISPR tests that produced a non-barcoded null mutant (Figure 4.3). The severe fitness defect was replicated in the screen but with an independently produced cell line including the 12-nucleotide barcode region that was added upstream of the drug resistance cassette in all knockout lines included in the bar-seq screen (Figure 4.19). Discovery of a morphological phenotype and its potential for further study due to a strong reduction in read count in the mouse model of infection led to the selection of Δ*LmxM.05.0850.* Δ*LmxM.25.0290* was chosen because of a reduced barcode read count in the promastigote stages as well as further growth defects during the mouse infections that could be investigated in the future. Unlike other RBP null mutants, Δ*LmxM.24.1570* had an increase in barcode reads compared to the control cell lines (during the stationary phase) so warranted further investigation. This cell line also had dramatic reductions in barcode reads during the amastigote stages of the experiment.

A paper by Fiebig, Kelly and Gluenz (2015) identified many genes in *Leishmania* that have a possible N-terminal extension. The 67 knockout candidates were compared to the findings of this paper and several genes were identified with possible N-terminal extensions (*LmxM.15.0130, LmxM.30.0250,*

LmxM.22.1500, LmxM.34.2270, LmxM.14.1140, LmxM.04.1170, LmxM.29.1110 and *LmxM.32.0260*). Due to the possible N-terminal extension for gene LmxM.15.0130, the upstream forward and downstream reverse primers were redesigned to correspond to the larger predicted CDS. Previous attempts to tag the short CDS were unsuccessful (data not shown). In the upstream reverse primers, the 30bp of homology corresponds to the first 30bp of the target CDS but with the ATG mutated to ATC (Met-Ile) to avoid translation starting from the endogenous start codon. For the purposes of producing the barcoded null mutants, in the previous chapter, the entire extended CDS was replaced in all cases.

PCR amplified tagging cassettes containing *PUR* and *BSD* genes were nucleofected into T7/Cas9 expressing *L. mexicana* (JM6571) along with 5' sgRNA cassettes produced as for knockout. However, after selection with both Blasticidin and Puromycin for two weeks in flasks, and extraction of protein followed by anti-HA western blot in log phase, only one tagged protein could be seen, LmxM.25.0290 (Figure 5.6). It was hypothesised that the low protein levels observed were due to partial tagging where although multiple copies had been tagged in some cells, others may have additional untagged copies or may only have one integrated copy but escaped selection. To overcome this, tagging was repeated using single selection with Puromycin to increase transfection efficiency. Both the N- and C-terminus of each gene were targeted in case the tag was interfering with protein levels or was occluded by the protein. After six hours recovery in flask, parasites were diluted in 96 well plates and selected with Puromycin to isolate clonal populations. Subsequent PCR screening was used to select clones with only tagged copies of the CDS.

DNA extracted from the populations in 5.2.2 was used as a template for PCRs to identify the presence or absence of remaining untagged copies of the gene (Figure 5.7). At least one fully tagged clone was identified for each candidate RBP. C-terminally tagged clones were produced for *LmxM.15.0130* and *LmxM.24.1570* (C12 and C8). An N-terminally tagged clone (F3) was produced for *LmxM.05.0850* and an unclear result at the C-terminus. *LmxM.25.0290* was the most successful, with an N-terminally tagged clone (D7) and multiple C-terminally tagged clones produced (B10, D11 and H11 which was used from here on). These results matched what was expected based on the location of key RNA-binding domains (Figure 5.8). DNA from positive clones was checked using PCR and sequenced to confirm the integrity of the tag. Tagging constructs integrated into the genome were diagrammed for each cell line in Figure 5.9.

5.5 Expression patterns of HA-tagged RBPs

Western blots were used to confirm expression of the tagged RBPs and to check their molecular masses. Anti-HA mouse primary and anti-mouse HRP secondary antibodies detected all four tagged candidate genes from 1×10^7 log phase *L. mexicana* (Figure 5.10). Protein levels were compared to the anti-N-Myristoyltransferase (NMT) loading control from the same membrane stripped and re-blotted. T7/Cas9 cells (JM6571) showed no bands demonstrating no observable cross reactivity of the primary or secondary antibodies used. LmxM.15.0130-3HA was observed as a single band at

~100kDa compared to a molecular mass (including a linker and three HA tags) of 89kDa predicted using the Expasy server (https://web.expasy.org/compute_pi/). Successful tagging here, compared to previously unsuccessful attempts to tag the short CDS (data not shown), suggests that the extended CDS described earlier, may be the predominant form of this protein. 3HA-LmxM.05.0850 was observed at ~75kDa (predicted 69kDa) with two lower bands possibly representing degradation products. Use of weaker sample buffer (lane 8) rather than stronger (lane 9) can eliminate these bands. 3HA-LmxM.25.0290 and LmxM.25.0290-3HA showed stronger expression than any of the other proteins with a molecular mass below 75kDa (predicted 66kDa) for both N-terminally and C-terminally tagged clones. Weak signals for degradation products were also seen. LmxM.24.1570-3HA was observed as two distinct molecular masses, one at ~100kDa and one between 75 and 100kDa (predicted 76kDa).

Having confirmed the expression of all four candidate RBPs, it was important to further investigate their expression in different lifecycle stages. A time-course was designed to grow all five tagged cell lines (3HA-LmxM.05.0850, LmxM.15.0130-3HA, 3HA-LmxM.25.0290, LmxM.25.0290-3HA and LmxM.24.1570-3HA) from log to stationary phase and then differentiate them from promastigotes to axenic amastigotes. Parasites were grown in Grace's medium for seven days at which point they were transferred to 33°C to differentiate. Protein was extracted at daily time points and $7.5x10^6$ cells were loaded per lane on SDS-PAGE gels. Western blots using anti-HA primary and anti-mouse HRP secondary antibodies as before, showed expression profiles for all four candidate RBPs across three major lifecycle stages in comparison to constitutively expressed NMT (Figure 5.11). LmxM.15.0130-3HA showed strongest expression in the first 24hrs of log phase (Figure 5.11b). In addition to detection at 100kDa as previously, it was weakly detected just below this and further below 75kDa. Expression of these smaller peptides was seen at day two and three but then dramatically reduced in stationary phase (days 4,5,6 and 7). Although reduced, faint protein expression can be seen in stationary phase whereas in amastigotes (days 8,9 and 10) no expression could be seen. 3HA-LmxM.05.0850 was also expressed most in early log phase, with expression severely reduced by day four and none visible by day five (Figure 5.11c). This protein was detected just above 75kDa as expected. 3HA-LmxM.05.0850 appeared to show no expression in amastigotes. LmxM.25.0290, whether tagged at the N- or C-terminus, was most strongly expressed at 24h, in early log phase (Figure 5.112d,e). Reduced expression was present on days two and three, but this further reduced by day four and remained barely visible for the remainder of the time course. The similarity between N- and C-terminally tagged expression profiles suggests that 3'UTR alterations in the C-terminally tagged LmxM.25.0290 haven't impacted protein expression dramatically. As expected, this protein was detected at ~75kDa throughout the time course. LmxM.24.1570-3HA was also expressed most strongly at day one at ~100kDa, and more at several lower molecular masses (Figure 5.11a). A consistent signal was present on days two, three and four but only at ~100kDa. After day four no signal was visible. In the majority of these western blots the NMT control expression appeared to reduce over time, especially in the last three days corresponding to the axenic amastigote

stage of the experiment. While enough expression was present for comparison across most if the blot, amastigote specific expression should be investigated further.

5.6 Imaging HA-tagged RBPs

Whilst the HA-tag was mainly intended to facilitate RNA-immunoprecipitations, its strength and specificity make it a good choice for immunofluorescence experiments. The subcellular localisation of a protein can reveal information relevant to protein function, and in the case of RNA-binding proteins, can help to provide context for the RNAs that it interacts with. To investigate the subcellular localisation of the four tagged RBP candidates, anti-HA immunofluorescence staining was carried out on fixed cells from early log phase culture to coincide with high expression levels. Both rabbit and mouse commercial anti-HA primary antibodies (Table 7.1, A190-108A, 26183) were used with AlexaFluor secondary antibodies. For both LmxM.15.0130-3HA and 3HA-LmxM.05.0850, no signal was seen above background, possibly because endogenous expression levels weren't high enough to be detectable. For LmxM.24.1570-3HA a weak cytoplasmic signal was seen but was hard to differentiate from background level staining without further repeat experiments (data not shown). Both N- and C-terminally tagged LmxM.25.0290 were detected as a strong, diffuse cytoplasmic signal, above background levels and with both rabbit and mouse antibodies (Figure 5.12). As with the other cell lines, some background signal was present but the difference between the controls and all tagged RBP expressing preparations allows for confidence that the staining pattern of this protein has been visualised. A cytoplasmic distribution also matches the data from the TrypTag database for the *T. brucei* homolog of this protein tagged at the N-terminus (Figure 3.5). This cell line also had the highest tagged protein signal when 1×10^7 cells were loaded per lane on a western blot (Figure 5.10).

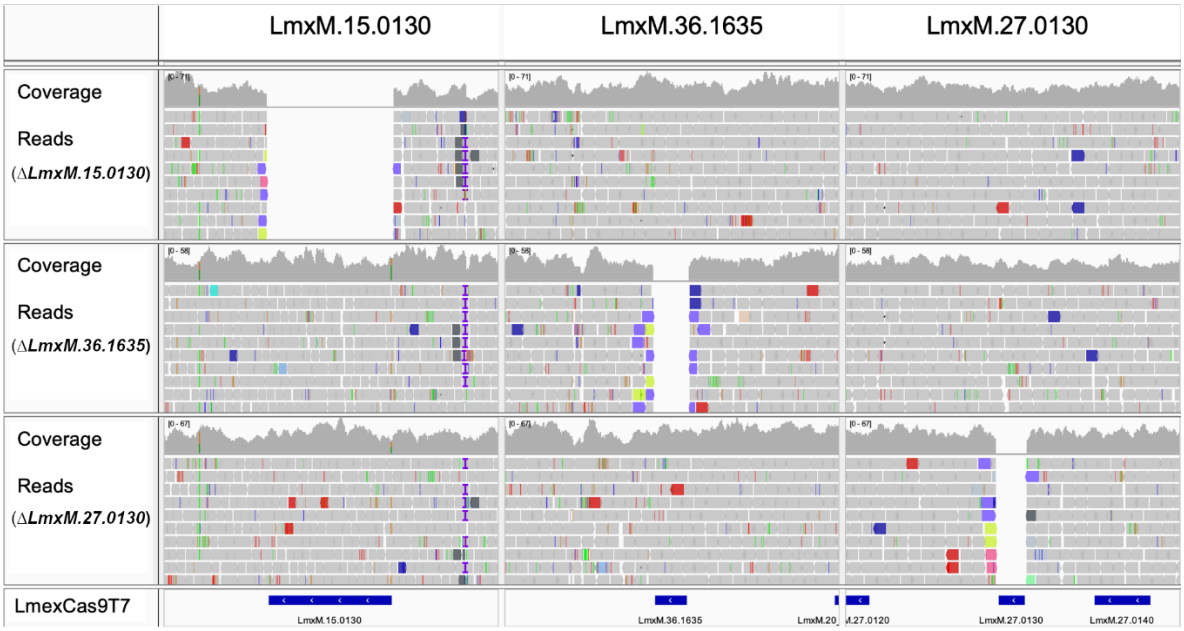
Although the protein could not be detected using immunofluorescence, Δ*LmxM.05.0850* had a clear morphological phenotype that was detected during the cloning process. When day three parasites from the knockout cell line were compared to the T7/Cas9 (WT) control under the light microscope there were distinct differences (Figure 5.13). The flagellum was shortened to the point where the cells were immotile, clumped together easily and aggregated at the bottom of the flask. The remaining portion of the flagellum was only visible in some cells and often had a loop or protrusion near the flagellar attachment zone. Despite being severely affected by *LmxM.05.0850* deletion, the cells still appeared to divide, often being seen paired post division due to their lack of motility. The relative lengths of cells from this cell line and the T7/Cas9 controls were measured in ImageJ and compared in Figure 5.14. The T7/Cas9 and Δ*LmxM.05.0850* cell lengths were compared, with means of 11.0 and 9.00. A Welch's two-samples t-test showed that the difference was statistically significant: $t(233.5) = 6.65$, p-value = 2.03×10^{-10}. Due to the width data for the T7/Cas9 not conforming to a normal distribution, an independent 2-group Mann-Whitney U Test was carried out with the results as follows: $U = 1194.5$, p-value < 2.2×10–16. This means the null hypothesis of no difference between the ranked independent datasets can be rejected with the width of

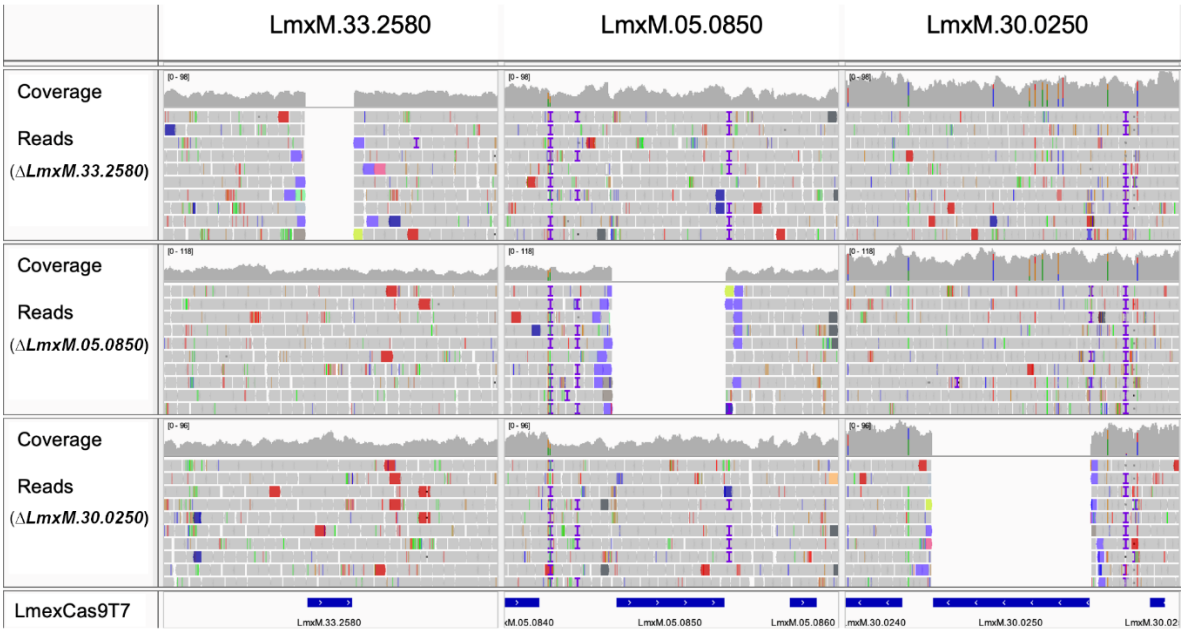
Δ*LmxM.05.0850* cells being significantly greater than that of the T7/Cas9 cells with wildtype morphology.

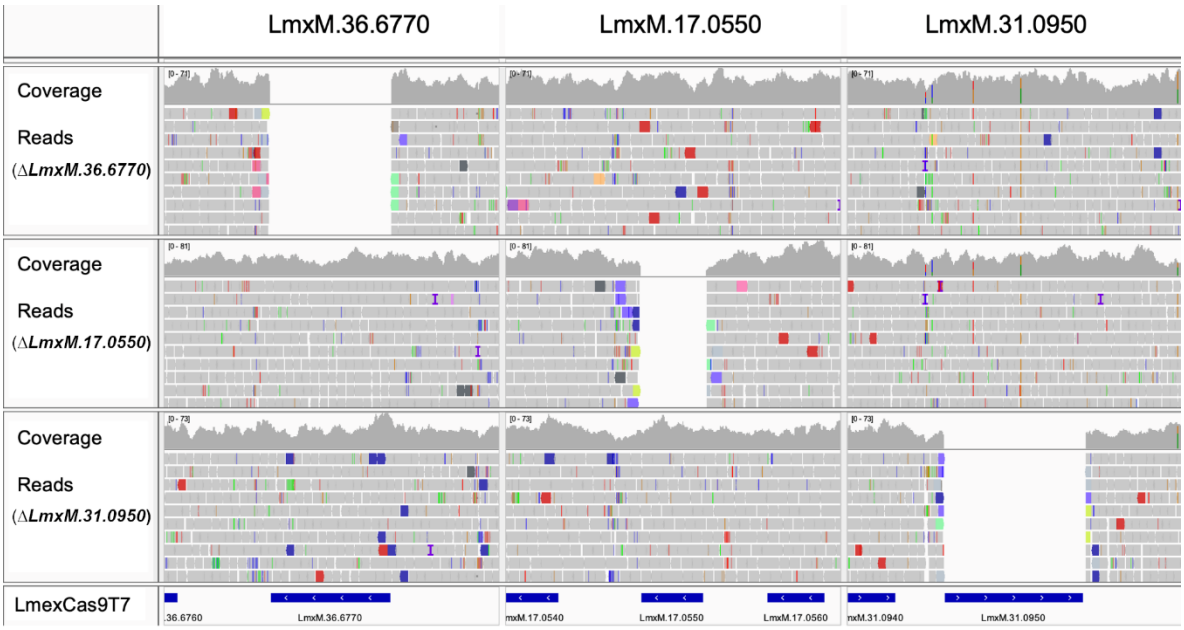
In addition to the morphological defect during log and stationary phase, incubating Δ*LmxM.05.0850* cells for more than three weeks at 26°C resulted in extensive differentiation to smaller cells with amastigote-like morphology. This is in contrast to the T7/Cas9 cell-line where most cells died after this long incubation. To test if the small amastigote-like cells were alive, cells were preincubated with Mitotracker which requires an active mitochondrial membrane for uptake. The cells were washed and fixed in paraformaldehyde before being stained with DAPI and mounted on slides (Figure 5.15). It is clear that there are many more live cells in the Δ*LmxM.05.0850* culture than in the T7/Cas9 and that those with the most staining have amastigote-like morphology. These data give an indication that, despite being mostly expressed in the log phase, deletion of *LmxM.05.0850* has an impact on differentiation or cell remodelling that extends to later lifecycle stages.

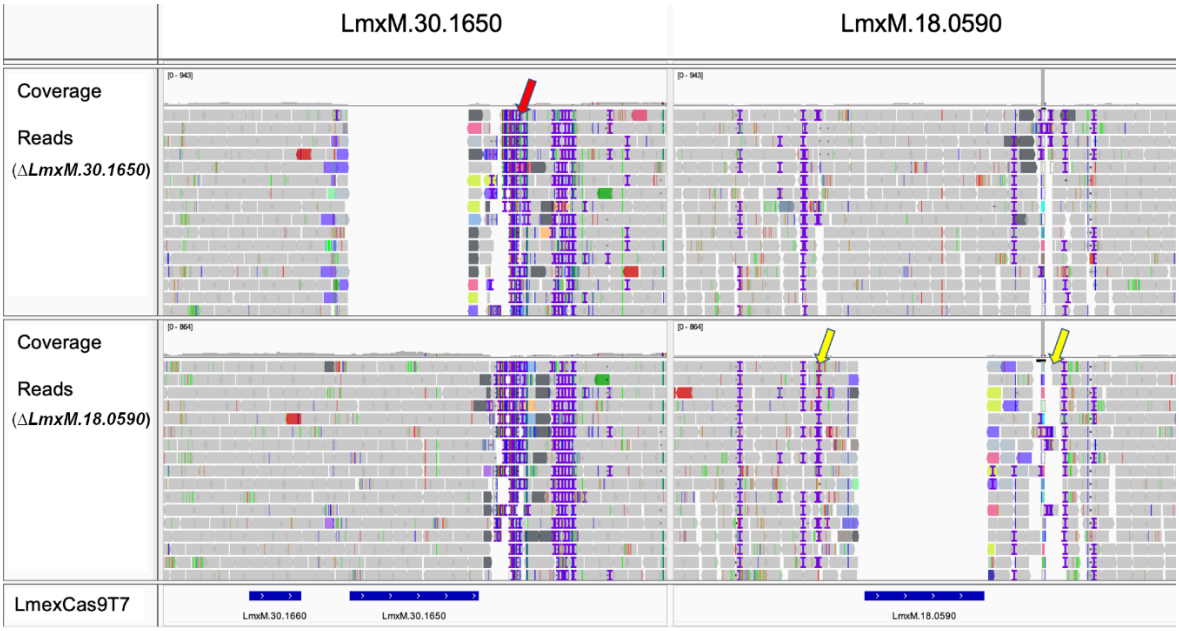
5.7 Conclusions

In conclusion, three out of the four RBP knockout cell lines investigated (Δ*LmxM.15.0130*, Δ*LmxM.05.0850* and Δ*LmxM.25.0290*) had phenotypes in the bar-seq screen that were validated through individual growth experiments. The cell line Δ*LmxM.24.1570* did not show the increase in fitness in stationary phase that was seen in the knockout screen. All four candidate RBPs were successfully HA-tagged and visualised for the first time by western blotting. Using the tagged cell lines, expression patterns for each protein were determined across the three major lifecycle stages. All four RBPs showed predominantly procyclic stage expression. Anti-HA immunofluorescence demonstrated that LmxM.25.0290 has a cytoplasmic distribution and that LmxM.24.1570 may also be cytoplasmic. *LmxM.05.0850* deletion caused a dramatic reduction in flagellar length as well as significantly reducing the length of the cell and increasing the width. A potential link to amastigogenesis was also found for Δ*LmxM.05.0850* cells with many live, amastigote-like cells surviving after long incubations at 26°C. However, this result requires repetition and further experimentation to make solid conclusions about amastigogenesis. Finally, the HA-tagged cell lines can now be used in other experiments such as RNA-immunoprecipitation to find mRNA targets of the RBPs under investigation. RNA-immunoprecipitations were carried out at the end of this study but analysis of results and confirmatory q-PCRs are still in progress so this work will be presented elsewhere.


LmxM.15.0130
LmxM.36.1635
LmxM.27.0130
Coverage
Reads
(ΔLmxM.15.0130)
Coverage
Reads
(ΔLmxM.36.1635)
Coverage
Reads
(ΔLmxM.27.0130)
LmexCas9T7
LmxM.15.0130
LmxM.36.1635
LmxM.27.0130
LmxM.27.0140


LmxM.33.2580
LmxM.05.0850
LmxM.30.0250
Coverage
Reads
(ΔLmxM.33.2580)
Coverage
Reads
(ΔLmxM.05.0850)
Coverage
Reads
(ΔLmxM.30.0250)
LmexCas9T7
LmxM.33.2580
LmxM.05.0850
LmxM.05.0860
LmxM.30.0240
LmxM.30.0250


LmxM.36.6770
LmxM.17.0550
LmxM.31.0950
Coverage
Reads
(ΔLmxM.36.6770)
Coverage
Reads
(ΔLmxM.17.0550)
Coverage
Reads
(ΔLmxM.31.0950)
LmexCas9T7
LmxM.36.6770
LmxM.17.0540
LmxM.17.0550
LmxM.17.0560
LmxM.31.0940
LmxM.31.0950


LmxM.30.1650
LmxM.18.0590
Coverage
Reads
(ΔLmxM.30.1650)
Coverage
Reads
(ΔLmxM.18.0590)
LmexCas9T7
LmxM.30.1660
LmxM.30.1650
LmxM.18.0590

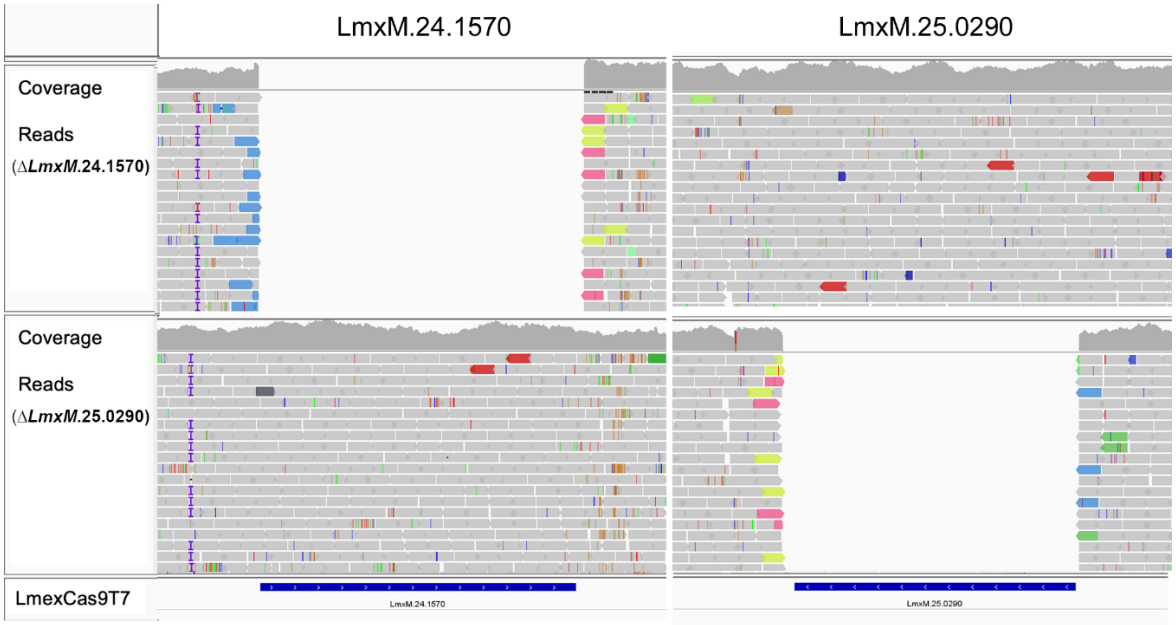


Figure 5.1. Whole genome sequencing of RBP null mutant clones. In total 13/28 RBP knockout cell lines were sent for whole genome sequencing using the Illumina HiSeq platform. Individual reads are shown as grey arrows within the IGV interface, mapped to the genome of the parental T7/Cas9 expressing cell line JM6571. Each row represents reads from a different RBP knockout cell line. Where read partners were separated by large distances they are coloured, with each colour representing a different location. Read coverage is represented graphically above each set of mapped reads. A red arrow shows an example of insertions compared to the reference genome. Since these were detected in all cell lines it is likely this is due to divergence of these cell lines from the sequenced parental line. Likewise, regions of low read quality (yellow arrows) were found at specific locations in all cell lines. Figures from Kyriacou 2021 with additions of *LmxM.24.1570* and *LmxM.25.0290* data.

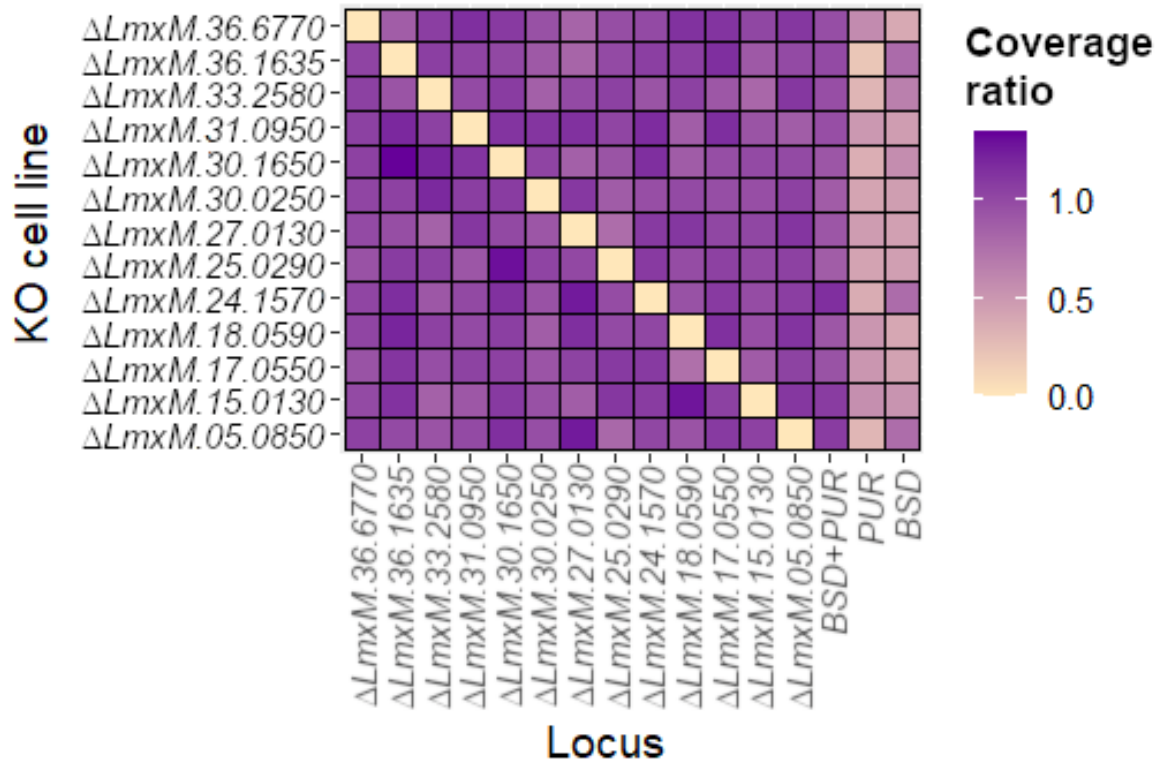


Figure 5.2. RBP-locus read coverage compared to chromosomal average. A coverage ratio was calculated comparing the read coverage at the genomic loci of each of the 13 RBP knockouts to the median coverage of the relevant chromosome. A coverage value of one represents the same coverage at the locus of interest as the median of the whole chromosome containing the RBP of interest. Coverage of zero represents no coverage at the region of interest. The coverage of the *BSD* and *PUR* genes was compared for each cell line to the median coverage of the chromosome containing locus they were integrated into. The column '*BSD+PUR*' refers to the sum of the proportional coverage of both *BSD* and *PUR* which should sum to one if no off-target integrations have occurred.

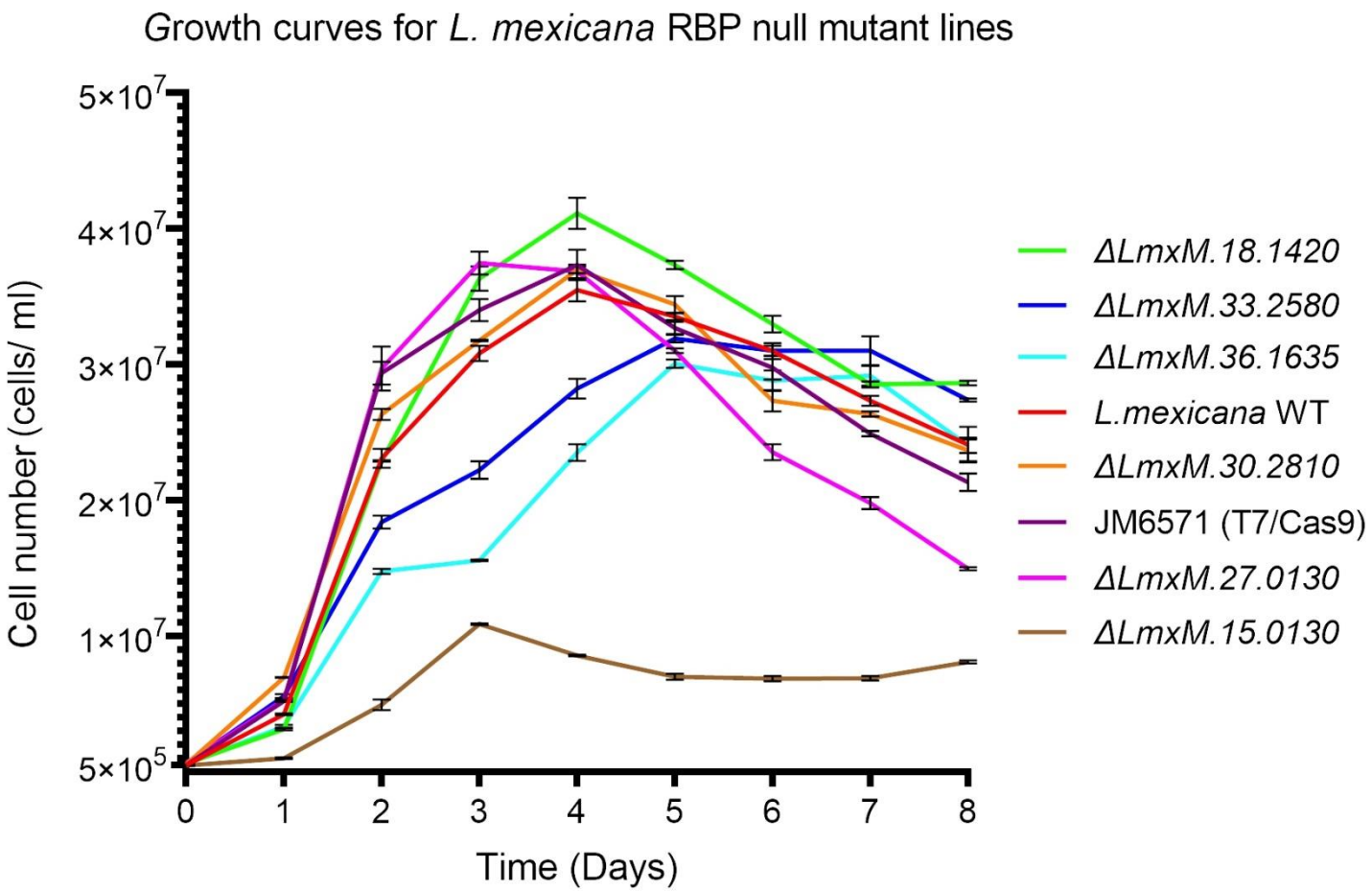
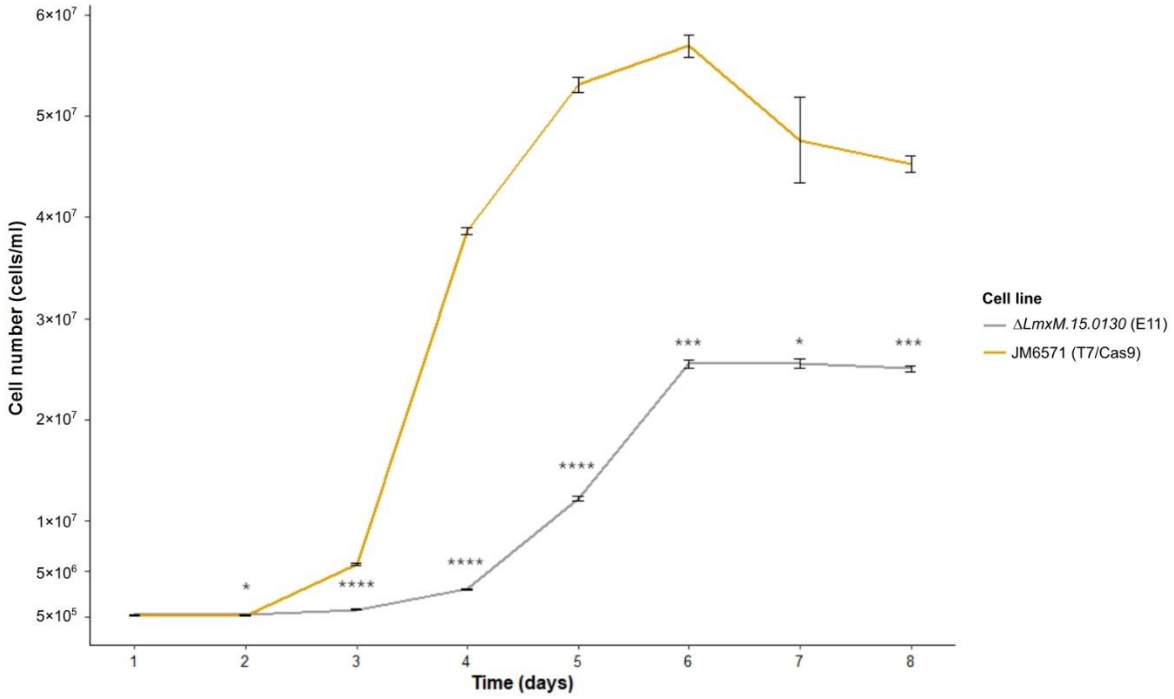
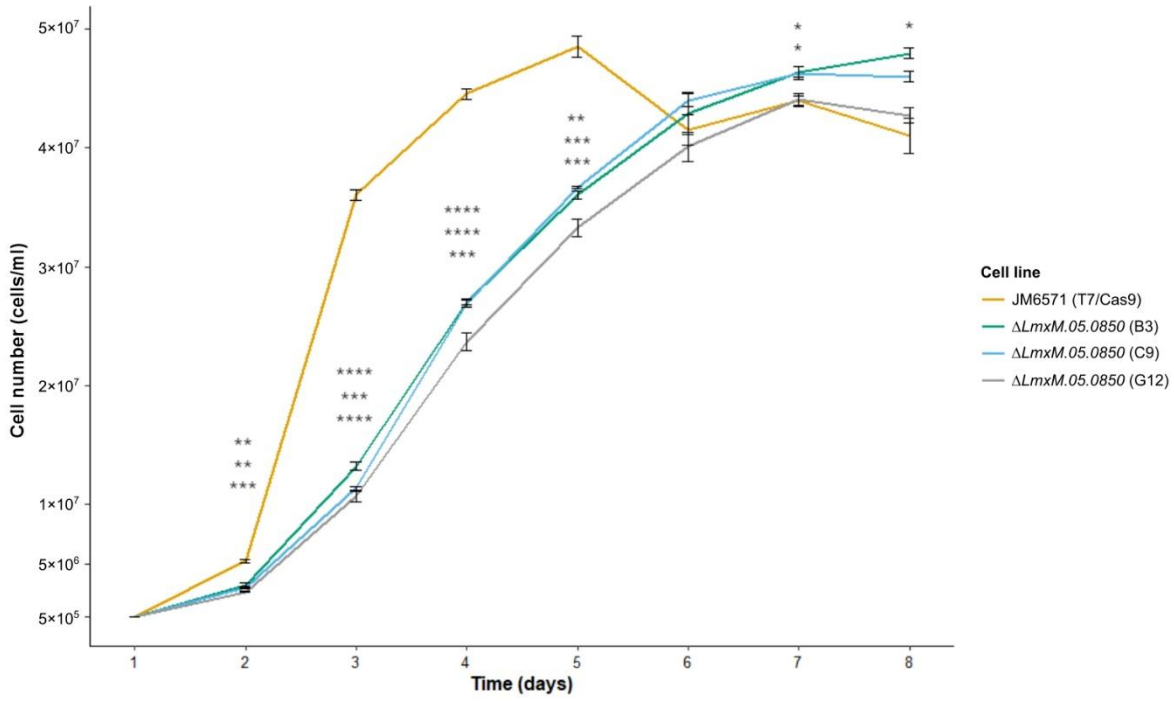


Figure 5.3. Promastigote growth of null mutant cell lines. RBP null mutant cell lines were grown in M199 media, in triplicate, from an initial concentration of $5x10^5$ cells/ml. The parental T7/Cas9 expressing *L.mexicana* cell line (JM6571) was used for comparison along with an *L.mexicana* WT cell line (M379). Cells were counted daily with a Beckman-Coulter Z2 series particle counter. Error bars show standard error around the mean of triplicate cell counts.

A)


Cell number (cells/ml)
6×10^7
5×10^7
4×10^7
3×10^7
2×10^7
1×10^7
5×10^6
5×10^5
1
2
3
4
5
6
7
8
Time (days)
Cell line
ΔLmxM.15.0130 (E11)
JM6571 (T7/Cas9)

B)


Cell number (cells/ml)
5×10^7
4×10^7
3×10^7
2×10^7
1×10^7
5×10^6
5×10^5
1
2
3
4
5
6
7
8
Time (days)
Cell line
JM6571 (T7/Cas9)
ΔLmxM.05.0850 (B3)
ΔLmxM.05.0850 (C9)
ΔLmxM.05.0850 (G12)

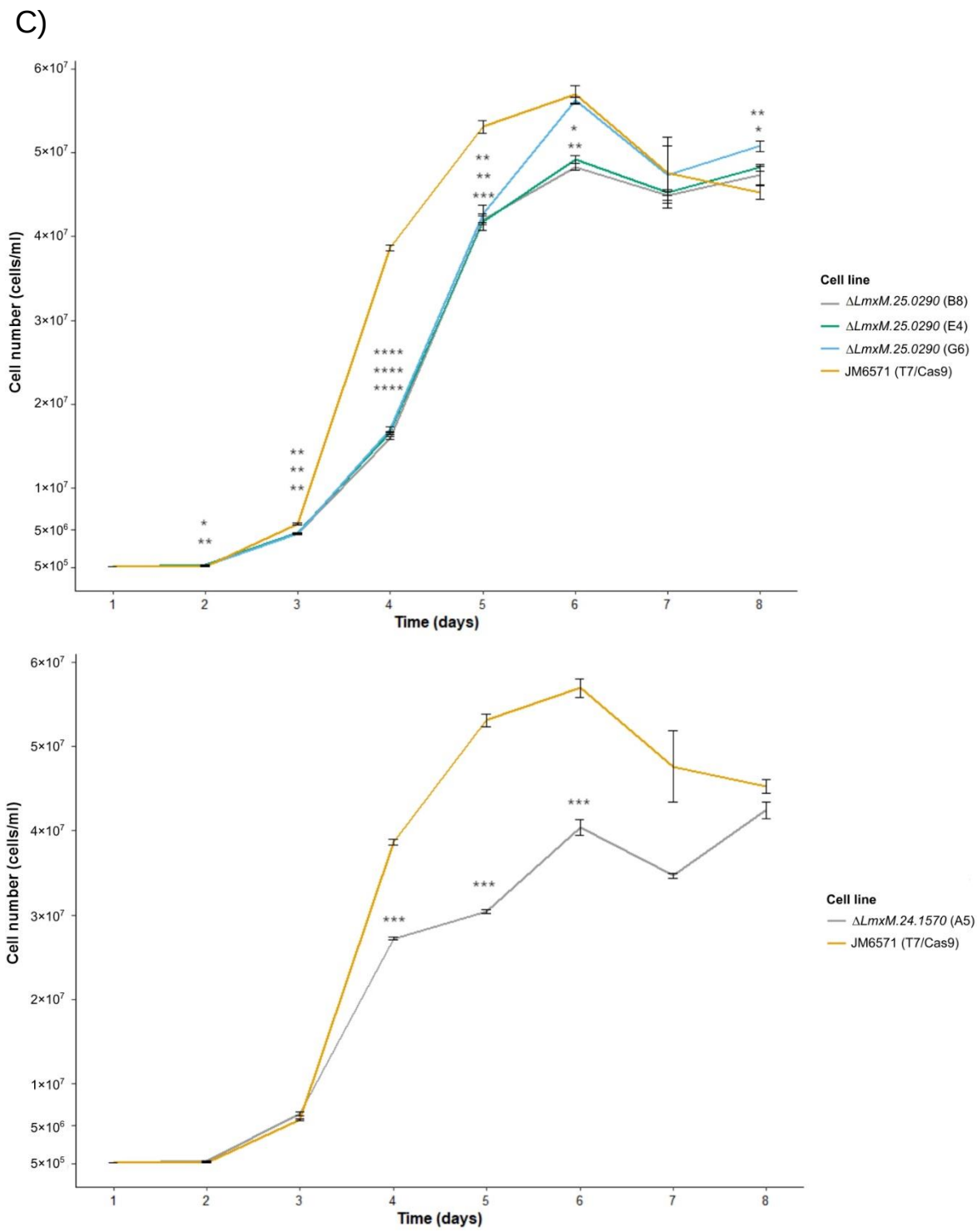
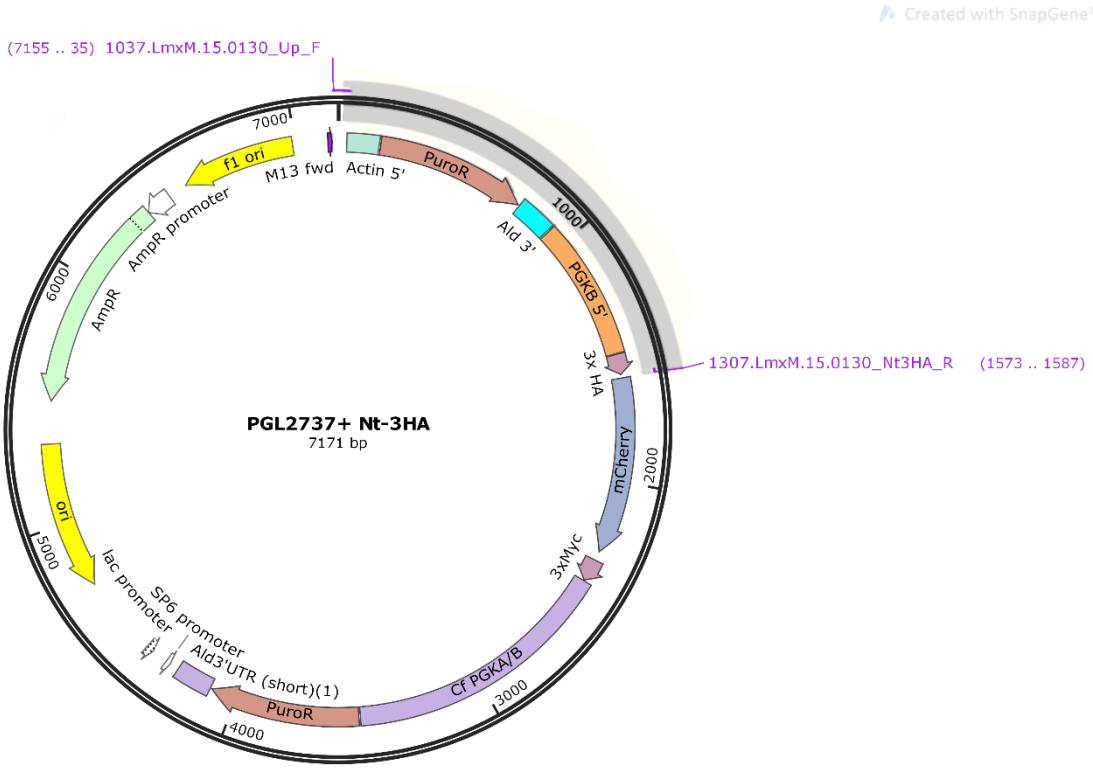
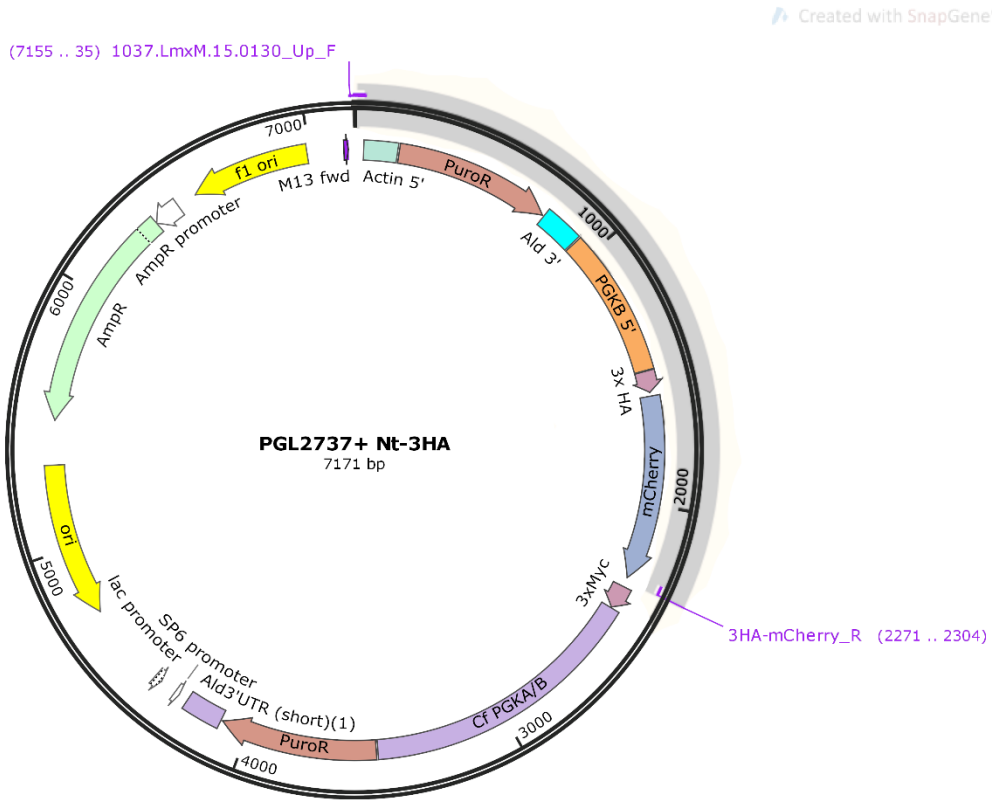


Figure 5.4. Promastigote growth of selected RBP null mutants. All four RBPs chosen for validation of the knockout screen and further characterisation were grown independently to assess promastigote growth. All cell lines were grown in triplicate in M199 medium. Statistics represent an unpaired t-test comparing the mean of the control cell lines to the mean of the null mutant lines at each time point with a Bonferroni correction for false discovery. Significant values are marked where * = P≤ 0.05, ** = P≤0.01, *** = P≤0.001, **** = P≤ 0.0001. Statistics are shown at each time point from cell lines with the highest to lowest cell count at that point from top to bottom. All data tested but non-significant results not shown on graphs. A) LmxM.15.0130, B) LmxM.05.0850, C) LmxM.25.0290, D) LmxM.24.1570.

A)


Created with SnapGene®
(7155 .. 35) 1037.LmxM.15.0130_Up_F
PGL2737+ Nt-3HA
7171 bp
f1 ori
M13 fwd
Actin 5'
PuroR
Ald 3'
PGKB 5'
3x HA
mCherry
3xMyc
Cf PGKA/B
PuroR
Ald3'UTR (short)(1)
SP6 promoter
lac promoter
ori
AmpR
AmpR promoter
1307.LmxM.15.0130_Nt3HA_R (1573 .. 1587)

B)


Created with SnapGene®
(7155 .. 35) 1037.LmxM.15.0130_Up_F
PGL2737+ Nt-3HA
7171 bp
f1 ori
M13 fwd
Actin 5'
PuroR
Ald 3'
PGKB 5'
3x HA
mCherry
3xMyc
Cf PGKA/B
PuroR
Ald3'UTR (short)(1)
SP6 promoter
lac promoter
ori
AmpR
AmpR promoter
3HA-mCherry_R (2271 .. 2304)

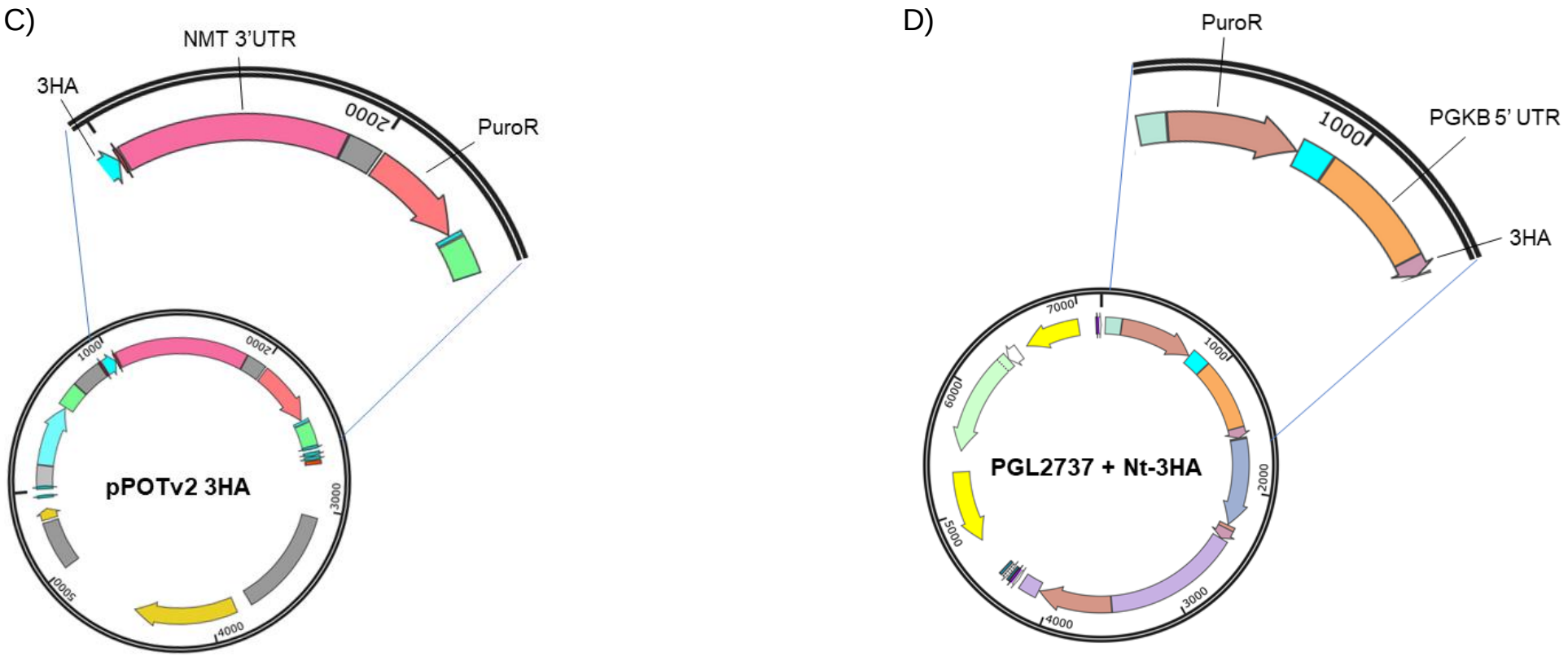


Figure 5.5. Production of tagging constructs. A) The plasmid pGL2737 was edited to contain a triple HA epitope tag for N-terminal protein tagging. Using a reverse primer in the unique linker region, a construct can be amplified to add only the HA epitopes instead of an epitope-fluorescent protein fusion as in B. C) Another construct with a pPOTv2 backbone was produced and adapted for C-terminal triple HA-tagging of RBPs. The amplified tagging construct is shown magnified and labelled. D) The N-terminal tagging construct amplified from the construct shown in A/B is shown here for comparison. Tagging constructs shown in C/D were flanked by 30bp homology regions to the gene of interest that were introduced in the primers. Once transfected, the constructs were inserted either upstream (N-terminal) or downstream (C-terminal) of the CDS.

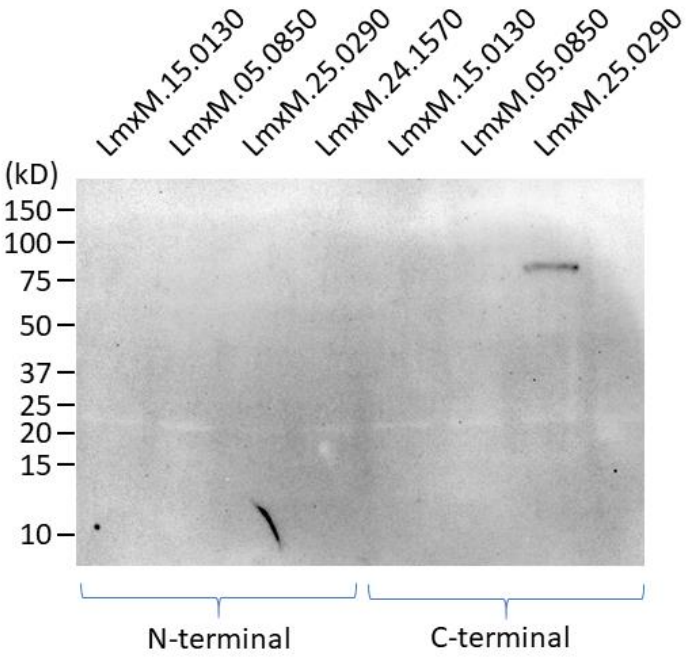


Figure 5.6. Double selection tagging attempted in populations. Selection with Blasticidin and Puromycin was used post nucleofection of both N- and C-terminal triple HA tagging cassettes. The resulting populations were passaged once and proteins were extracted in Laemmli buffer (Laemmli, 1970). $1x10^7$ cells were loaded per well and transferred to PVDF membrane. The resulting western blot is shown using a mouse anti-HA antibody and visualised with a HRP secondary on a Chemidoc (Biorad) transilluminator.

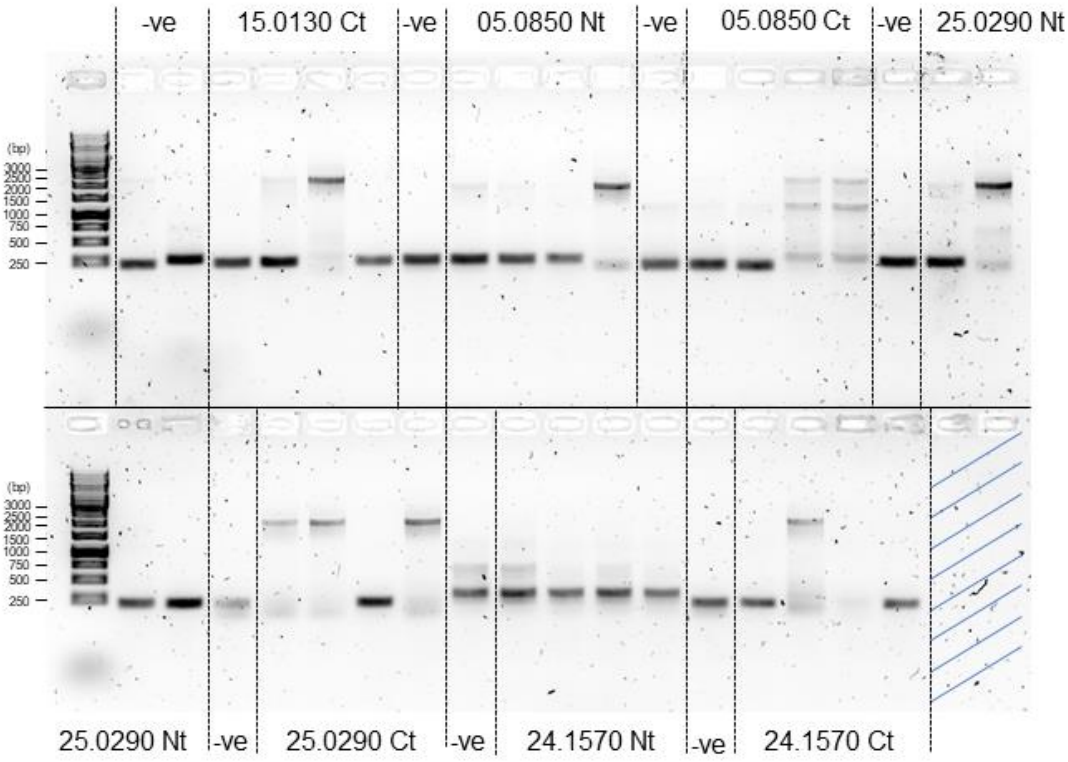


Figure 5.7. PCR screening of HA-tagged RBP clones. JM651 cells were nucleofected with triple HA tagging cassettes containing *PUR* and selected with Puromycin for two weeks in 96-well plates until clones could be transferred to flasks. PCR was carried out using DNA extracted from cloned, nucelofected *L.mexicana* as a template. Negative control PCRs (-ve) show the band corresponding to unsuccessful tagging using JM6571 DNA as a template. Primers upstream and downstream of the integration site were used, producing a larger amplicon upon successful integration.

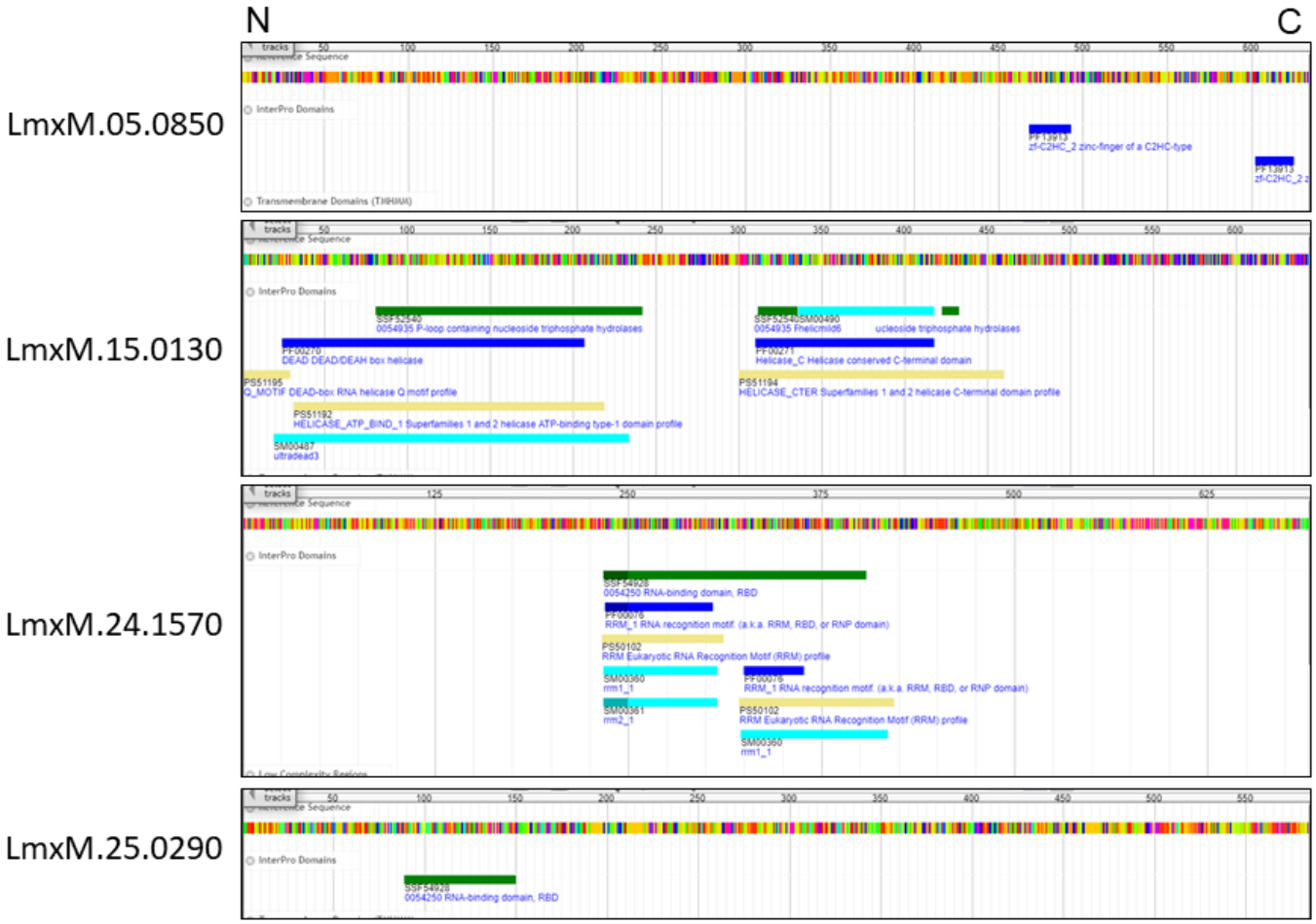
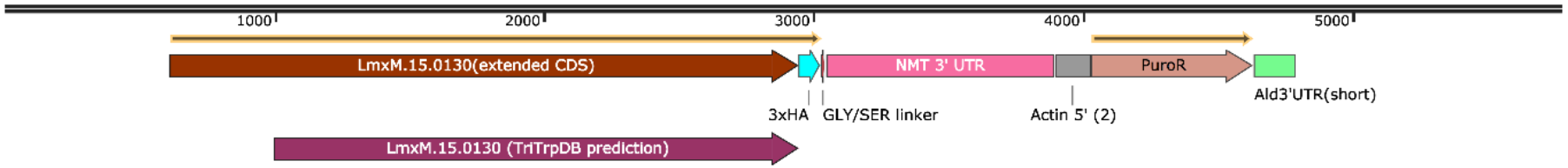
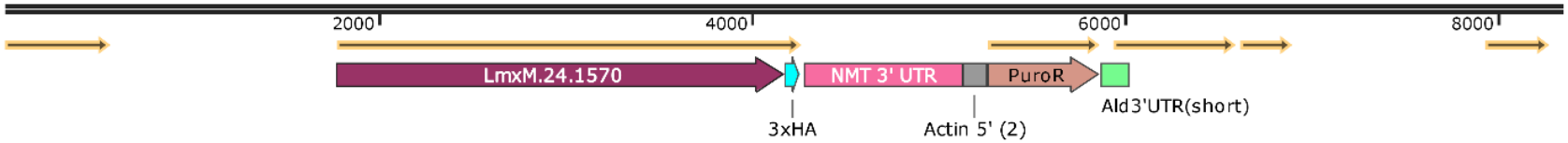


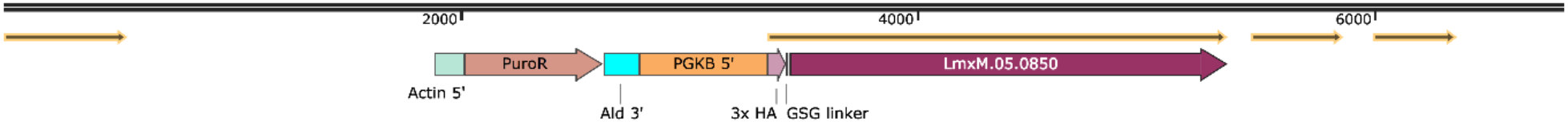
Figure 5.8. Location of RNA-binding domains. Full amino acid sequences from TriTrypDB are presented with annotations showing the location of predicted protein domains.



LmxM.15.0130Ct3HA_Puro
5770 bp



LmxM.24.1570Ct3HA_Puro
8328 bp



LmxM.05.0850Nt3HA_Puro
6821 bp

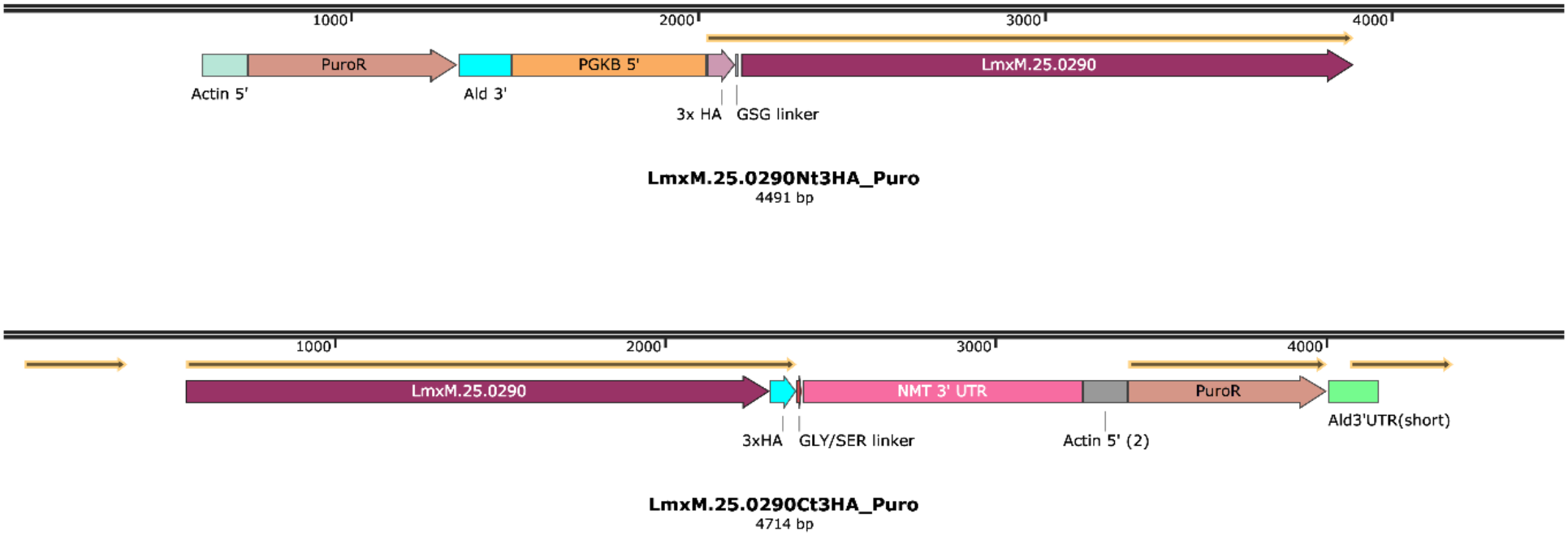


Figure 5.9. Maps of RBP tagging constructs. RBP tagging constructs, integrated successfully and isolated as clones in 4.5.3.1, are shown with annotations of all core parts. The open reading frame (ORF) containing the gene of interest is shown with thin green arrows highlighting translated regions. For C-terminal constructs the *PUR* gene is in the same ORF as the CDS but for N-terminal constructs it is on a separate ORF so is not shown as translated. The predicted CDS of LmxM.15.0130 from TriTrypDB is shown below the extended CDS that was targeted in tagging attempts.

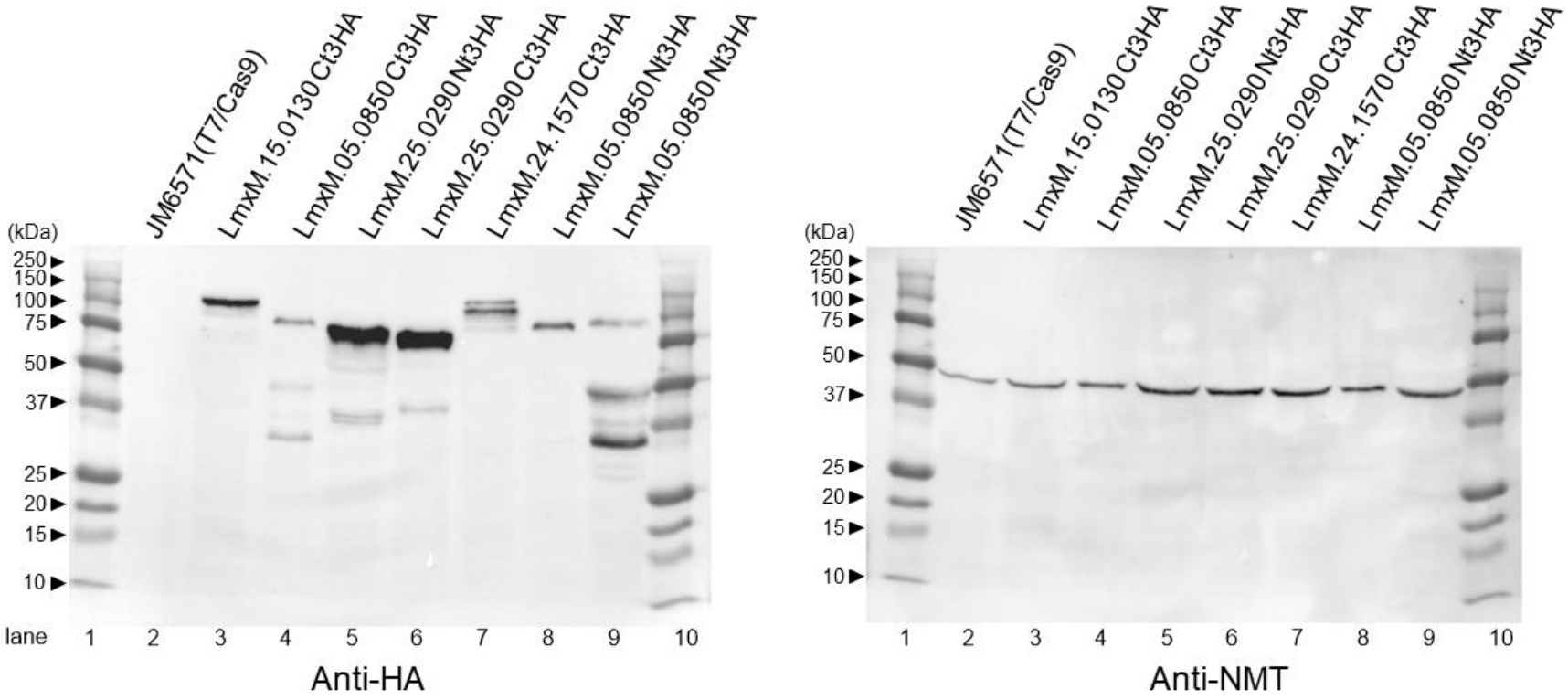
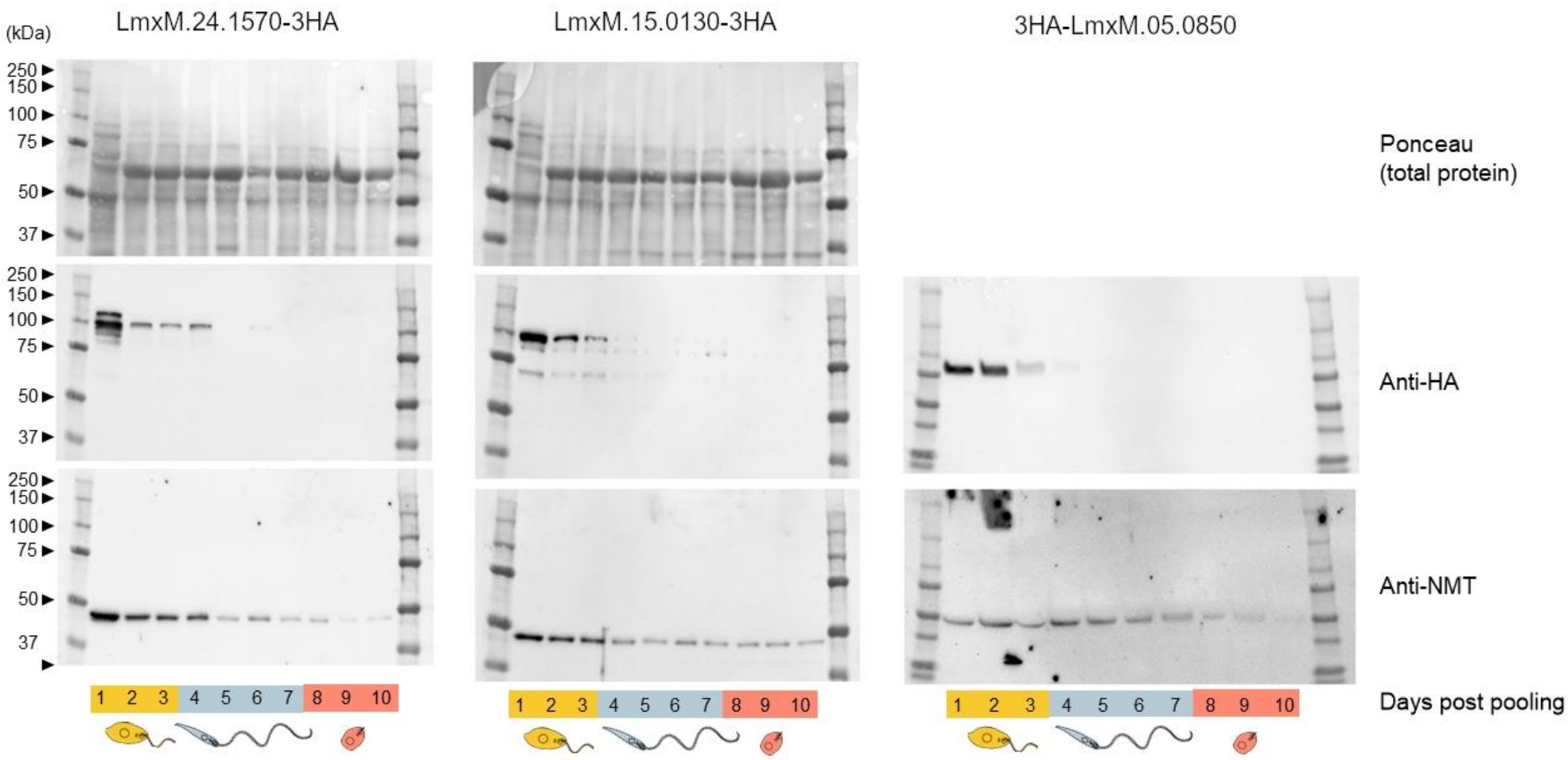


Figure 5.10. Visualisation of HA-tagged RBP candidates. Each of the successfully tagged RBP clonal cell lines was grown in individual flasks and protein was harvested in log phase and separated on a SDS PAGE gel with 1×10^7 cells per lane. Anti-HA western blots were performed as before, followed by stripping of the membrane which was re-probed using anti-NMT primary antibody and visualised again using HRP secondary antibody and a Chemidoc (Biorad) transilluminator. LmxM.05.0850 cells extracted using weaker and stronger Laemmli buffer are shown in wells eight and nine respectively.


LmxM.24.1570-3HA
LmxM.15.0130-3HA
3HA-LmxM.05.0850
(kDa)
250
150
100
75
50
37
Ponceau
(total protein)
Anti-HA
Anti-NMT
1 2 3 4 5 6 7 8 9 10
Days post pooling

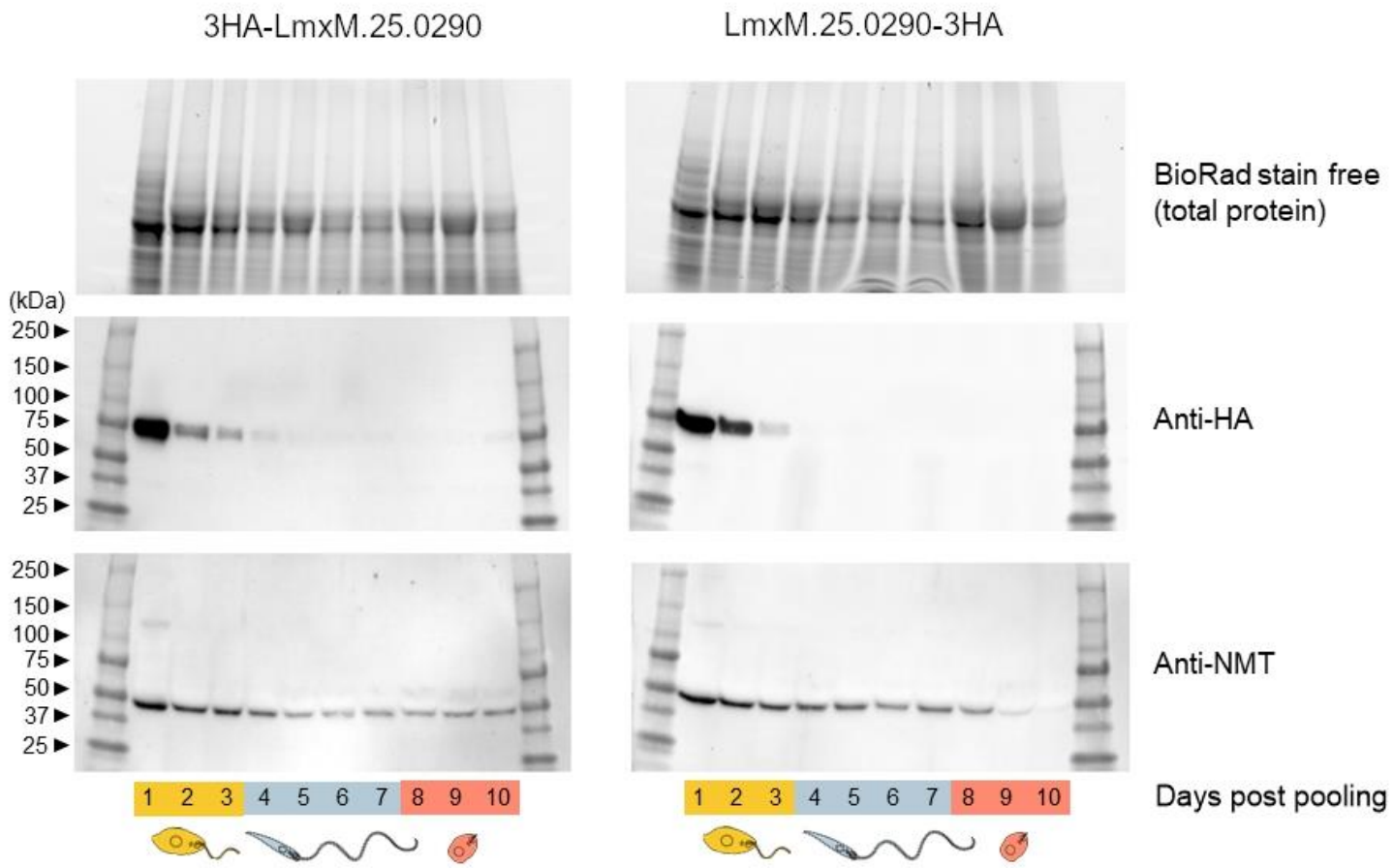


Figure 5.11. Expression profiles of RBP candidates in three lifecycle stages. All five HA-tagged RBP candidate cell lines were grown in Grace's media for seven days before being transferred to 37°C to differentiate into axenic amastigotes. Protein samples were taken at 24h intervals to assess changes in protein expression. 7.5×10^6 cells were loaded per well in an SDS PAGE gel and western blots were carried out as in Figure 5.10. Visualisation of HA-tagged RBP candidates. Log phase = yellow, stationary phase = blue and amastigotes = red.

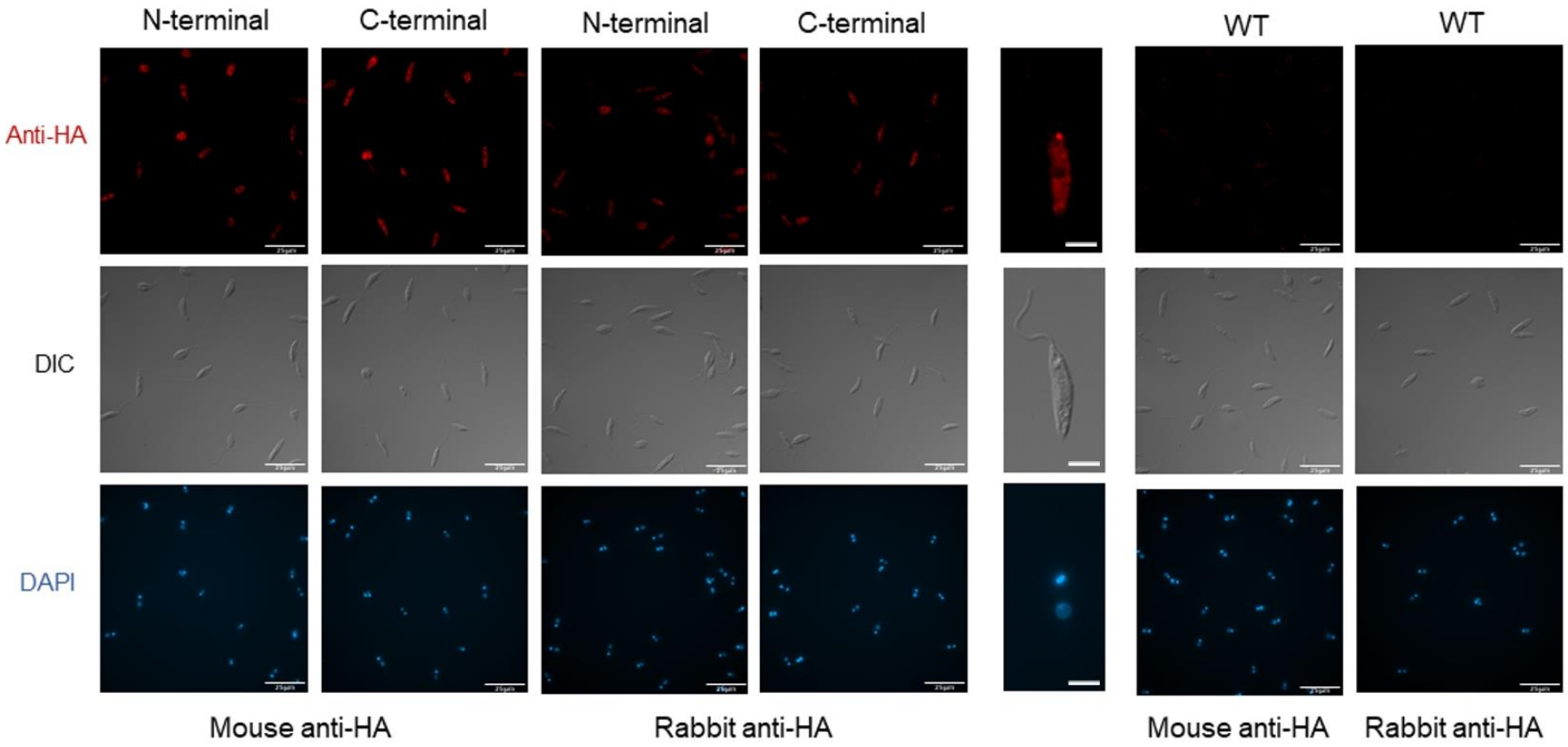


Figure 5.12. Visualising LmxM.25.0290 subcellular localisation. Log phase HA-tagged LmxM.25.0290 cells were fixed in 4% paraformaldehyde and permeabilised before primary anti-HA antibody incubation followed by secondary Alexafluor antibody incubation. After application of DAPI stain, slides were produced and imaged using a Zeiss Axiobserver microscope. Both N- and C-terminally tagged LmxM.25.0290 showed a distinct signal above background levels distributed evenly in the cytoplasm, with some cells showing posterior foci near the base of the flagellum. The signal seen was also similar when detected with two different primary antibodies (supplementary table 1, rabbit: 26183 and mouse: A190-108A). WT cells are JM6571 (T7/Cas9) *L. mexicana*. Scale bars = 25µm for full images and 5µm for enlarged images.

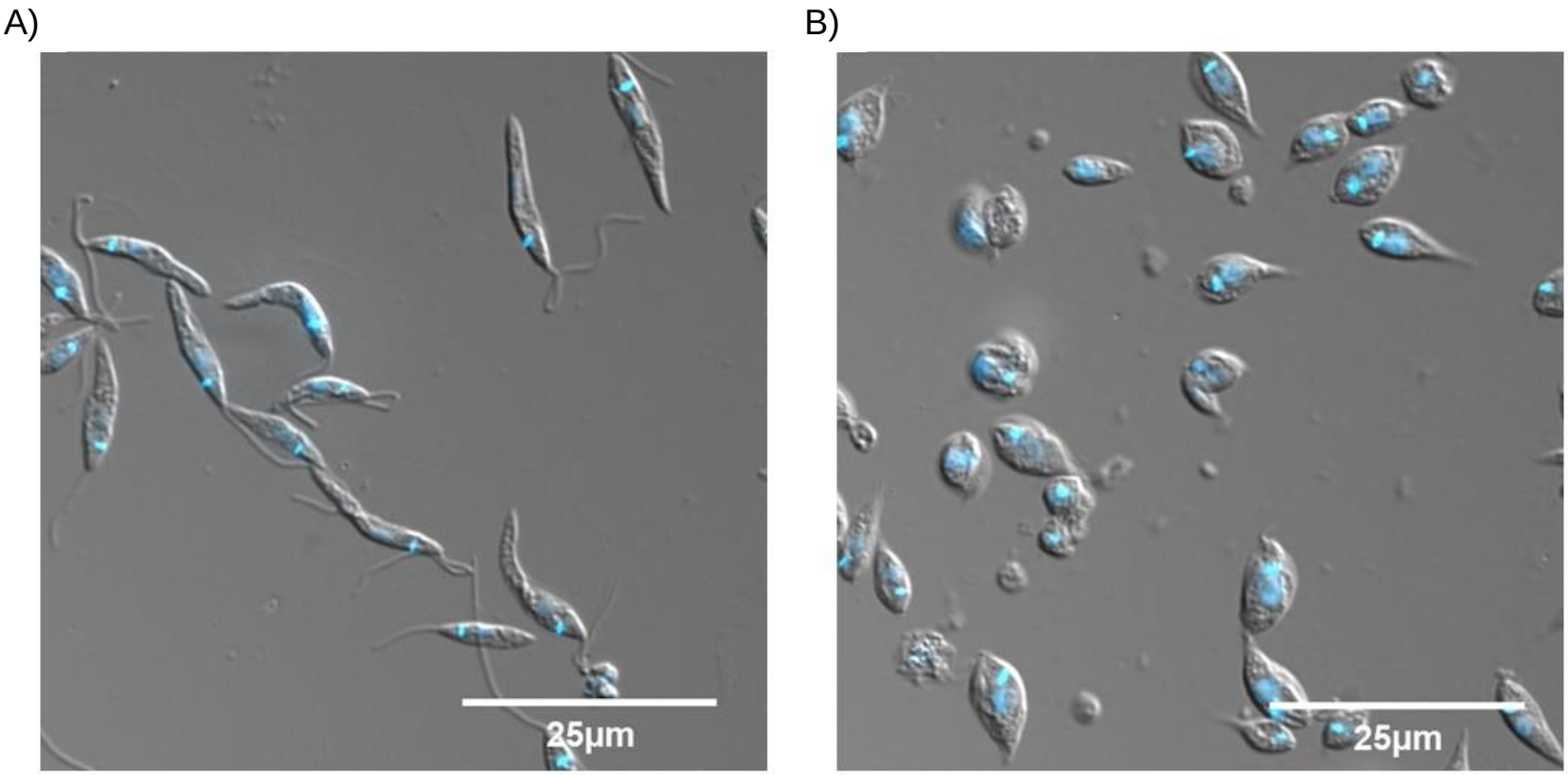


Figure 5.13. Morphological phenotype of LmxM.05.0850 null mutants. The parental T7/Cas9 expressing cell line, JM6571 (A), and Δ*LmxM.05.0850* (B) were grown for three days before cells were harvested from both lines and fixed, stained with DAPI and mounted onto slides. Images were taken on a Zeiss Axiobserver at a magnification of 1000×.

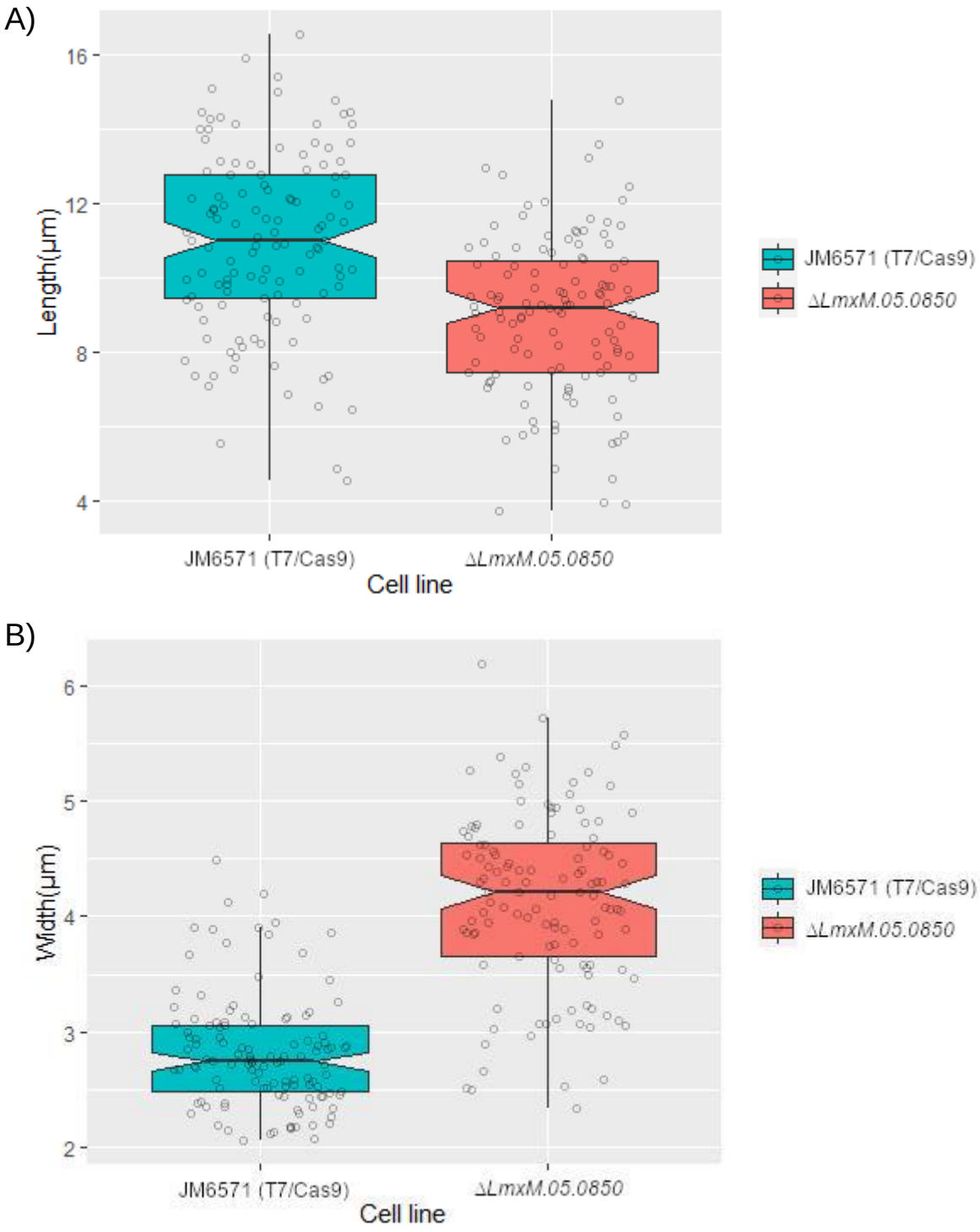


Figure 5.14. Quantification of Δ*LmxM.05.0850* morphological phenotype. Fixed cells from both JM6571 and Δ*LmxM.05.0850* cell lines mounted on slides as in Figure 5.6.2 were measured using ImageJ for comparison. Lengths (A) were measured using the segmented line tool from the base of the flagellum to the cell apex with all lines drawn passing through the nucleus. Cells that did not lie flat on the plane of view were excluded from counts. Cells were counted from multiple fields of view from multiple replicates. Widths (B) were measured at 90° to the length measurement at the widest point of the cell. Data rearrangements, boxplots and statistics were produced using RStudio. A Welch two-samples t-test showed that the difference in length was statistically significant: $t(233.5) = 6.65$, p-value = 2.03×10^{-10}. To compare widths, an independent 2-group Mann-Whitney U Test was carried out with the results as follows: $U = 1194.5$, p-value < 2.2×10^{-16} suggesting a significant difference.

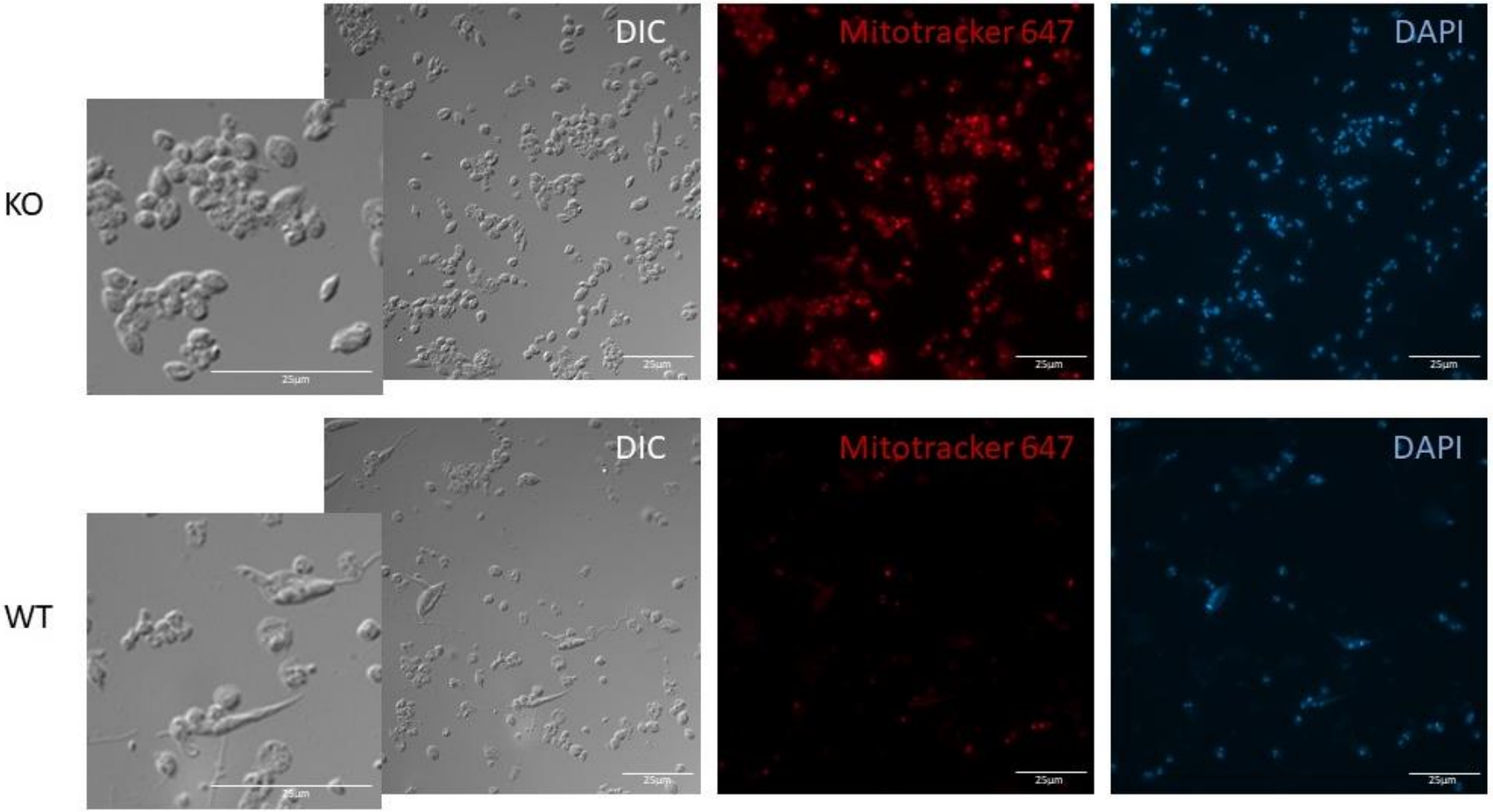


Figure 5.15. Amastigote-like cells forming at 26°C in Δ*LmxM.05.0850* lines. Cells were left in M199 media at 26°C for three weeks before incubation with Mitotracker 647 to stain active mitochondria. After fixation in 4% paraformaldehyde, cells were mounted on slides and imaged on a Zeiss AxioObserver at 1000×. Scale bars represent 25µm.

5.8 Chapter 5 discussion

5.8.1 Sequencing of RBP null mutant lines

The results of the RBP bar-seq screen reveal distinct phenotypes linked to the deletion of *Leishmania mexicana* RBPs. However, when investigating the phenotypes of null mutants, it is crucial to prove that the effects seen are a direct consequence of the single target gene being disrupted. This is especially the case for clonal populations as any off-target effects or compensatory mutations would be present in all cells. The most commonly used strategy is to add the gene of interest back into the knockout cells and test for phenotypic recovery. In *Leishmania* this has previously been achieved through introduction of an episomal expression vector or integration of a stable expressing construct into a repetitive region of the genome such as the ribosomal locus ((Mißlitz *et al.*, 2000; Tetaud *et al.*, 2002)). When add-backs are successful, they provide confirmation that the target gene is the sole cause of a given phenotype. However, there are many reasons why addbacks can result in false negative results. One reason is that not all phenotypes are reversible after reintroduction of the gene of interest. Many RBPs regulate pathways in a tightly controlled manner that depends on other specific environmental or subcellular events. Morphological phenotypes can be especially hard to reverse using addbacks due to the cascade of successive processing stages that often underpins morphological changes. Another factor that can produce false negative results is a mismatch in expression levels between the addback gene and the native gene. High copy number of transfected episomes can dramatically overexpress the addback gene which may change the phenotype rather than returning it wildtype. This problem can also occur when integrating constructs into the ribosomal locus due to the highly active ribosomal promoter. Despite these issues, add-backs are still the most commonly used and most robust method for validating knockout cell lines in kinetoplastids. Variable add-back expression levels can be largely solved by returning the gene of interest to its original locus using the CRISPR/Cas9 system. However, the design and implementation of this approach can be time consuming for a large number of target genes as it is necessary to optimise them individually.

Recently, the reduced cost of whole genome sequencing has provided another option for validating the correct gene deletion in knockout cell lines. Viewing the whole genome sequence provides a huge amount of data about the cell line that is otherwise hard to access. By mapping sequencing reads to a known genome, the number of reads mapped to the replaced CDS can be compared with the surrounding regions. A lack of reads in the CDS region indicates successful knockout, as seen in Figure 5.1. Although not all the cell lines in the screen were sequenced, all 13/28 showed evidence of specific and complete replacement of the target gene with the donor DNA. This suggested that the method used for producing knockout clones was robust and that the PCR screening was sufficient for detecting CDS removal. Having access to the full pairwise sequencing data meant off target integrations could be investigated by tracking paired read partners. Because each read pair has a unique barcoded linker added during PCR amplification for sequencing, forward and reverse reads for the same template region can be matched in the final data. If there was only integration of the

donor DNA in the correct location then sequencing reads mapped to the 3'UTRs in the donor cassette should only be found at the locus of interest and their endogenous location. Off-target integrations would be detectable in these clonal populations with reads mapping to both the endogenous location and the locus of interest as well as the off-target locus. As no off-target integrations were found for almost half of the screened cell lines, it can be assumed that the CRISPR/Cas9 RBP deletions were highly accurate overall. In addition to confirming the accurate replacement of target RBPs, the read coverage of the *BSD* and *PUR* gene cassettes used to replace the RBP of interest revealed multiple integrations likely occurred at the intended locus. Despite this, smaller off target effects such as SNPs and indels introduced by Cas9-facilitated off-target double stranded breaks can't be ruled out as they were not actively searched for across the sequencing data. In order to detect these small changes, as well as any potential compensatory mutations, sequencing with a greater read depth across multiple null mutant clones would be required. If compensatory mutations have occurred they can distort the phenotypic data seen in screens or assays of these lines. Further investigation of these possibilities is underway for several RBP null mutants including Δ*LmxM.05.0850*, Δ*LmxM.15.0130*, Δ*LmxM.24.1570* and Δ*LmxM.25.0290* where sequencing at greater read depth of several clones (Δ*LmxM.05.0850* and *LmxM.25.0290*) will be searched for evidence of compensatory mutations by the Jeffares lab.

Whole genome sequencing reads across *BSD* and *PUR* regions explain the greatly increased barcode counts of some cell lines immediately after pooling and justify day zero normalisation of the bar-seq data as the best method for representing the relative fitness of different null mutants. High levels of genomic plasticity have been previously described in *Leishmania* and other trypanosomatids (El-Sayed *et al.*, 2005; Ivens *et al.*, 2005; Rogers *et al.*, 2011; Reis-Cunha, Valdivia and Bartholomeu, 2017; Sinha *et al.*, 2018). Most chromosomes will have two alleles of each gene on average but, in *Leishmania,* copy number variations and aneuploidy are common. For example some chromosomes are commonly aneuploid, for example the tetraploid chromosome 30 in *Leishmania mexicana* M379 (predecessor of JM6571) (Rogers *et al.*, 2011). All *Leishmania* chromosomes show variation between cells in a population and can also have variable copy number within a chromosome (Sterkers *et al.*, 2011; Damasceno *et al.*, 2020). It is likely that this natural variation accounts for the odd ratios of *BSD* to *PUR* reads for some cell lines as the CRISPR/Cas9 system is highly efficient and the donor DNA is in excess. Read coverage for *BSD* and *PUR* compared to the average reads for the relevant chromosome (containing the RBP of interest) added to one, confirming again that off-target integrations did not explain the odd drug resistance read ratios.

5.8.2 RBP null mutants selected for further characterisation

Following the results of the bar-seq screen and validation of the RBP null mutant cell lines through DNA sequencing, four candidate genes were selected for validation of promastigote phenotypes. As explained in the introduction to Chapter 4.10, there is competition between cell lines in a pool. Also, because of the progressive nature of the screen, defects seen in preceding stages will have an impact

on the number of parasites (and so the number of barcodes reads) in later stages. If it were possible to reliably produce RBP null mutants in amastigotes, the results of the screen could be improved by subsequently starting from this stage and differentiating back into procyclic promastigotes before differentiating to metacyclic promastigotes. However, transfections across multiple membranes to reach intracellular amastigotes are inefficient and the cell count of null mutant lines being extracted from mouse footpads at six weeks post-infection is highly variable. This variation would lead to the majority of RBP null mutants not being present in subsequent back-transformation assays.

To validate the bar-seq screen as a method for identifying phenotypic defects in *Leishmania* due to the deletion of RBPs, the four selected cell lines were grown individually and growth was compared to a parental cell line (JM6571). For reasons discussed in Chapter 3.10, the null mutant parasites included in the screen were diluted in 96 well plates to produce clonal or near clonal parasite lines. As each clone only represents a number of CRISPR/Cas9 integration events equal to the number of alleles of the target gene they are not as representative as a mixed population with thousands of different individual integration events. However, the sequencing shown in section 5.2 does demonstrate the precision of the CRISPR system used, with no off-target integrations of the donor cassettes detected in 13/28 clones. To increase confidence in the reliability of the results, multiple clones were used (when available) when producing individual growth curves. It should be noted that Δ*LmxM.05.0850* was cloned after several passages through serial dilution so while the lines are likely to be clonal, they do not necessarily represent different integration events. Although all four RBPs are poorly characterised, available background information is discussed below.

The first of these was LmxM.15.0130. This protein can be classified as a DEAD-box helicase, retaining some of the key conserved features of this family such as the motif LVLDEADRML, which is present with two amino acid substitutions (LVLDE**C**D**K**ML) in the core helicase domain (supplementary Figure 7.1a). It is unusual amongst *Leishmania* DEAD-box helicases for the DEAD motif itself to contain a divergent amino acid. Q-motifs were also found intact in the core helicase domain. Downstream of these elements there is evidence of other conserved motifs in the helicase C-terminal domain (supplementary Figure 7.1a) (Caruthers, Johnson and McKay, 2000; Schütz *et al.*, 2010). When the full amino acid sequence was aligned to other DEAD-box helicases in *L. mexicana*, as well as some well-studied helicases from humans, it most closely grouped with human DDX27, with the next closest ortholog being DDX6 (Figure 7.1b). This is also reflected in BLASTp searches in the human genome using the whole amino acid sequence for LmxM.15.0130. Human DDX27 matched with 86% cover and 41% identity compared to 73% cover and 26% identity for DDX6. Due to the vast evolutionary divergence, mapping the true relationship between the *L. mexicana* DEAD-box helicases and those in better characterised model organisms is a difficult task requiring a more thorough bioinformatic and experimental investigation.

The second protein chosen was LmxM.05.0850. This protein has two C2HC zinc finger domains towards its C-terminus. Unlike other zinc finger domains such as CCCH that are commonly found in RNA-binding proteins, the C2HC domain is associated with E3 ubiquitin ligases. Despite having

no obvious RNA-binding domain, this protein was isolated in the XL-RBPome and its interesting phenotype and its reduced fitness in the mouse model of infection make it a target worth pursuing. Due to its relative enrichment in the XL-RBPome compared to the whole cell proteome it is likely to be at least associated with mRNA regulatory complexes if it is not a direct RNA-binding protein. No references to studies on a closely related protein could be found with the top BLASTp (Altschul *et al.*, 1990; Oyama *et al.*, 2008; Agarwala *et al.*, 2016) hits in *S.cerevisiae, D.melanogaster* and *H. sapiens* all being uncharacterised proteins containing the same domain. The closest studied protein appears to be the mammalian protein RNF125 which acts as an E3 ligase where the C2HC domain facilitates interaction with the E2 enzyme UbcH5a (Bijlmakers *et al.*, 2016)

Another protein selected, LmxM.25.0290, is the direct ortholog of RBP43 in *T. brucei*, which has been isolated in pulldowns of various factors involved in the cap-binding translation initiation complex. Close interactors appear to be EIF4G1, PABP2 and EIF4E5 (Freire *et al.*, 2014; Zoltner *et al.*, 2018; Clayton, 2019). Despite being isolated several times, RBP43 appears not to have been studied directly to date and there is no available data about this protein in *Leishmania* parasites. BLASTp results using the *L. mexicana* RBP43 sequence as a query identified no significant similarity using default parameters in the *H. sapiens, D. melanogaster* or *S. cerevisiae* genomes. RBP43 has a distinctive RRM domain in the N-terminal region linking it clearly with RNA-binding functions.

The final protein, LmxM.24.1570, has been named double RRM binding domain protein 13 (DRBD13) in *T. brucei* after it was predicted as an interactor of RBP6 and studied using mRNA immunoprecipitations (Najafabadi *et al.*, 2013). DRBD13 was found (along with DRBD12 and RBP6) to interact with adenylate/uridylate-rich elements (AREs) that play a major part in controlling lifecycle stage regulation. It was also shown that DRBD13 directly regulates RBP6 and has a role in regulating essential trypanosome coat proteins (Jha *et al.*, 2015). DRBD13 appears to be distantly related to ELAV-like proteins which were first characterised in *D. melanogaster* as tissue specific RBPs involved with the regulation of pre-mRNA splicing (Koushika, Lisbin and White, 1996).

For three out of the four of these cell lines, the phenotypes matched those observed in the knockout screen, with reduced barcode counts corresponding to reduced growth compared to the parental cell line (JM6571). For Δ*LmxM.15.0130*, a strong growth phenotype was also detected before the screen was carried out (using independently produced clones with no barcode), suggesting that this result is repeatable. The fact that Δ*LmxM.24.1570* showed a decrease in cell count during stationary phase, rather than an increase, appeared to contradict the findings of the screen. The dramatic increase in barcode reads in this cell line in stood out against all other cell lines at the stationary stage of the screen but this is likely to be an artefact of barcode sequencing and not indicative of a true biological effect on cell number.

5.8.3 Tagging and subcellular localisation of *Leishmania* RBPs

All four RBPs selected showed a phenotype as promastigotes although effects of RBP deletion were also observed later in the life cycle for all of these cell lines. Phenotypes that indicate regulons in the human infectious stages have been disrupted are of the highest importance for improving our understanding of *Leishmania* pathology. Amastigotes are the most obvious stage to study in this sense as they are the primary disease-causing stage and as such, further investigations into amastigote phenotypes are required. However, establishment of infection is relatively poorly understood and the contributions of the metacyclic promastigotes to infection need more thorough investigation. It is important to further understand how metacyclic promastigotes (which are superficially more similar to procyclic promastigotes than amastigotes) can withstand the hostile environment in a mammalian host where their procyclic form progenitors cannot establish an infection. Furthermore, it is important to investigate whether the reductions in barcode reads seen during the stationary phase (seven days post inoculation) of the screen were due to RBPs active during the procyclic stage or later forms such as the infectious metacyclic promastigotes. Visualising protein expression in several lifecycle stages was pursued to further understand the timing of the observed phenotypes.

Very few RBPs have been studied individually in *Leishmania* to date; consequently very few subcellular localisations have been determined. The localisation of a tagged protein in the cell gives crucial insight to the organelles and structures it associates with, possibly leading to new information about its function. For regulatory RBPs this information can reveal which pathways and processes the protein may be regulating. Unlike in *Leishmania*, RNA-binding proteins have been studied more in *T. brucei* and many of them have been tagged due to the genome wide tagging project TrypTag (Dean, Sunter and Wheeler, 2017). Until a project of a similar scale is completed in *Leishmania*, the TrypTag database represents the most information available about the localisation of kinetoplastid RBPs. However, despite often being compared, the large degree of evolutionary separation between *Trypanosoma* and *Leishmania* means that inferring the function of one protein from the function of its ortholog should be treated with caution. The TrypTag localisations of the 67 RBPs included in the bar-seq screen are mostly cytoplasmic (Figure 3.5). Since mRNA is mostly regulated and translated in the cytoplasm or associated with the endoplasmic reticulum, this result is consistent with the inclusion of factors isolated on mRNA from the XL-RBPome. In TrypTag, several factors were successfully tagged at both the N- and C-terminus, allowing for more confidence in their localisation patterns. However, some proteins have different localisations depending on the tag position, which suggests that the tag interfered with the protein function and localisation has changed. Unfortunately, in these cases, determining the localisation of the native protein can be hard. The fact that the tag is a large fluorescent protein may also affect the localisation of many small RBPs as well as affecting their ability to bind RNA. For RBPs in particular, it is preferable to use small tags where possible. Where tagging has failed to produce any healthy clones, it is important to note that the RBP may be difficult to tag in *Leishmania* and hard to study without a specific antibody.

5.8.4 *Leishmania* RBP expression patterns

The expression patterns of RBPs in kinetoplastids are often tightly regulated and many are lifecycle stage specific. With this in mind, it was important to see if protein expression correlated with the growth phenotypes seen in section 5.3. It was also important to understand when these proteins are expressed to optimise further characterisation experiments. As noted previously, pulldowns of mRNA-bound proteins from several different lifecycle stages showed the majority of RBPs in *Leishmania* are most abundant during the procyclic promastigote stages (Figure 3.4). The RBPs chosen for further investigation all had phenotypes that were identified in the promastigote stages during the RBP knockout screen. Although some cell lines (Δ*LmxM.25.0290 and* Δ*LmxM.02.0850*) had defects during stationary phase and after metacyclic purification, it was important to investigate if these effects were due to log or stationary phase protein expression. Protein expression profiles produced for all four candidates showed a predominantly log phase expression pattern, with maximum expression at 24 hours after media inoculation. At this time the majority of the cells are rapidly dividing procyclic promastigotes. This suggests that the phenotypes seen are due to changes early in the lifecycle progression, despite downstream effects in later lifecycle stages. For example, expression of Δ*LmxM.05.0850* was barely detected beyond day four but showed a consistent morphological phenotype throughout promastigote growth.

Whilst the lack of RBP expression in later stages is certainly true to some extent, N-Myristoyltransferase (NMT) levels showed that comparison between promastigote and axenic amastigote samples (Days 8,9 and 10) was not straightforward. When re-blotted with anti-NMT antibody, a signal was seen that reduced from early to late promastigotes and even further in the axenic amastigotes. Protein samples for these blots were loaded with an absolute number of cells per well. NMT has been previously characterised as constitutively expressed, but using lesion-derived rather than axenic amastigotes (Price *et al.,* 2003). The reduced body size and protein expression in metacyclic promastigotes may explain the reduction in total protein (ponceau stain or Bio-Rad stain-free). It is possible that either non-expressing or dead cells increased during stationary phase and weren't detected during counting explaining the reduced levels of NMT. The protocol used for producing axenic amastigotes also may not have produced stable dividing cells. To improve this, culture conditions could be optimised for these cell lines and anti-SHERP or anti-δ-amastin western blots used to confirm the differentiation to metacyclic promastigotes or amastigotes respectively (Coulson and Smith, 1990; Teixeira *et al.*, 1994; Rochette *et al.*, 2005; Doehl *et al.,* 2017). Alternatively, macrophage derived amastigotes could be used in place of axenic. Despite the issues discussed here, the combination of an NMT and a total protein signal can be used to determine that the expression of all four RBP candidates was predominantly during the log phase (procyclic promastigotes).

Some tagged RBPs were seen to exist as several different peptide lengths. LmxM.15.0130-3HA was initially detected between 75 and 100kDa. The consistent appearance of two slightly different molecular masses while analysing expression levels over time may indicate a post-translational

modification of this protein. Further experiments, including more highly mass-separated western blots would be required to confirm this hypothesis. The other RBP that was observed with molecular masses was LmxM.24.1570-3HA, being most obvious at the 24h timepoint where this protein was most expressed. The additional fainter band observed above 100kDa could represent post translational modification. Several lower molecular mass peptides between 100kDa and 75kDa may be naturally occurring products of peptide cleavage or degradation products of the full-size peptide formed during protein extraction.

5.8.5 Morphological phenotypes

Morphological phenotypes can be particularly useful for forming hypotheses about the involvement of a protein in different cellular pathways. For example, it is likely (but by no means certain) that an abnormal flagellar phenotype is caused by the target protein's involvement with pathways relating to the production or maintenance of the flagellar structure. Phenotypes like this could also be linked to reduced energy provision, as producing and maintaining the functional flagellum requires large quantities of ATP. The phenotype of Δ*LmxM.05.0850*, with a greatly reduced flagellum, was the only obvious morphological phenotype to be identified during the production of the null mutant cell lines. It is possible that other less dramatic phenotypes exist and it would be worthwhile pursuing a microscopy screen of all knockout cell lines to systematically identify these. One predicted effect of the increased size of *LmxM.05.0850* null mutants was that they were removed during a metacyclic purification, which depends on sedimentation based on size and density. Another way of identifying parasites with morphological phenotypes would be to assess other cell lines that had reduced barcode counts during the metacyclic purification stage of the screen. Of these, cell lines with no obvious structural abnormality are likely to have difficulty differentiating into or surviving as metacyclic promastigotes.

The manipulation of RBPs in *T. brucei* also resulted in some different, but interesting, morphologies. Perhaps the best described is the 'nozzle' morphology that presents during the overexpression of *Tb*ZFP2 (Hendriks *et al.*, 2001). The phenotype was later reported following overexpression of the related protein *Tb*ZFP3 (Paterou *et al.*, 2009). The phenotype consists of a polar extension of the cytoskeleton posterior to the kinetoplast due to the unregulated extension of existing microtubules. In a similar manner to Δ*LmxM.05.0850*, these cells showed a reduced growth in procyclic stage. The effects seen were also cell cycle stage dependent, with later stages showing other phenotypes relating to abnormal karyotypes or numbers of organelles. While unlikely to be related to the phenotype seen for Δ*LmxM.05.0850*, this example does show how manipulation of an RBP can have lasting consequences on the cytoskeletal structure of the parasite.

Similar morphologies to that of Δ*LmxM.05.0850* have been described elsewhere in *Leishmania* research. Protein kinase MPK3 and MKK null mutants were reported to show very similar morphology with the flagellum reduced to one fifth of the size of the WT cell line (Erdmann *et al.*, 2006). Additional protein kinase null mutants with similar morphologies (especially Δ*stk36*, Δ*ulk4*,

Δ*LmxM.29.0600*, Δ*LmxM.02.0570* and Δ*mpk3*) were identified by Baker *et al.* (2021) and shown to have severely impacted motility, making them unable to colonise the sand fly vector. In general, mutants with impaired motility have been linked to poor success during sand fly colonisation experiments (Beneke *et al.,* 2019). Both MPK3 and MKK are part of a MAP-Kinase signalling cascade that has a role in flagellar morphogenesis (Erdmann *et al.,* 2006). Because of the similar phenotypes, it is worth investigating whether LmxM.05.0850 is involved in the same signalling pathway and whether it is actively regulated by phosphorylation in *Leishmania* parasites. The impact of *LmxM.05.0850* deletion and its effect on the morphology of the parasite has been shown in this study to affect the parasite's success in a mouse model of infection, but not as severely when macrophages are directly infected *ex-vivo*. It is plausible that, whilst differentiation to amastigotes is possible, and the mutants can successfully invade macrophages when directly incubated with them, the lack of motile metacyclic promastigotes perturbs their ability to establish infection in the mice. Alternatively, or in conjunction, other aspects of the host immune system may be able to clear the mutant parasites more effectively than the control cell lines. Either of these scenarios would demonstrate how removal of even procyclic form specific RBPs can have significant deleterious effects in the human infectious stages six weeks after they should have been expressed.

The differentiation to amastigotes and survival of *LmxM.05.0850* null mutants at 26°C is an interesting preliminary result. It may suggest that some similarities of this morphological mutant to amastigote forms, the lack of flagellum and shorter, wider body, are accompanied by underlying physiological changes that enable long term survival and trigger differentiation. However, so far this has only been observed once and needs to be repeated with a range of experimental conditions. For example, a range of incubation temperatures from the 26°C of the sand-fly midgut to the 37°C of a human host should be investigated as well as observations over time to record when the differentiation occurs. Investigation of the effects of pH on differentiation of the Δ*LmxM.05.0850* line compared to WT cells may also yield results, although pH may contribute less to differentiation than changes in temperature (Barak *et al.*, 2005; Alcolea *et al.*, 2010). For confirmation that these smaller, amastigote-like cells are the same as amastigotes described in WT cells, amastin or HASPB levels should be tested for on both the protein and mRNA level (Teixeira *et al.*, 1994; Wu *et al.*, 2000; Rochette *et al.*, 2005; Doehl *et al.*, 2017).

6 General discussion

A primary goal for research into *Leishmania* parasites is to work towards better treatment or prevention of the disease they cause, which affects over a quarter of a million people each year. Crucially, more treatment options and the potential for a *Leishmania* vaccine must be explored in order to improve the quality of life for those exposed to it. In order to do this, we need a much greater understanding of the pathways and mechanisms underlying parasite infectivity, virulence and ability to differentiate to infectious forms. Considering the lack of canonical transcription factors in *Leishmania*, identifying post transcriptional regulators of differentiation or infectivity is crucial to

revealing the most important pathways. The lack of classical transcriptional regulation also makes *Leishmania* parasites a unique model organism for examining the regulatory effects of RNA-binding proteins with more clarity than is possible in many commonly used eukaryotic models. However, when making comparisons between RBPs in *Leishmania* and other eukaryotes it is crucial to consider their ancient divergence from the common ancestor of metazoan organisms. Comparison to other kinetoplastids often yields more obvious parallels, although caution must be taken when assuming similar function for orthologs even within this group. In *T. brucei* several RBPs have been characterised as master regulators of differentiation where ablation or overexpression is sufficient to promote or block differentiation (Hendriks *et al.*, 2001; Hendriks and Matthews, 2005; Kolev *et al.*, 2012; Jha *et al.*, 2015; Alcantara *et al.*, 2018). Several of these proteins have direct orthologs in *Leishmania* (e.g. ZFP3, Alba1 and Alba3, RBP16). However, the vast majority of these have yet to be confirmed as having the same role in these two parasites and it is likely that many do not (or at least their roles differ in major ways). In addition, due to the considerable evolutionary divergence, there are many RBPs in *Leishmania* with relatively low homology to any protein in *T. brucei*, which require investigation for involvement in *Leishmania* specific processes (Figure 3.2). As discussed in Chapter 1, the lack of an RNAi system meant that studying *Leishmania* orthologs of *T. brucei* or *T. cruzi* RBPs has only been possible through overexpression, inducible deletion such as DiCre or by genetically engineering null mutant parasites. Prior to the introduction of CRISPR/Cas9, the production of null mutant cell lines in trypanosomatids was laborious with most studies focusing on a single protein. However, improved CRISPR protocols now allow for knockouts to be produced reliably and quickly (Peng *et al.*, 2015; Sollelis *et al.*, 2015; Zhang and Matlashewski, 2015; Beneke *et al.*, 2017). With this in mind the major goal of this study has been to screen a range of RBPs, both those that have been characterised in related species and novel ones, to further the understand their function. This goal has largely been achieved, with the results of the RBP bar-seq screen shedding light on the stage specific phenotypes of a wide range of RBP null mutants. To date, this is the largest bar-seq screen of RNA-binding proteins in kinetoplastid parasites. As well as the cell lines that were pooled and screened, the many RBP knockout attempts that were unsuccessful identify a large proportion of RBPs that may have essential functions in *Leishmania.* In addition, background information on these RBPs, and many more has been collated and presented graphically to aid future studies (Chapter 3). This includes further exploration of the RBPome data that informed much of the knockout candidate selection for this study. Several RBPs with phenotypes observed in the RBP bar-seq screen were investigated individually, validating the results of the screen and revealing potential avenues for further study.

The findings of the background research into the RBPome as well as the *L. mexicana* genome are consistent with previous research into the lifecycle of these parasites and other kinetoplastids. For example, it is not surprising that the vast majority of RBPs are most abundant in the procyclic promastigotes, where replication, transcription and translation are at their highest level (Kloehn *et al.*, 2015; De Pablos *et al.*, 2019). Because of this, the minority of RBPs that were most abundant in the human infectious stages are of particular interest, both for the bar-seq screen and for further study.

The large divergence in RBP amino acid sequences between kinetoplastids and the eukaryotic model organisms is also to be expected given their early branching from the eukaryotic lineage. One area that hasn't been well addressed by this study are the numerous uncharacterised proteins that were isolated in the XL-RBPome (De Pablos *et al.*, 2019). In depth bioinformatic analysis using hidden Markov models (HMMs) or structural predictions may reveal more information about many of these protein identities. Additionally, the production of more RNA-bound proteomes will allow cross comparison to identify those that are most closely associated with RNA. However, real progress will be made on these targets when the scale of knockout, tagging and cloning procedures can be expanded in these parasites to easily include hundreds of genes. The LeishGEM project (www.leishgem.org) will soon test a high throughput knockout screen in *Leishmania*, the first study to attempt this at the genome wide level. It is likely that this study and others will help to optimise high throughput CRISPR screens in *Leishmania*, potentially with automation of large parts of the screening of knockout clones.

A large proportion of all RBP knockout attempts failed to produce any recoverable clones or populations. In the absence of multiple repeat transfection attempts with different sgRNA cassettes, it is possible that some of these attempts failed due to poor sgRNA targeting or simply that not enough cells were screened. However, there are indications that this is not generally the case, such as subsequent repeats of some selected proteins producing the same results (Table 7.4). Being regulators of many crucial cellular pathways and often having multiple RNA targets, RBPs are very likely to have disruptive effects on the cell when deleted. This is even more likely to be the case in *Leishmania* parasites, where the lack of transcriptional control puts post-transcriptional regulation by RBPs at the forefront of gene regulation. For many RBPs, deletion attempts only resulted in parasites with both drug resistance cassettes integrated as well as additional copies of the CDS. This also likely reflects the essentiality of these proteins as it is improbable that no transfection events resulted in complete removal of the CDS given the large number of cells screened and the efficiency of the CRISPR/Cas9 system used. It is more likely that complete removal of many RBPs causes a significant fitness defect in promastigotes, as removal of key RBPs or transcription factors does in other eukaryotic cells (Koike-Yusa *et al.*, 2014; Joung *et al.*, 2017; Van Nostrand, Pratt, *et al.*, 2020). In this case, and under the harsh double drug selection, most transfectants did not survive the cloning process leaving only the few with extra CDS copies to survive. As discussed in 4.10.3, it appears that the proportion of RBPs that are essential in *Leishmania* based on this screen (34-58%) is larger than that of some other groups such as kinases and DUBs.

While the large proportion of RBPs that could not be deleted are likely to represent many factors that are essential in promastigotes, the remaining successful knockouts mostly have fitness defects at other stages of the *Leishmania mexicana* lifecycle. Screening these remaining proteins to determine their relative fitness in different lifecycle stages was the major accomplishment of this study. Many proteins have been identified for the first time as causing lifecycle specific fitness defects as a consequence of CDS deletion. Absence of many of these RBPs specifically affects the human

infectious stages and prevents growth in a mouse model of infection, implying their involvement in metacylic promastigote or amastigote specific processes. In this sense, RBPs with post-transcriptional *trans*-regulatory function can be a useful tool for uncovering molecular pathways that are crucial for *Leishmania* pathology. Specific reductions in fitness for RBP null mutants after metacyclic purification indicates that either there has been an undetected morphological defect associated with CDS deletion or that there are less metacyclic promastigotes available for purification. The semi-quiescent lesion derived amastigotes described in Kloehn *et al.*, (2015) may explain some of the phenotypic differences in phenotypes between the macrophage infections and mouse footpad infections in the bar-seq screen. The general trend of more pronounced fitness defects in the mouse model of infection may also be due to the added pressure of the host immune system acting on the pool.

Morphological phenotypes are strongly linked to a protein's function and can provide a lot of information when characterised individually. In addition to the continued investigation of Δ*LmxM.05.0850* and its comparison to other lines with similar morphology (Baker *et al.*, 2021), it may be worthwhile to visually screen all lines with low barcode read counts after metacyclic purification to identify any less obvious morphological changes. Additionally, screening of all lines with low metacyclic barcode reads for SHERP expression during stationary phase would identify those with true defects in metacyclogenesis (Doehl *et al.*, 2017). Identifying the differences between procyclic and metacyclic promastigotes regarding infectivity will improve our understanding of the subcellular processes involved in establishing an infection in the mammalian host. By grouping RBPs that negatively impacted fitness in similar lifecycle stages of the bar-seq screen and identifying their RNA-targets it may be possible to find common mRNA transcripts representing proteins that are crucial for differentiation or infectivity. One key area for further investigation that was beyond the scope of this study is to look for the effect of individual RBP deletion on virulence in the mouse model of infection. Identifying novel virulence factors or signalling pathways involved with virulence is a necessary precursor to the design or discovery of new drugs to combat leishmaniasis.

The tagged RBPs, selected for validation of the bar-seq screen, were produced using the PCR-based CRISPR cassette system to use the same sgRNAs for tagging as knockout (Beneke *et al.*, 2017). One advantage of this system is that, in cases were knockout attempts have failed, successful production of tagged clones could validate that the CRISPR/Cas9 system, the transfection and cloning method and the guides themselves are not responsible for the failed knockout attempts. The same tagged RBPs (3HA-LmxM.05.0850, LmxM.15.0130-3HA, 3HA-LmxM.25.0290 and LmxM.24.1570-3HA) were also investigated using RNA-immunoprecipitation to determine which RNA-targets are responsible for the observed phenotypes. However, due to time constraints and the disruption caused by the Covid19 pandemic, the results of these investigations weren't available for analysis for inclusion in this thesis. This and other further investigations of the RNA-targets of the screened RBPs may provide valuable insight into the mechanisms underlying the observed growth and morphological phenotypes. Further RNA-immunoprecipitation experiments on RBPs with a

metacyclic promastigote or amastigote expression pattern may be more likely to identify targets with direct influence on infectivity or virulence so should be pursued. Extensions to this work could include a combined iCLIP-/dCLIP-Seq approach were regulatory protein binding sites within target RNAs are identified as well as the RNAs themselves (König *et al.,* 2010; Marchese *et al.,* 2016; Rosenberg *et al.*, 2017). Through comparison with the XL-RBPome data, these techniques may help to separate proteins that bind RNA directly from those that are proximal interactors in RNA-bound complexes. Further characterisation of the RNP regulatory complexes involving key *Leishmania* regulators is also likely to produce valuable results. The reverse of the CLIP methods mentioned previously can be used to pull down RNA-bound, crosslinked protein complexes once RNA targets have been identified but other methods such as biotin labelling followed by mass spectrometry may also be applied, arguably providing greater depth (Branon *et al.*, 2018; Roux *et al.*, 2018; Geoghegan *et al.*, 2021).

In the future, research carried out to investigate RBP targets of potential interest from the screen can take many forms. While this study has revealed RBPs that may be essential in the infectious stages specifically, it is also important to investigate those RBPs that may be essential in multiple lifecycle stages. Many proteins that were likely to be essential, as the parasites did not tolerate their deletion, may have been important for multiple lifecycle stages but weren't functionally screened as no null mutant lines were available. One method for attempting this would be to perform further rounds of CRISPR based knockout in the amastigote stages that may differentiate between proteins that are essential throughout the lifecycle stages and those that are just essential in procyclic promastigotes. The difficulty with this approach is the challenging nature of achieving high throughput transfections in an intracellular parasite. It is possible that carefully cultured axenic amastigotes would be suitable for this purpose but attempts to date have shown poor results compared to procyclic promastigote nucleofections and would require significant optimisation. This system would still not produce viable cell lines for any proteins that are essential in both procyclic promastigotes and amastigotes so bar-seq screening would not be available for those RBPs. The most elegant solution to this problem would be the implementation of an inducible knockout or knockdown system in *Leishmania*, allowing for the true confirmation of essentiality for many genes. This would also facilitate studies on the effects of RBP depletion at different stages of the lifecycle. One method that has been successfully implemented in *Leishmania* is the destabilisation domain based on the dihyrofolate reductase gene from *Escherichia coli* (Podešvová, Huang and Yurchenko, 2017). While this system does facilitate inducible protein depletion, there are also several caveats. Firstly, for many genes, attempts at destabilising are unsuccessful, possibly because the large tag interferes with protein function or folding causing it to be degraded (data not shown). Secondly, compared to CRISPR null mutants where protein expression is effectively negated, destabilisation will not be as effective at lowering protein levels. Another robust solution that exists for *Leishmania* is the DiCre system (Duncan *et al.,* 2016). By flanking the gene of interest with LoxP sites that recombine upon rapamycin induction, gene deletion can be induced at will. This system has proven highly reliable and has been used to investigate genes that are otherwise hard to study (Santos *et al.*, 2017; Damianou

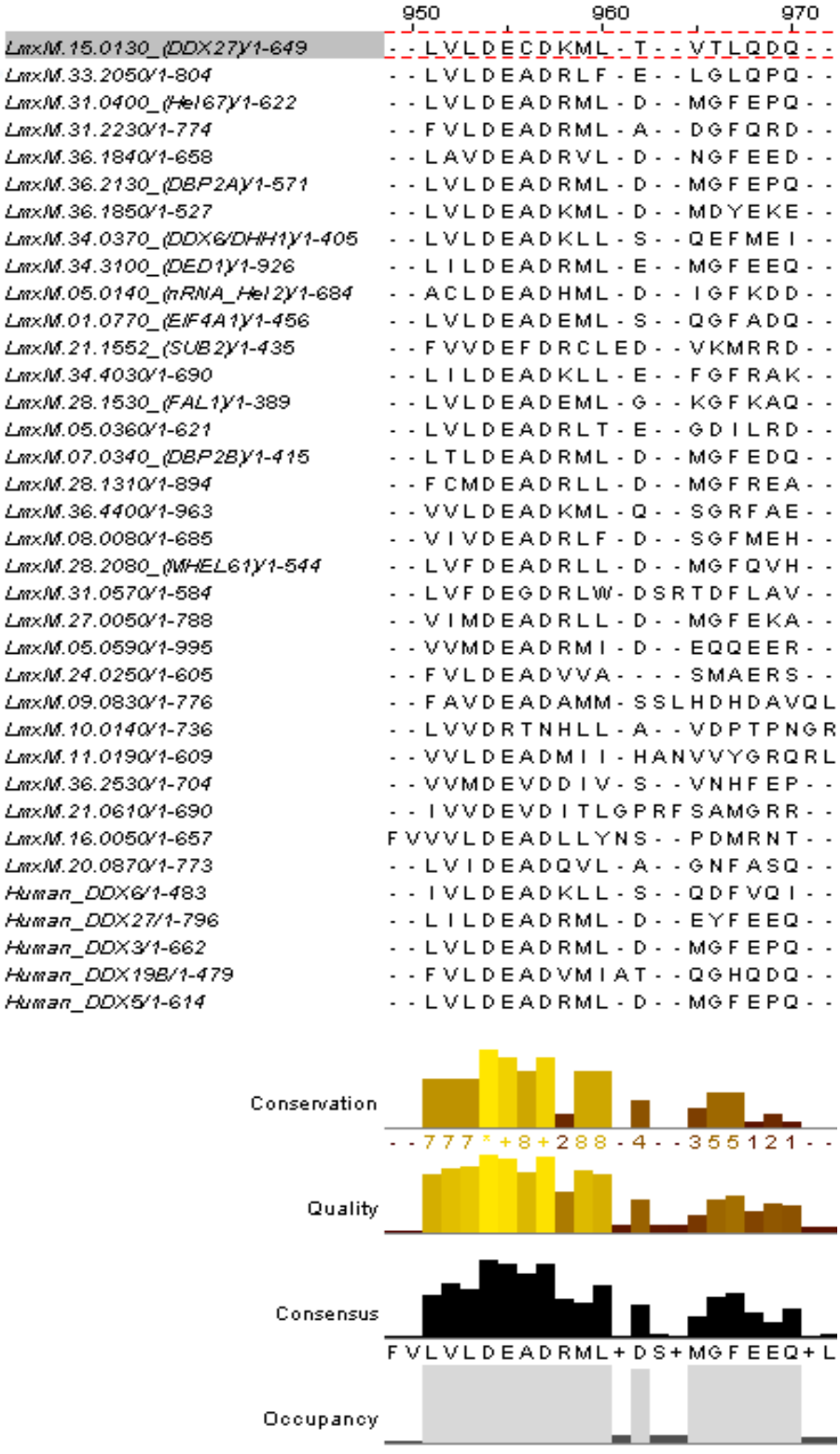
et al., 2020). Some drawbacks of this system are that it is slow to implement for many genes simultaneously, and that rapamycin may interfere with elements of the TOR pathway, although this has not been investigated in *Leishmania* (Barquilla, Crespo and Navarro, 2008; Barquilla *et al.*, 2012). The introduction of LoxP sites and ease of editing multiple proteins has recently been improved (Yagoubat *et al.*, 2020). An ideal solution that enables the study of essential proteins at high throughput would be an inducible CRISPR/Cas9 system that removes the need for introduction of LoxP sites. Recently a system using tetracycline based inducibility has been developed in *Trypanosoma brucei* but hasn't been adapted successfully for use in *Leishmania* (Rico *et al.*, 2018).

Given the range of RBPs included in the bar-seq screen and the fact that tagging primers are available for all 67 knockouts that have been attempted, it is likely that there will be many future research projects that benefit from the work presented here. Overall, our understanding of gene regulation in *Leishmania* is rapidly expanding as the new tools of molecular biology are applied to answering the key questions in this field. For example, since the start of this project, single cell RNA-sequencing technologies have become available and have been applied to kinetoplastids, unlocking a vast quantity of data on thousands of transcripts across several lifecycle stages (Vigneron *et al.*, 2020; Briggs *et al.*, 2021; Louradour *et al.*, 2022). The results of the bar-seq screen presented here and any further investigation of RBP targets compliment these datasets well as they provide evidence for the RBP function, related to stage specific transcripts. With advancing knowledge of the pathways involved in *Leishmania* differentiation, infectivity and virulence, it seems inevitable that, in the near future, the number of available protein targets for novel drugs or vaccines will increase. This will hopefully bring the field closer to the overarching goal of providing significant treatments for those suffering with leishmaniasis, or preventatives for those at risk of developing the disease.

7 Appendices

7.1 Supplementary figures

A)



B)

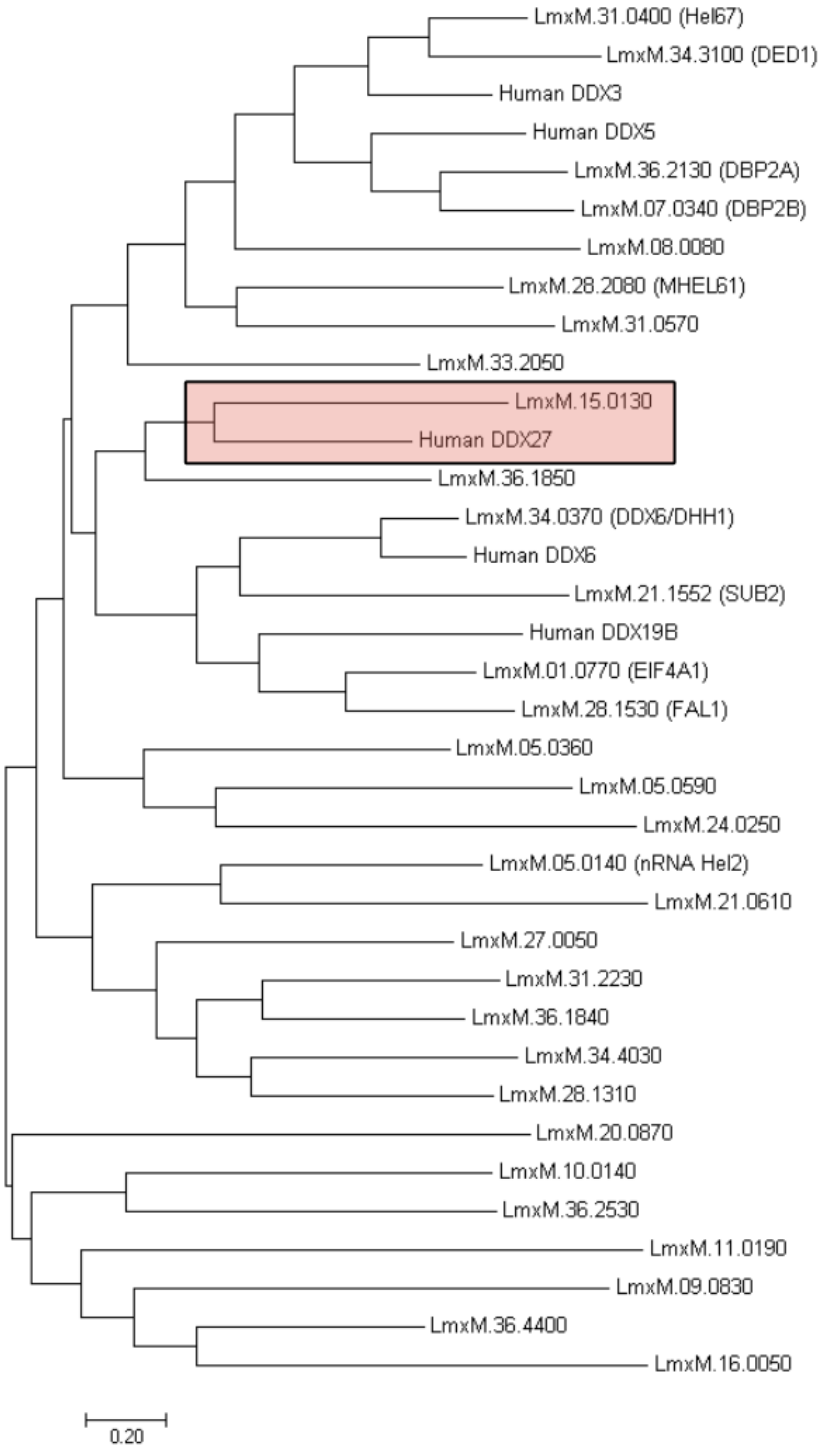
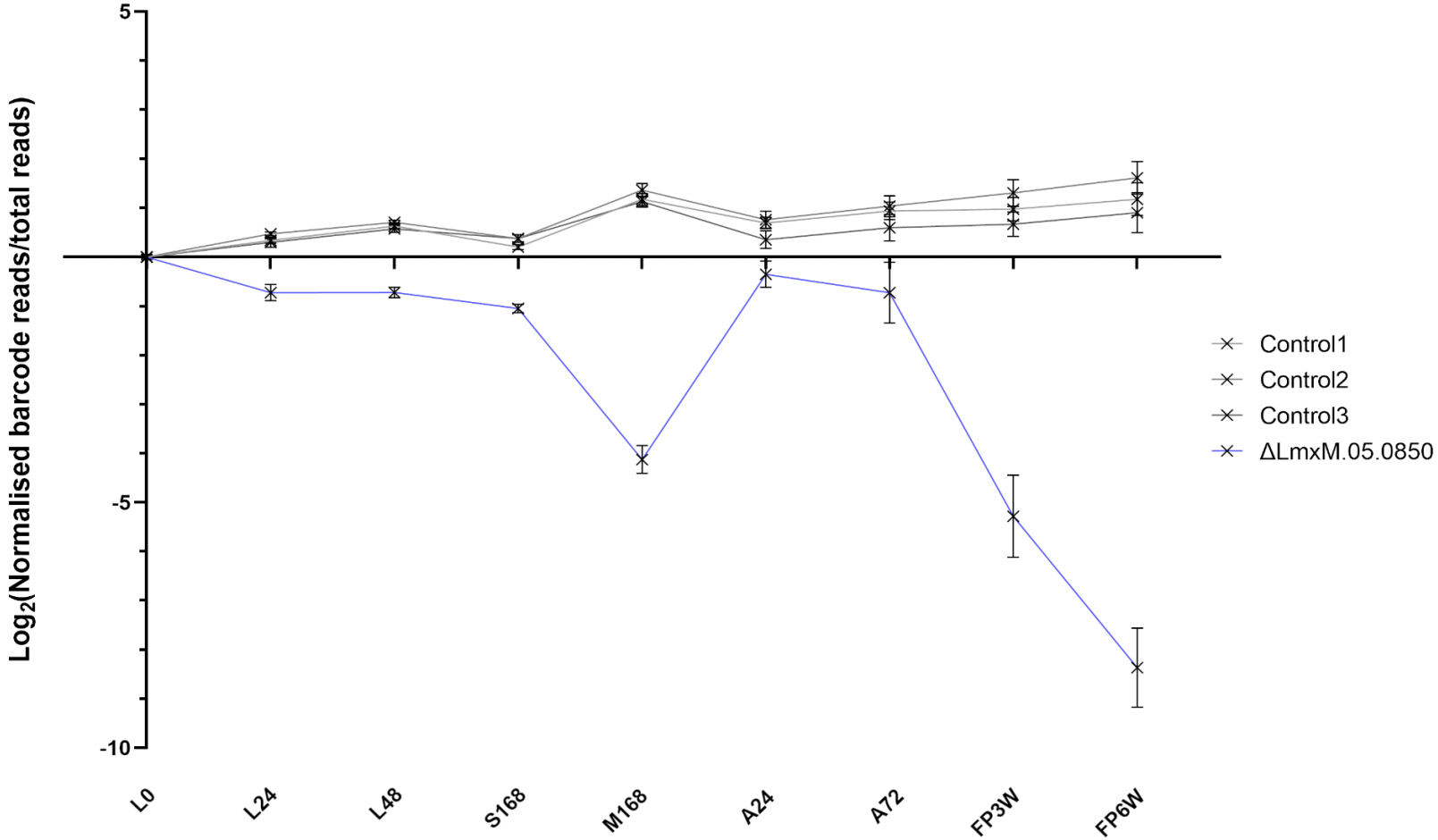
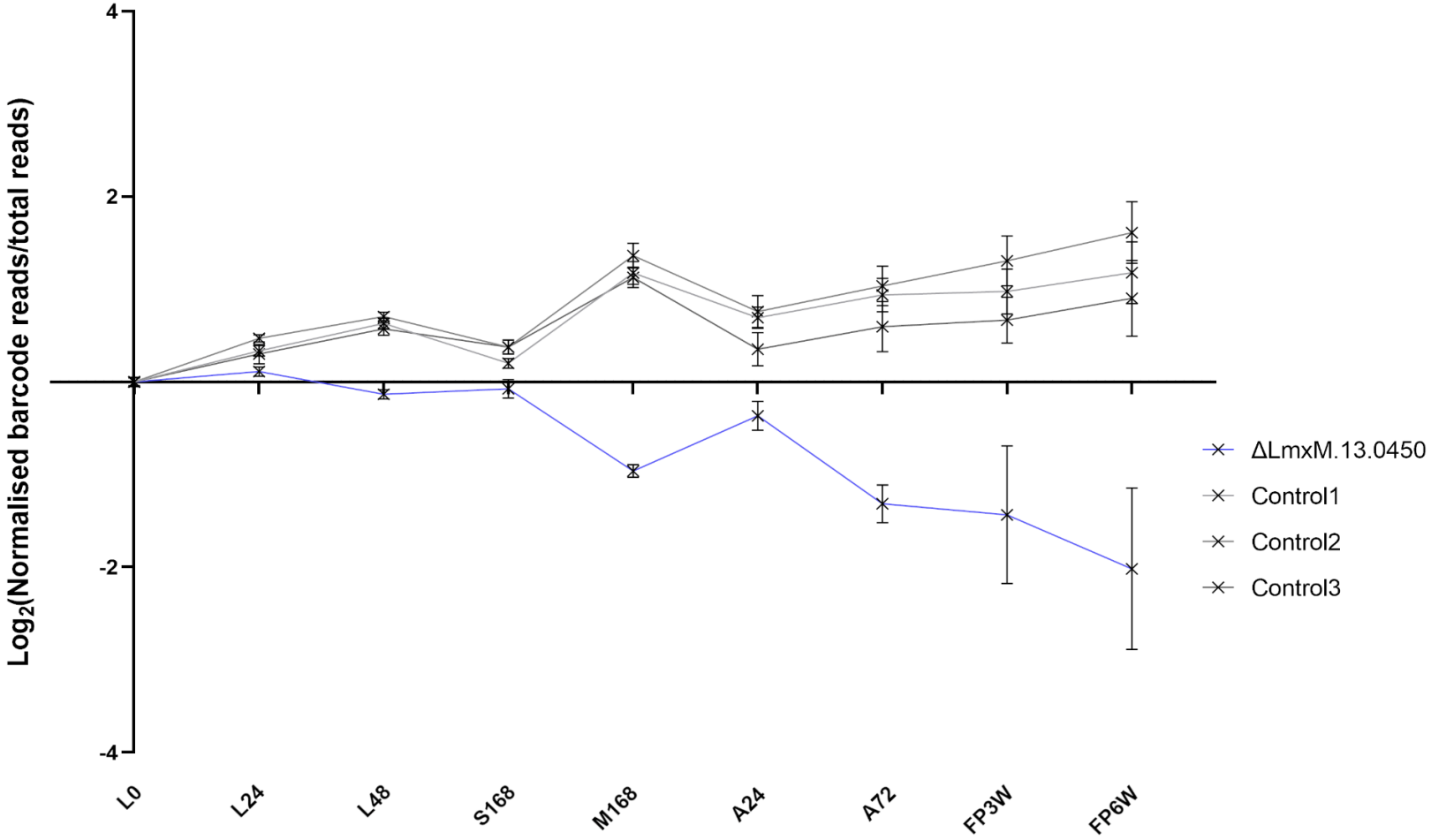
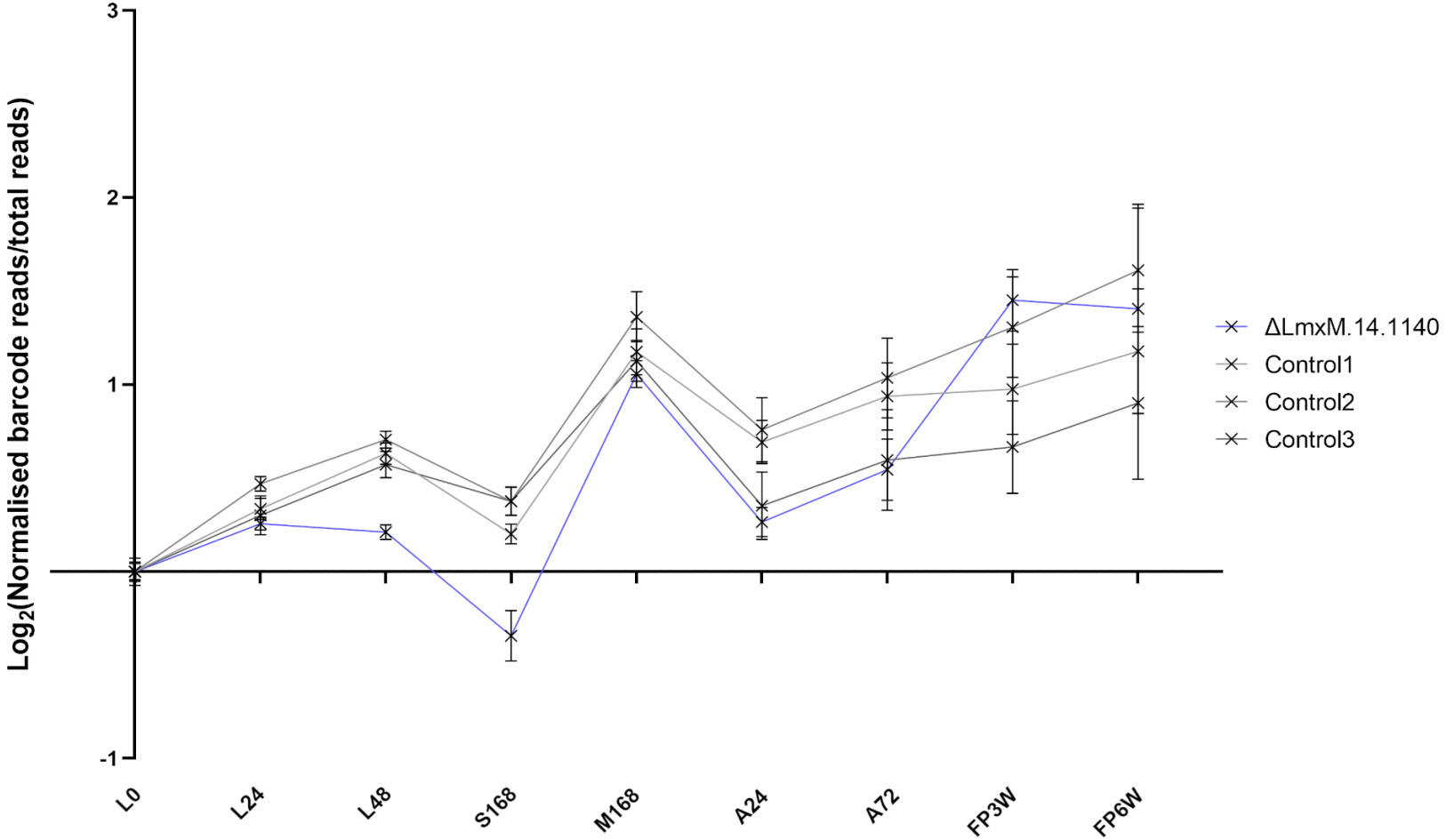
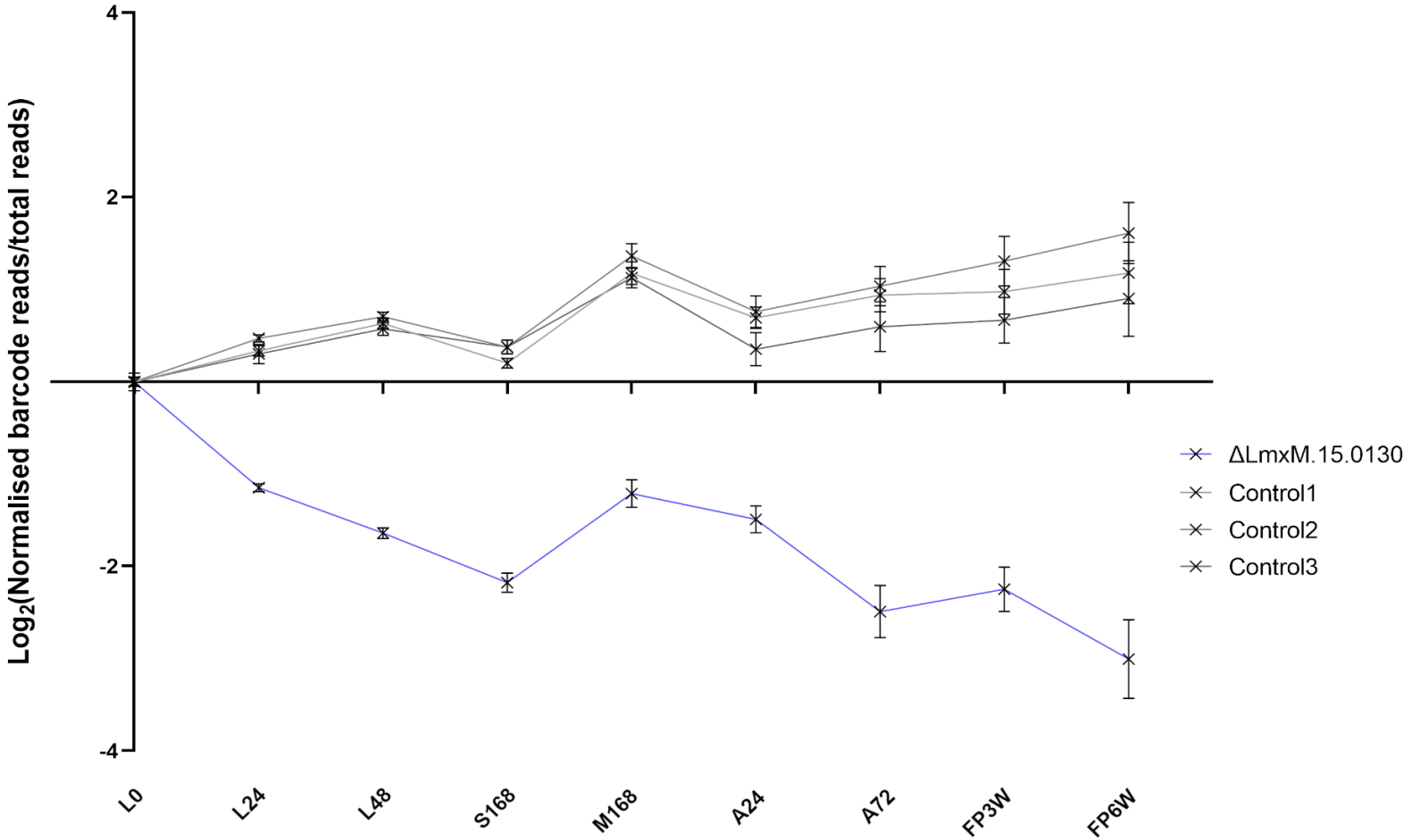


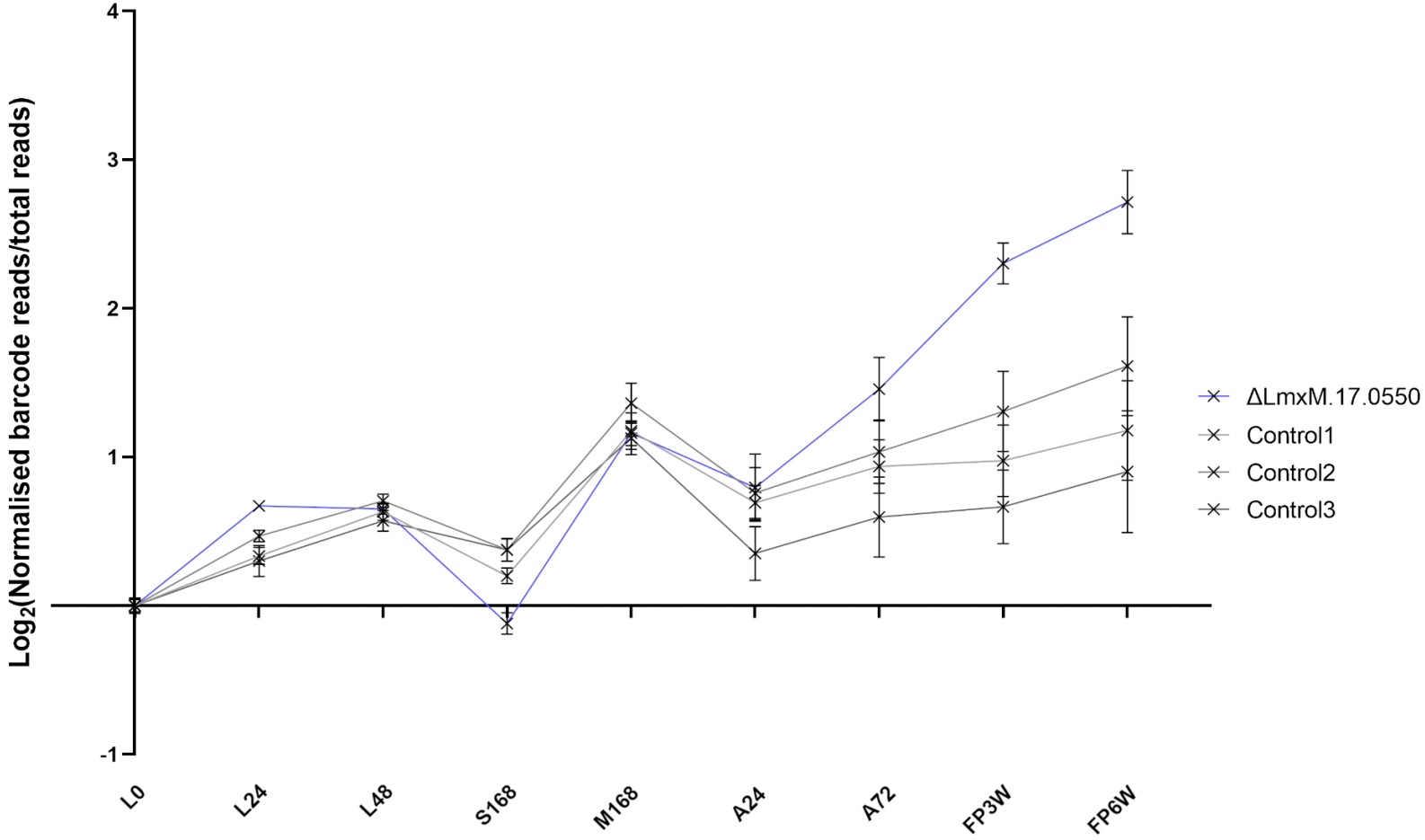
Figure 7.1. (Supplementary). Comparison of LmxM.15.0130 to other DEAD-box helicases. Full predicted amino acid sequences from the *L.mexicana* reference genome (MHOM/GT/2001/U1103) were downloaded from TriTrypDB and aligned in MEGA7 using the MUSCLE alignment tool with default parameters (see methods 2.1.2). A) The region showing the conserved 'DEAD' motif is shown below the figure, with amino acid position on the top x-axis and conservation and quality scores given on the lower x-axis. The motif is shown to be semi-conserved in LmxM.15.0130 with the sequence 'DECD'. Interestingly, the closest homolog to human DDX6 in *Leishmania*, LmxM.34.0370, shows an intact 'DEAD' motif. B) A phylogenetic tree produced from the same alignment in MEGA7 is shown with LmxM.15.0130 being most closely linked to human DDX27. The tree was constructed using 'maximum likelihood' as described in methods 2.1.2 and a scale of 0.2 amino acid substitutions/unit of branch distance.

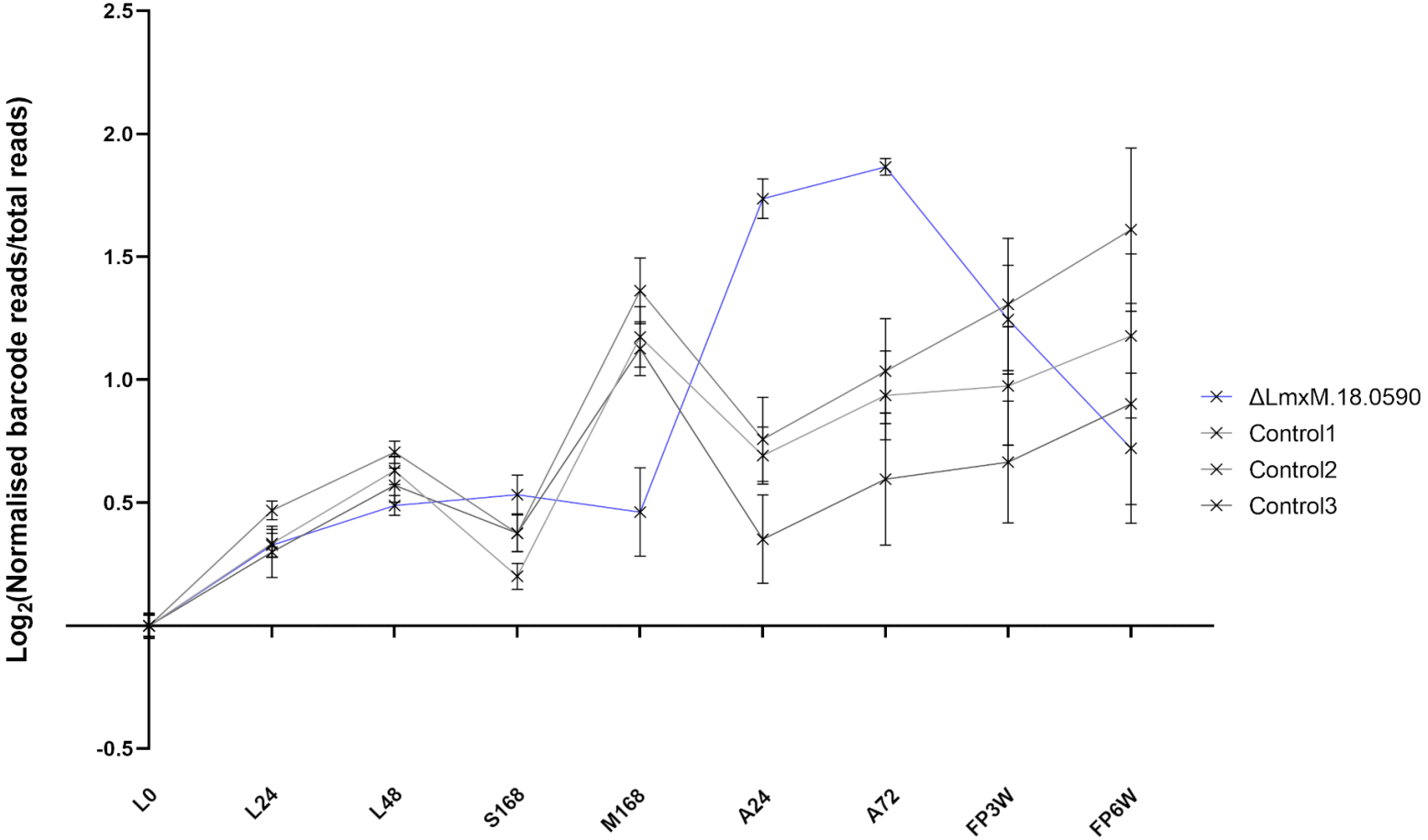

Log2(Normalised barcode reads/total reads)
5
-5
-10
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
Control1
Control2
Control3
ΔLmxM.05.0850

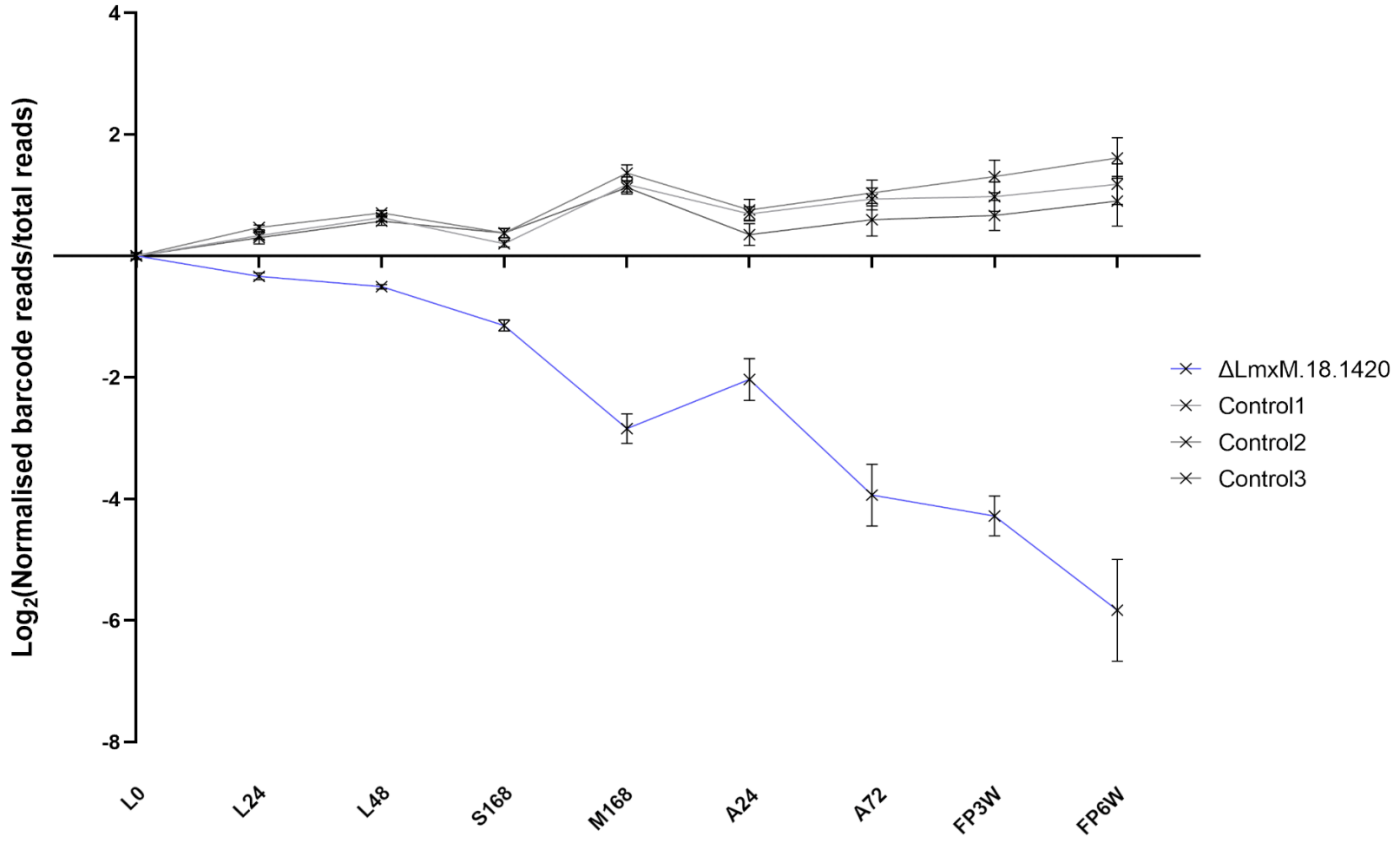

4
2
-2
-4
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.13.0450
Control1
Control2
Control3

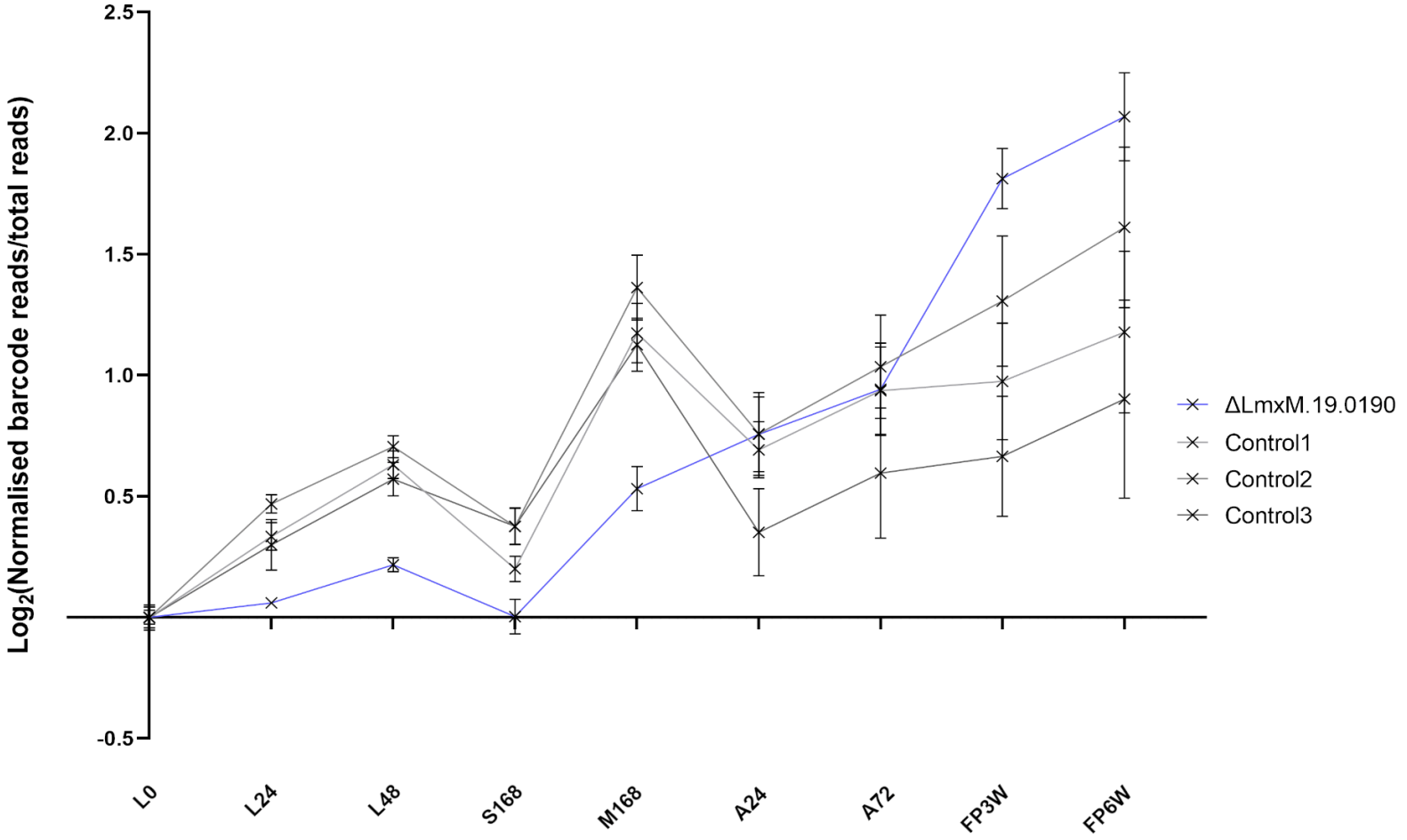

Log2(Normalised barcode reads/total reads)
3
2
1
-1
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.14.1140
Control1
Control2
Control3

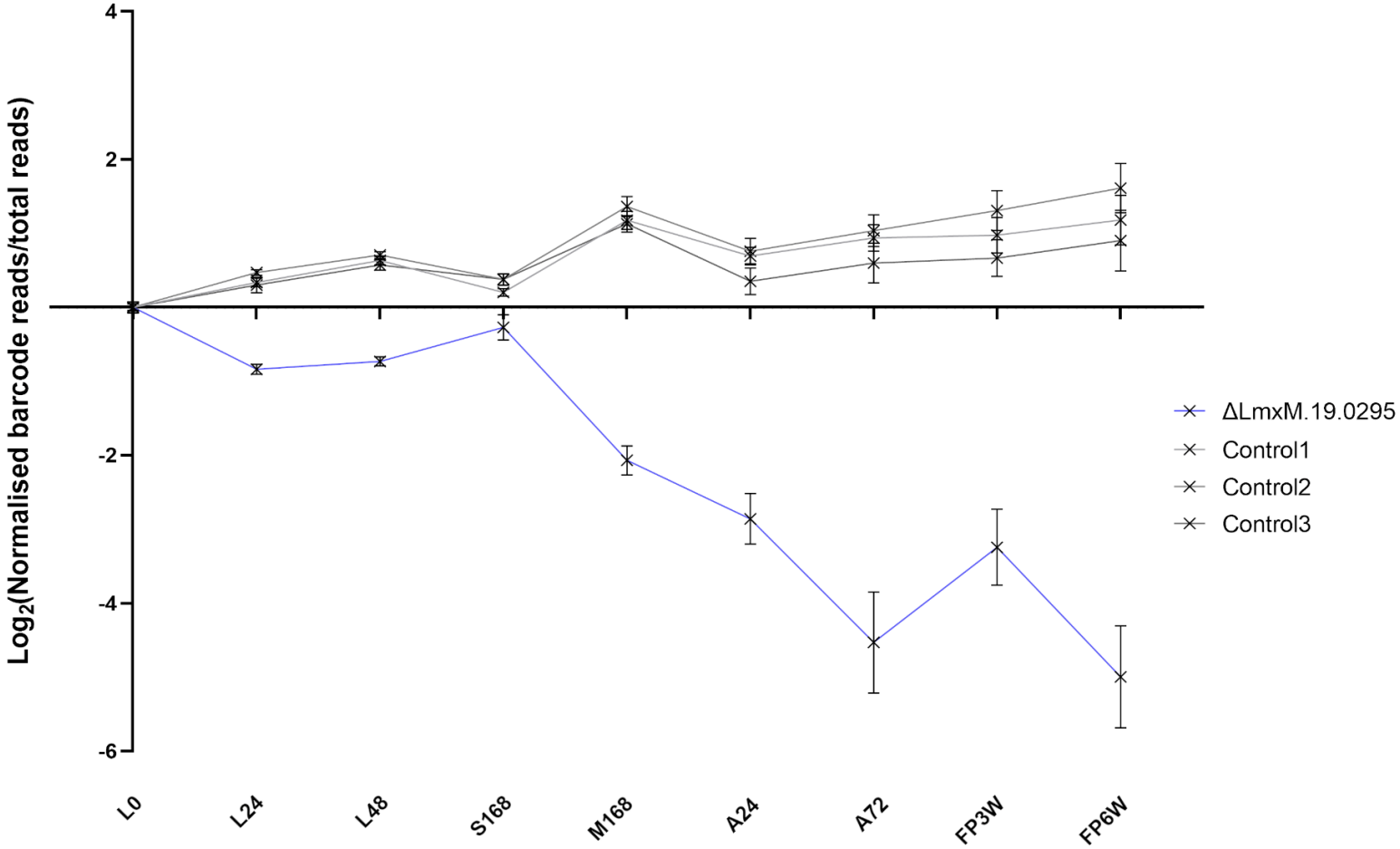

Log2(Normalised barcode reads/total reads)
4
2
-2
-4
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.15.0130
Control1
Control2
Control3

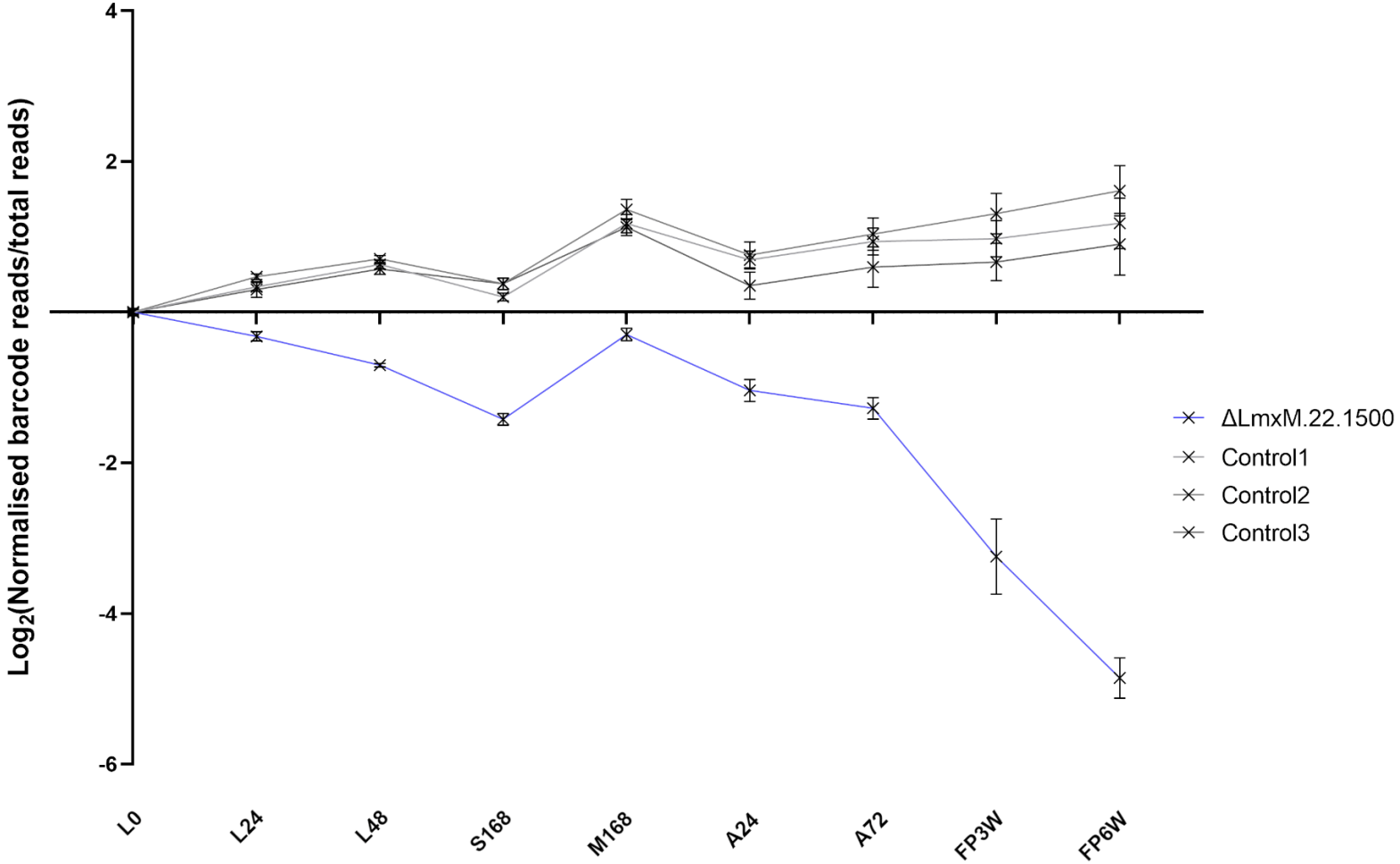

Log2(Normalised barcode reads/total reads)
4
3
2
1
-1
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.17.0550
Control1
Control2
Control3

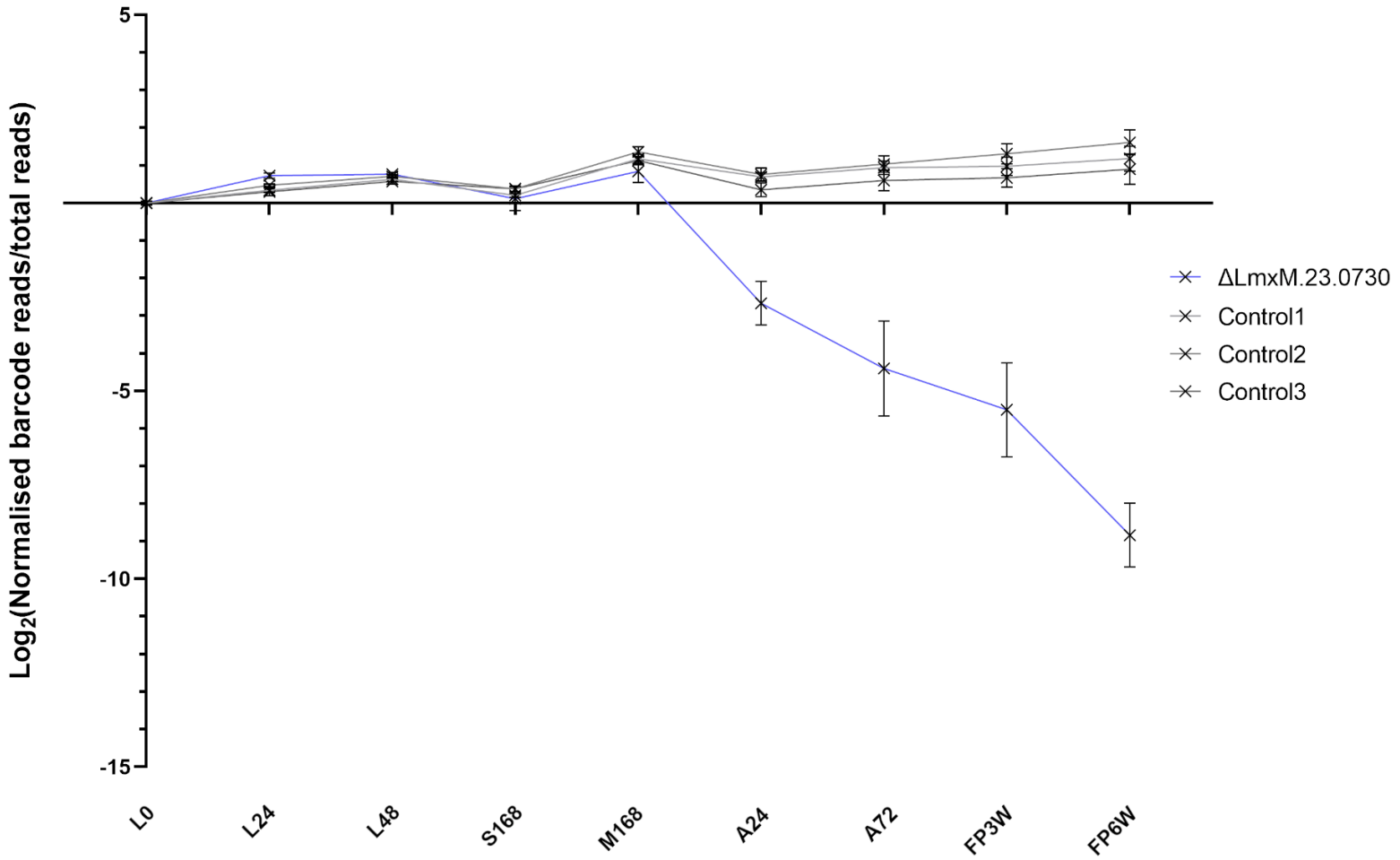

Log2(Normalised barcode reads/total reads)
2.5
2.0
1.5
1.0
0.5
-0.5
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.18.0590
Control1
Control2
Control3

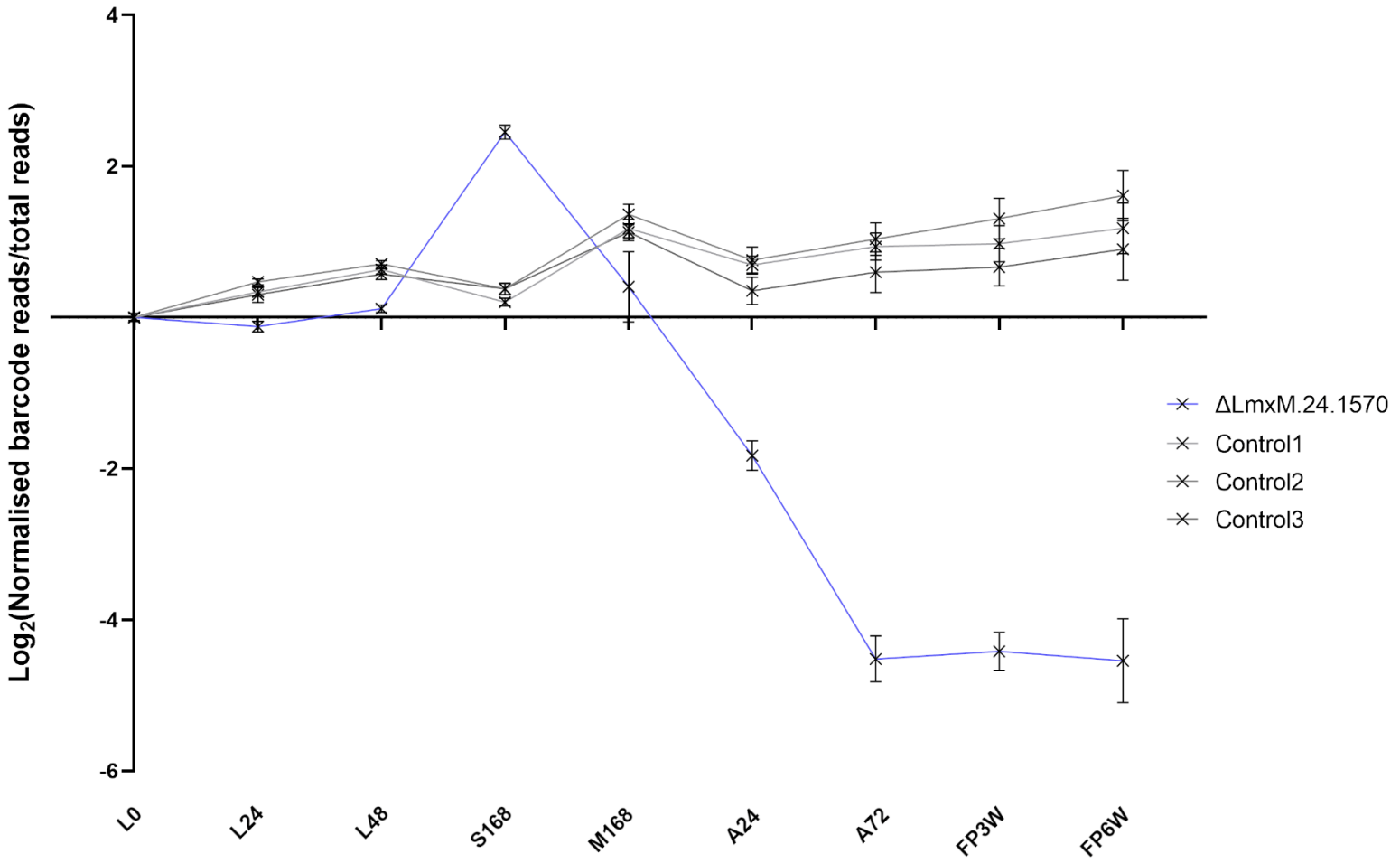

4
2
-2
-4
-6
-8
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.18.1420
Control1
Control2
Control3

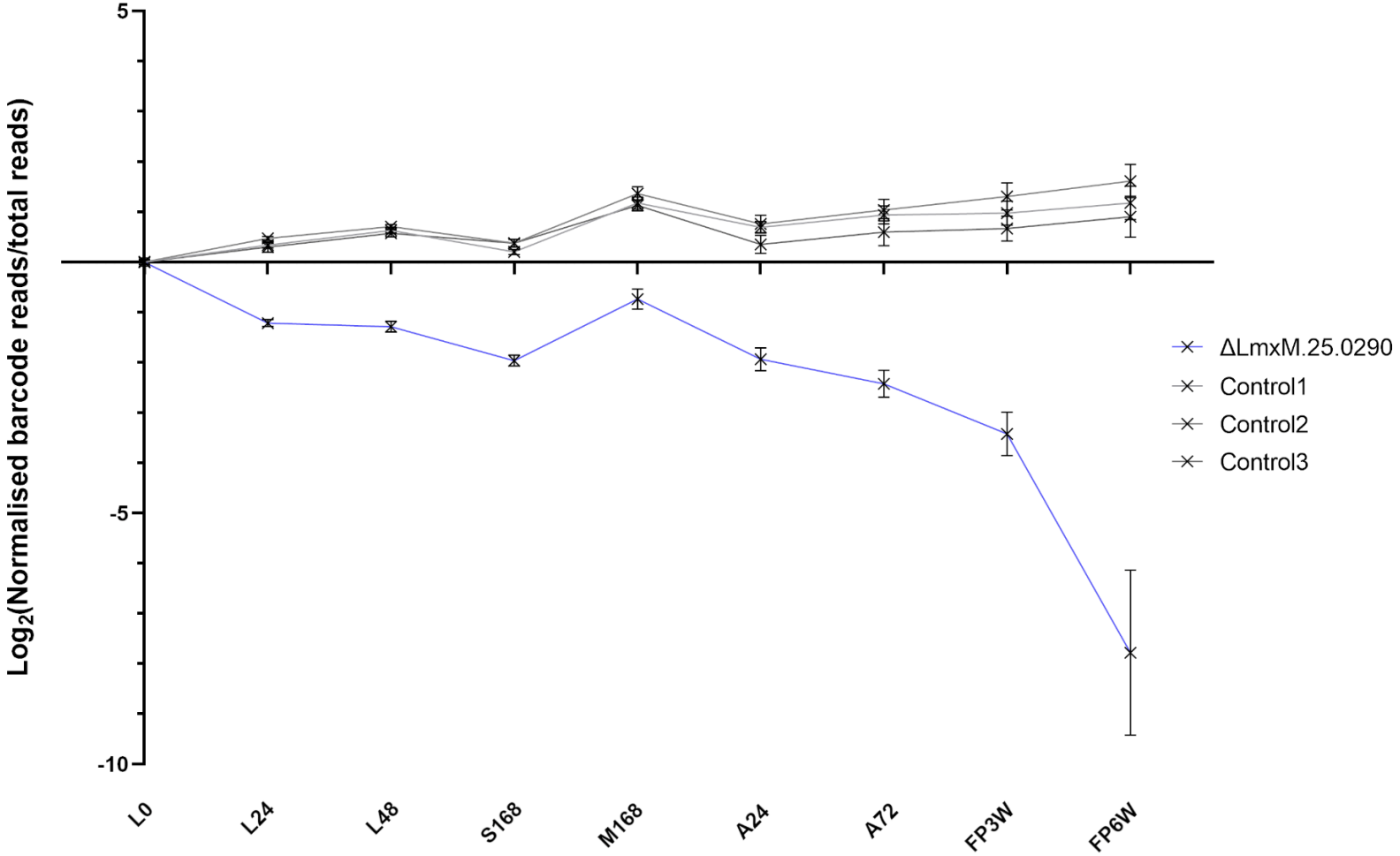

Log2(Normalised barcode reads/total reads)
2.5
2.0
1.5
1.0
0.5
-0.5
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.19.0190
Control1
Control2
Control3

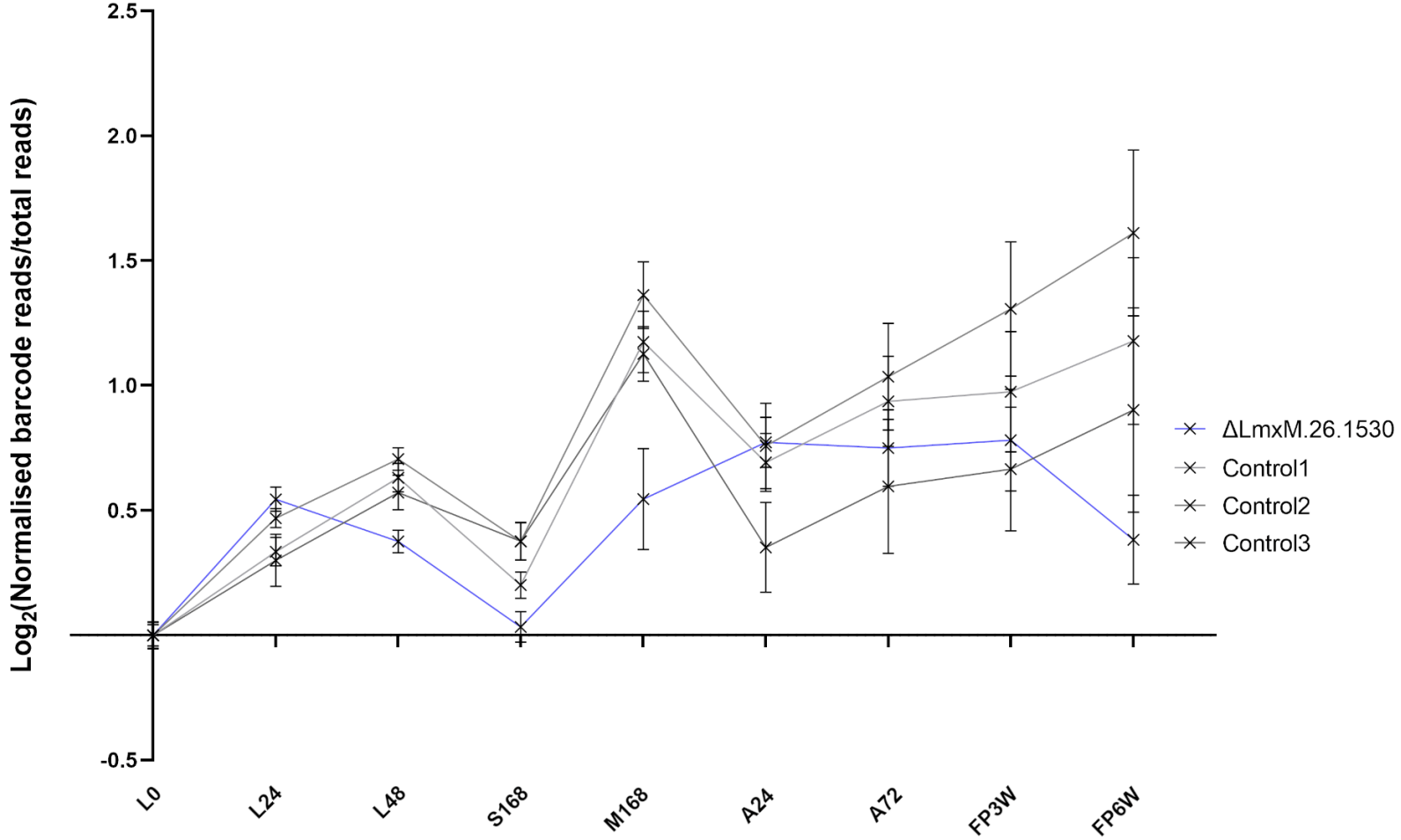

4
2
-2
-4
-6
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.19.0295
Control1
Control2
Control3

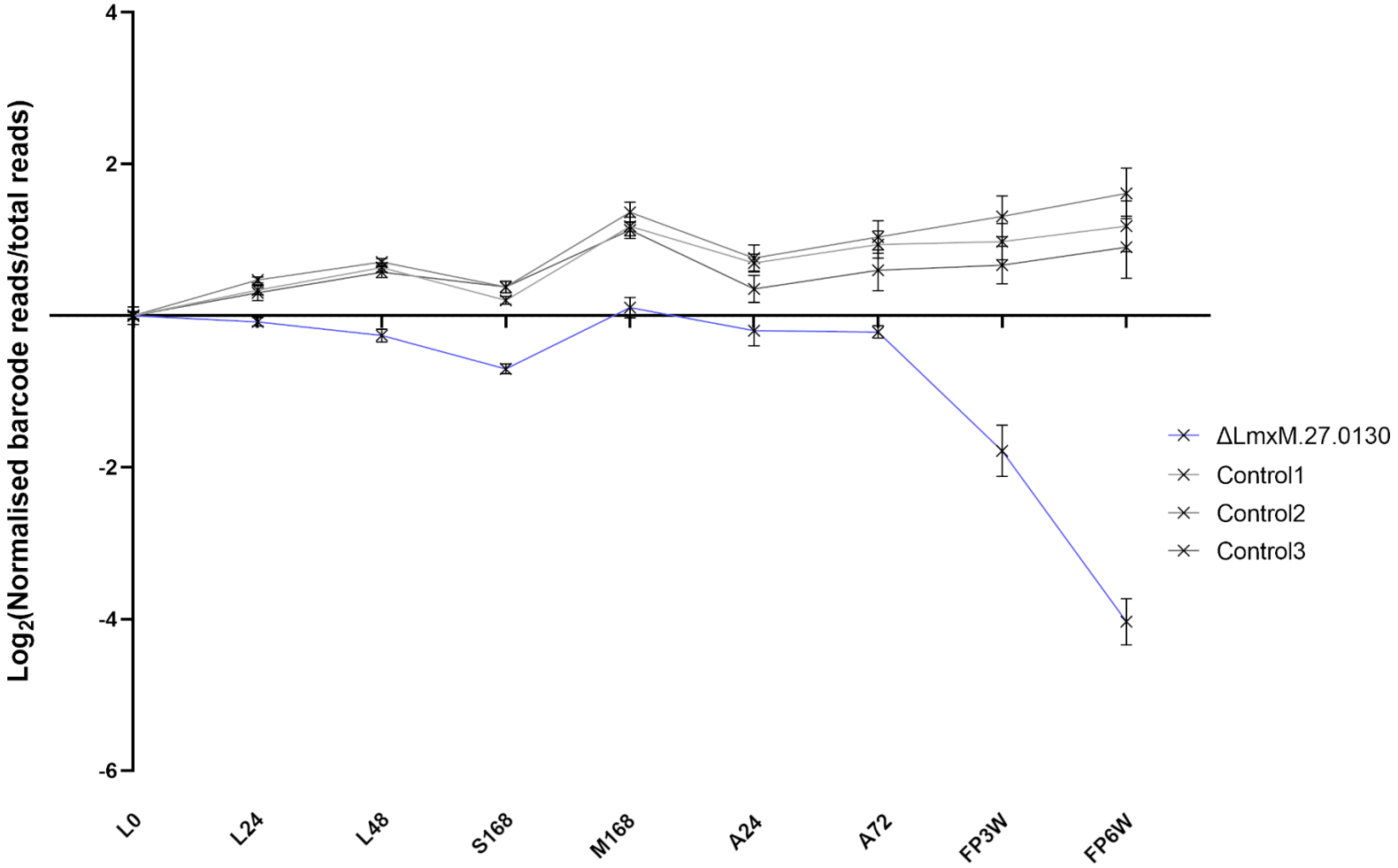

Log2(Normalised barcode reads/total reads)
4
2
-2
-4
-6
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.22.1500
Control1
Control2
Control3

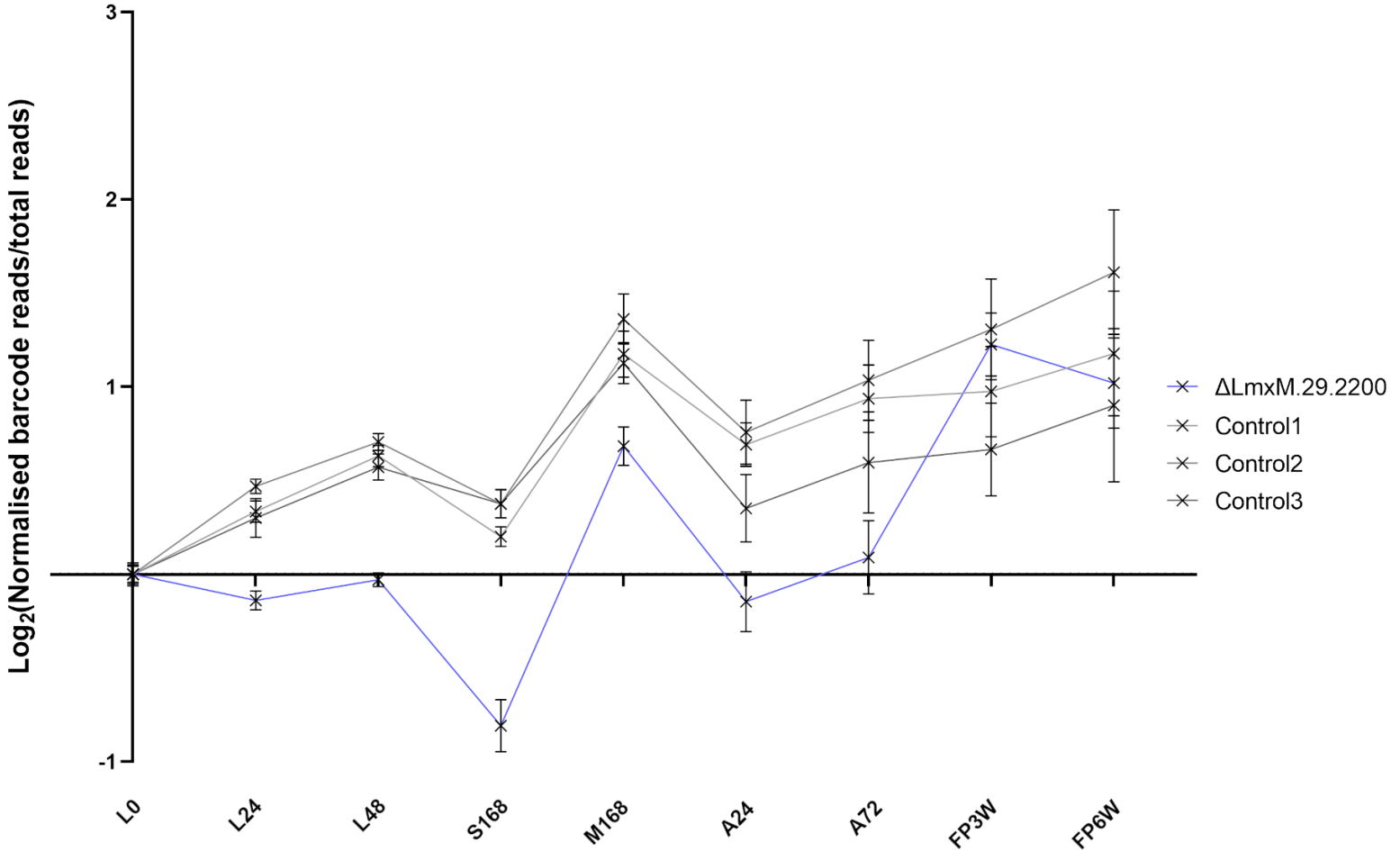

5
-5
-10
-15
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.23.0730
Control1
Control2
Control3

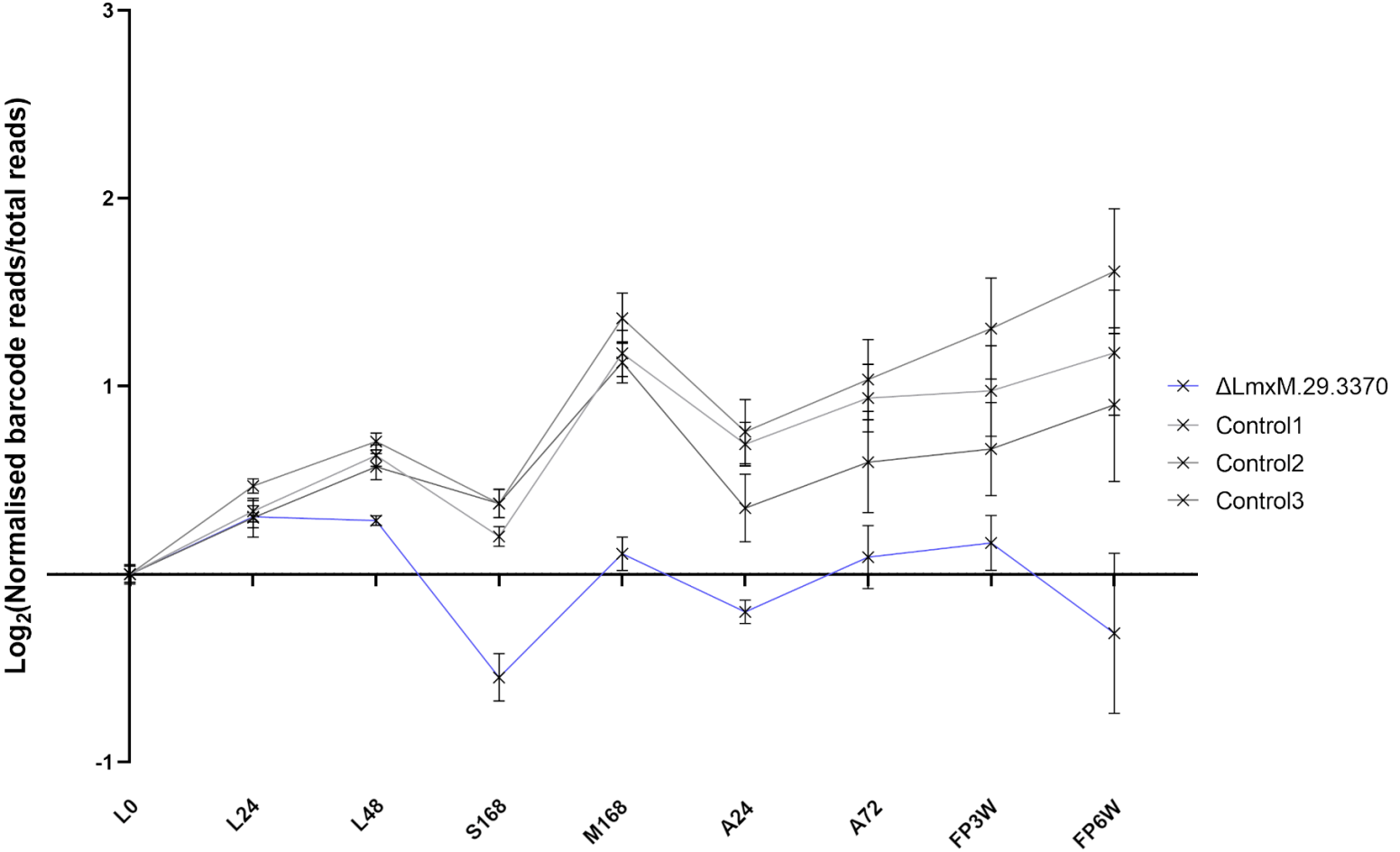

4
2
-2
-4
-6
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.24.1570
Control1
Control2
Control3

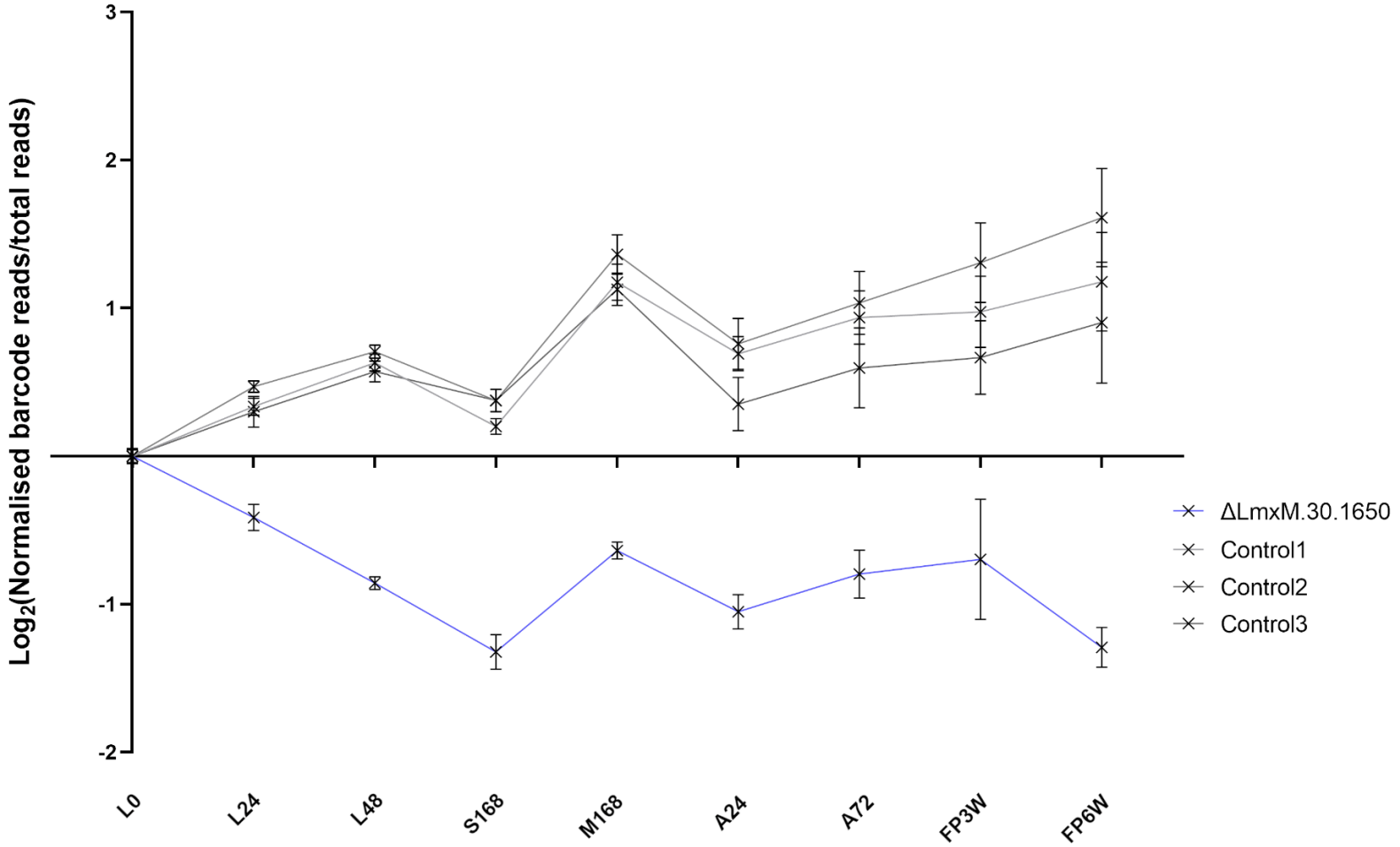

5
-5
-10
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.25.0290
Control1
Control2
Control3

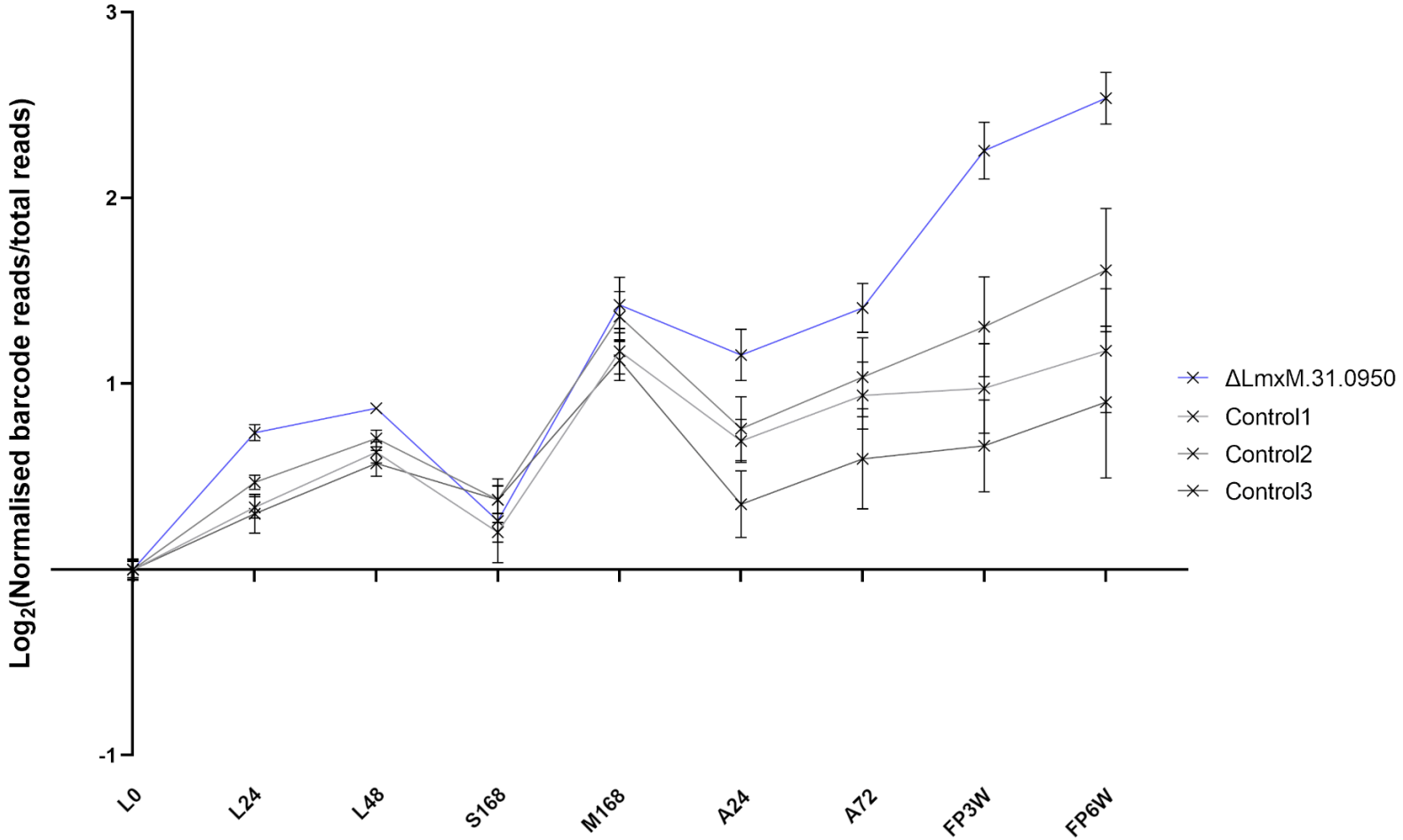

Log2(Normalised barcode reads/total reads)
2.5
2.0
1.5
1.0
0.5
-0.5
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.26.1530
Control1
Control2
Control3

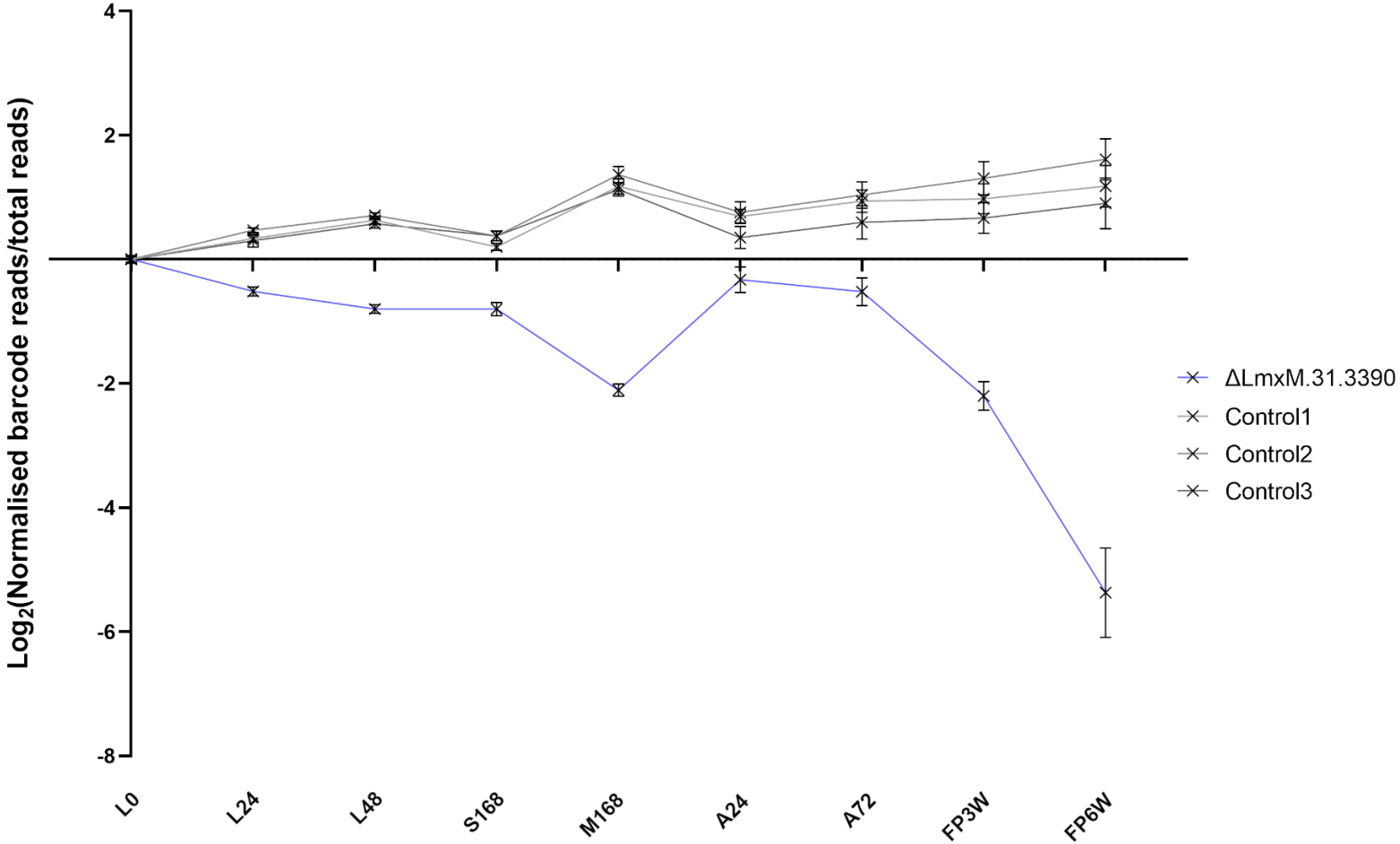

4
2
-2
-4
-6
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.27.0130
Control1
Control2
Control3

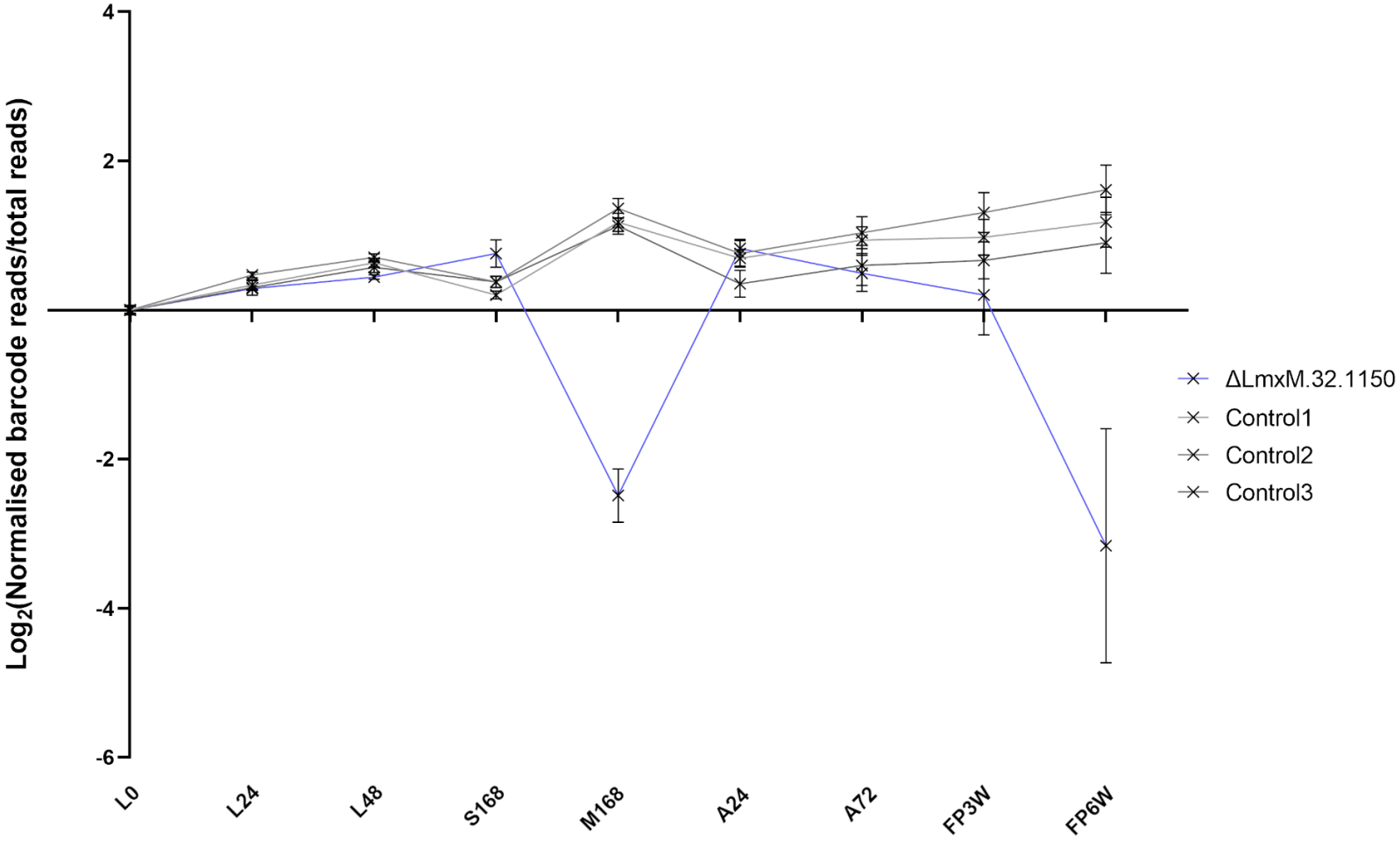

Log2(Normalised barcode reads/total reads)
3
2
1
-1
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.29.2200
Control1
Control2
Control3

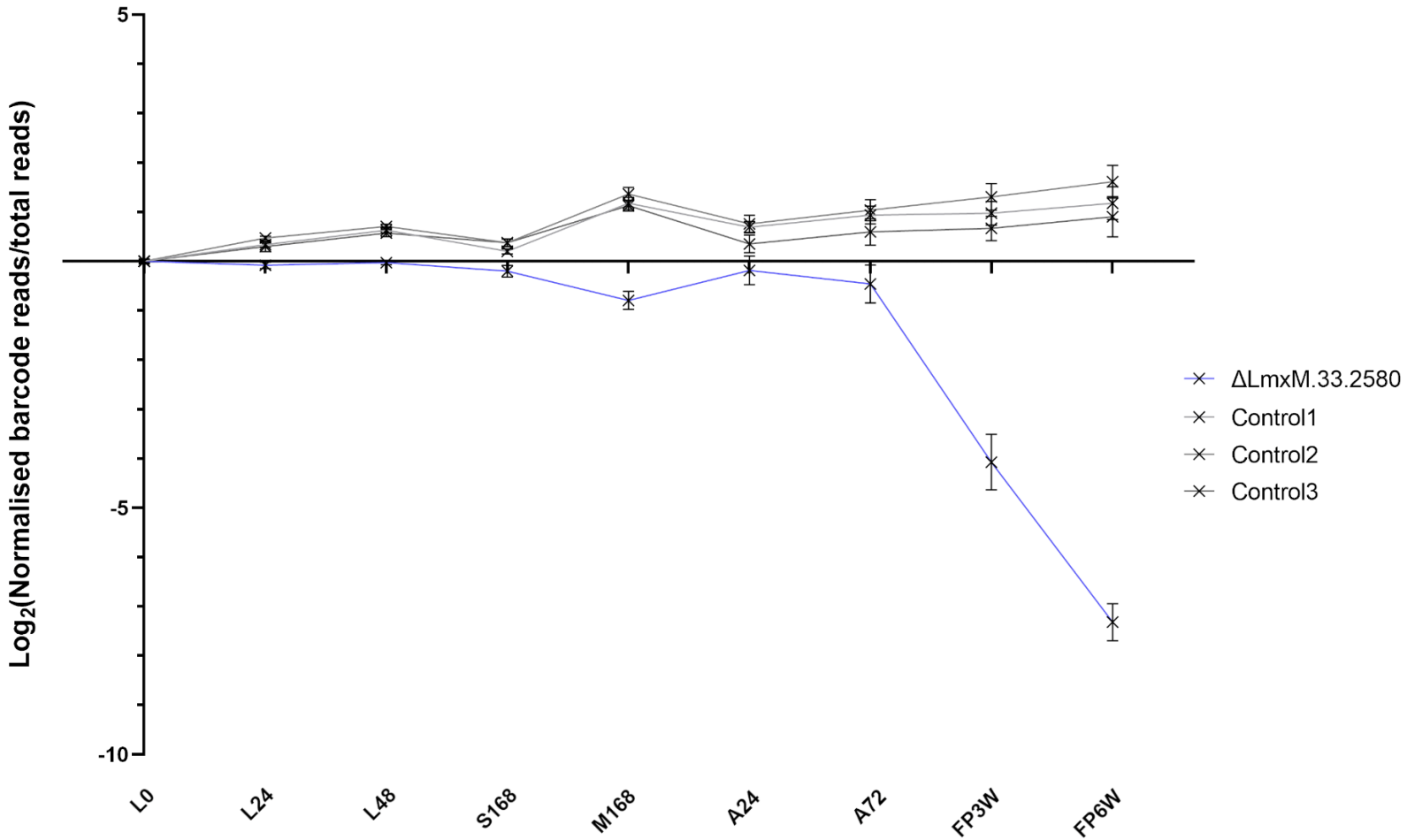

Log2(Normalised barcode reads/total reads)
3
2
1
-1
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.29.3370
Control1
Control2
Control3

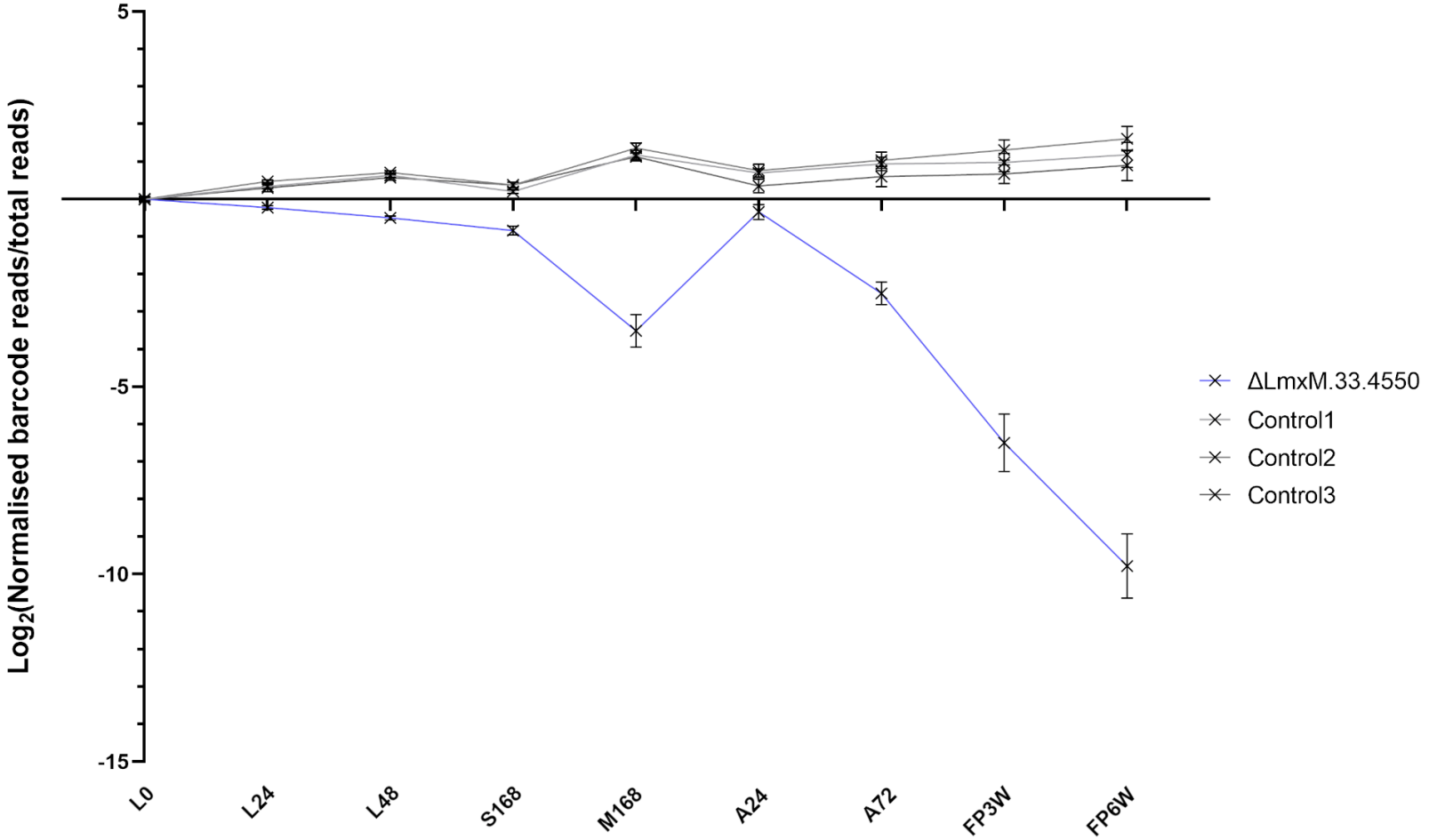

Log2(Normalised barcode reads/total reads)
3
2
1
-1
-2
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.30.1650
Control1
Control2
Control3

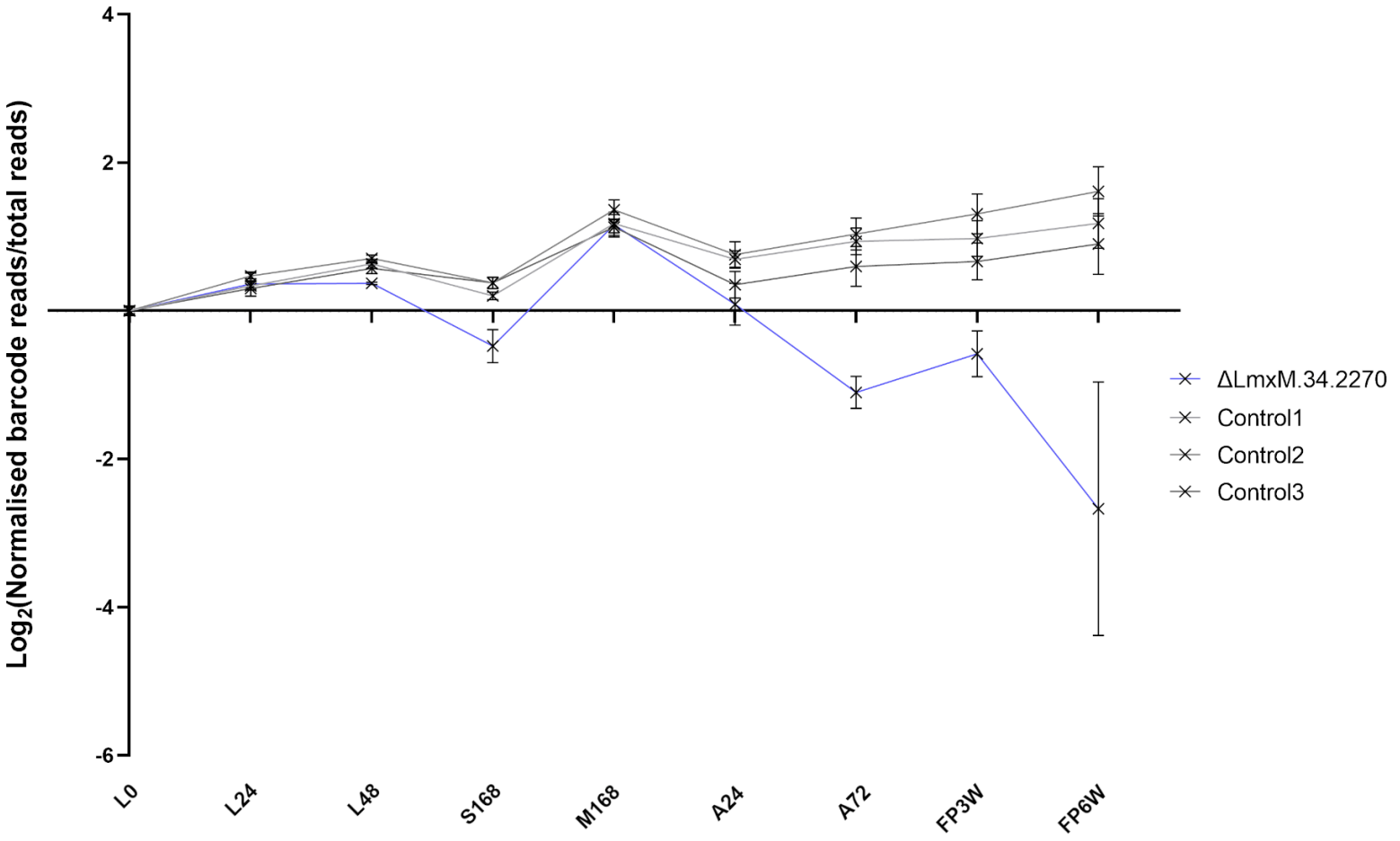

Log2(Normalised barcode reads/total reads)
3
2
1
-1
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.31.0950
Control1
Control2
Control3

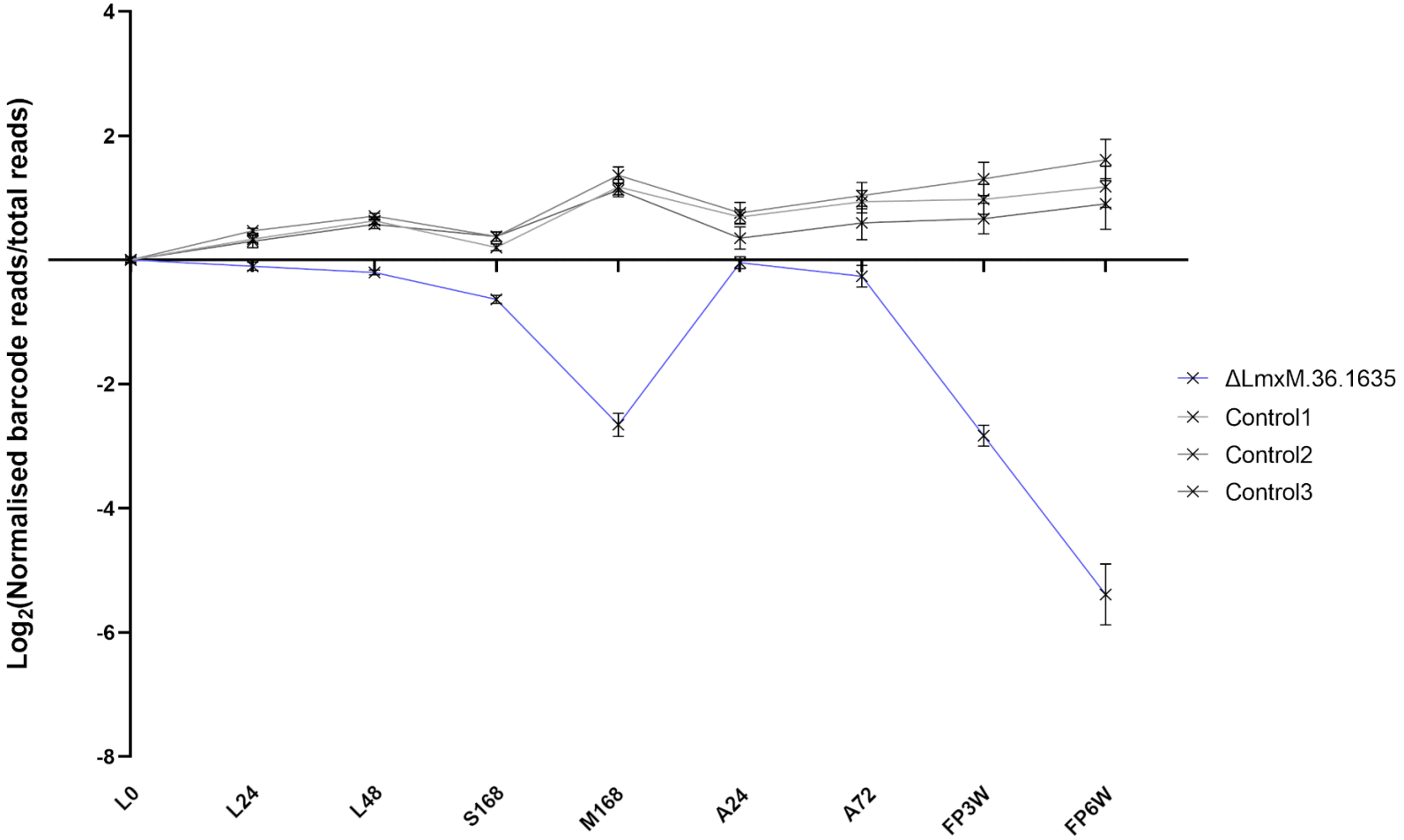

Log2(Normalised barcode reads/total reads)
4
2
-2
-4
-6
-8
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.31.3390
Control1
Control2
Control3

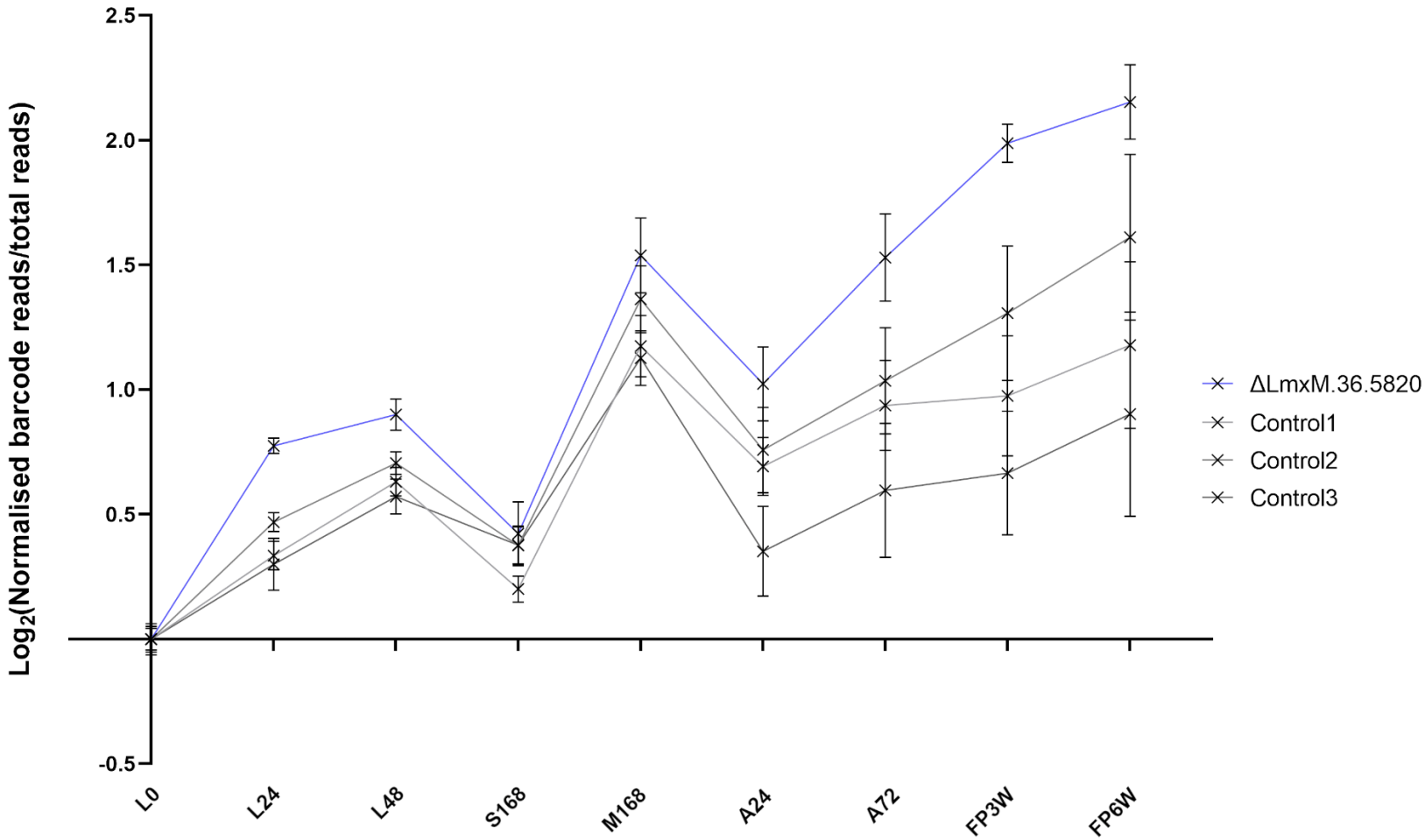

4
2
-2
-4
-6
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.32.1150
Control1
Control2
Control3


5
-5
-10
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.33.2580
Control1
Control2
Control3


Log2(Normalised barcode reads/total reads)
5
-5
-10
-15
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.33.4550
Control1
Control2
Control3


Log2(Normalised barcode reads/total reads)
4
2
-2
-4
-6
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.34.2270
Control1
Control2
Control3


Log2(Normalised barcode reads/total reads)
4
2
-2
-4
-6
-8
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.36.1635
Control1
Control2
Control3


Log2(Normalised barcode reads/total reads)
2.5
2.0
1.5
1.0
0.5
-0.5
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
ΔLmxM.36.5820
Control1
Control2
Control3

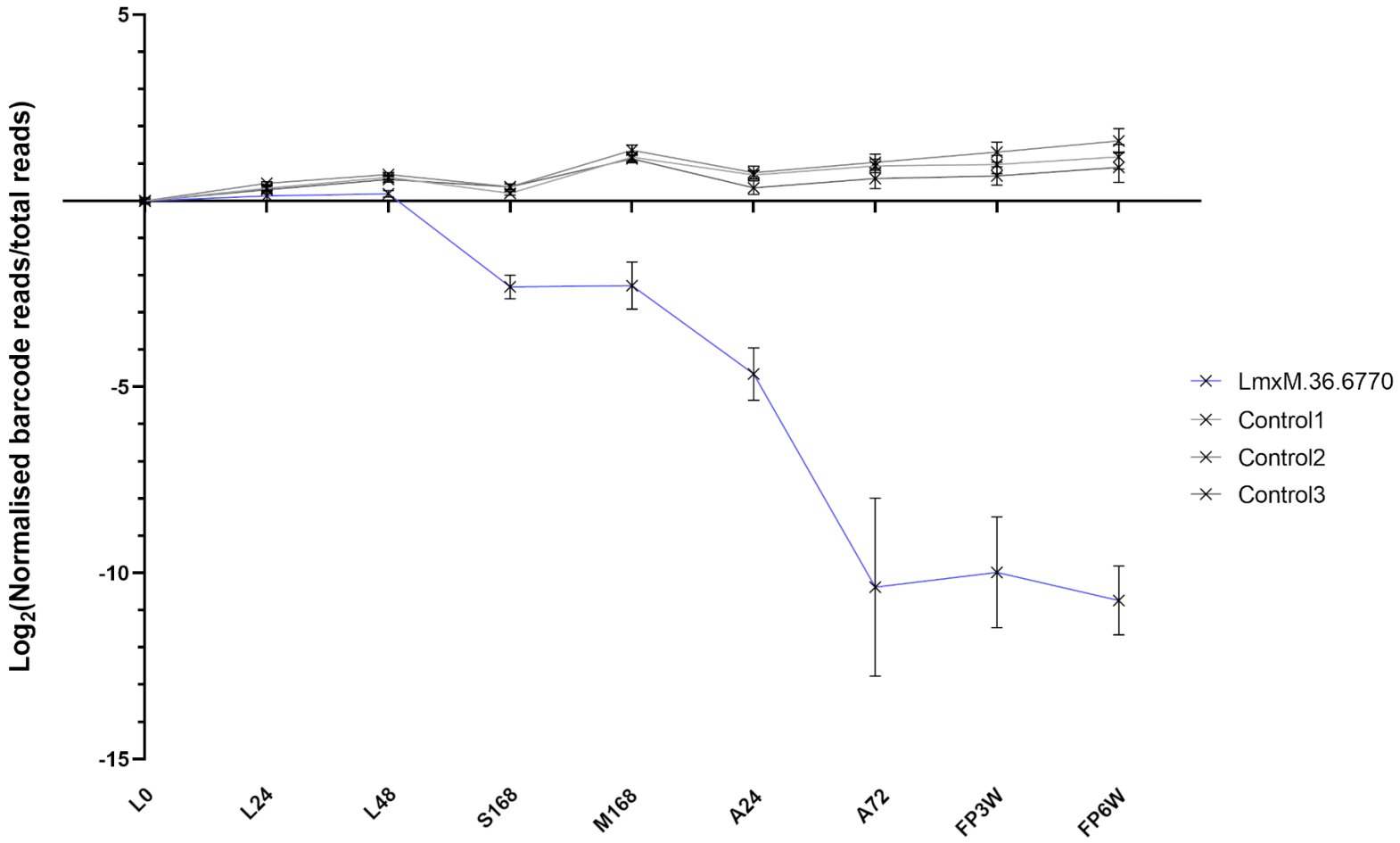

5
-5
-10
-15
Log2(Normalised barcode reads/total reads)
L0
L24
L48
S168
M168
A24
A72
FP3W
FP6W
LmxM.36.6770
Control1
Control2
Control3

Figure 7.2. (Supplementary). Trajectories of all RBP null mutant lines in the bar-seq screen. The individual trajectories of each RBP null mutant line during the bar-seq screen are plotted as the mean of the six replicates of the Log2 (normalised barcode reads/total reads) for each stage of the experiment (as in 4.8). Error bars represent the standard error around this mean. All trajectories are plotted showing each individual RBP null mutant line (blue) and the three control cell lines (grey, each containing a unique barcode but no fitness defect) for comparison. L0 = log phase 0h post-pooling, L24 = log phase 24h post-pooling, L48 = log phase 48h post-pooling, S168 = stationary phase 168h (7days) post-pooling, M168 = metacyclic purification 168h (7days) post pooling, A24 = amastigotes from bone marrow derived macrophages 24h post-infection, A72 = amastigotes from bone marrow derived macrophages 72h post-infection, FP3W = amastigotes from mouse footpads three weeks post-infection, FP6W = amastigotes from mouse footpads six weeks post-infection.

7.2 Supplementary tables

Table 7.1. (Supplementary). Antibodies. A table listing all antibodies used in this thesis with the animals they were raised in and the antigens they were raised against. Recommended dilutions based off the work in this thesis are shown for both western blot and immunofluorescence where tested.

ID/name	Type	Raised in	Raised against	Dilution (Western)	Dilution (imaging)
NIF825	secondary	goat	mouse	1:10,000	not tested
NIF824	secondary	goat	rabbit	1:10,000	not tested
05-235	Primary	mouse	EIF1α	1:10,000	not tested
abSK805	primary	rabbit	NMT	1:5000-1:10,000	not tested
A-11005	secondary	goat	mouse	not tested	1:1000
A-11012	secondary	goat	rabbit	not tested	1:1000
26183	primary	mouse	HA	1:10,000	1:500
A190-108A	primary	rabbit	HA	1:5000	1:250

Table 7.2. (Supplementary). Primers. A table showing all primers used in this thesis as well as others relevant to the project and its continuation. All primers were ordered through IDT and were stored at -20°C at 100µM. 10µM dilutions were used as stock for PCR reactions throughout, except for sgRNA cassette primers which were used at 100µM. Tagging primers were designed in the same way for all RBPs in the bar-seq screen despite only being used to tag four candidate RBPs in this study.

Oligo name	Oligo sequence (5' to 3')
931.DDX6_Up_F	TCTTTCCAGTTTCTCTTTTCCCGCTCTCCTGTAT AATGCAGACCTGCTGC
932.DDX6_Up_R_1xMyc+1xHA	AAGAGCCTTGCACAGCCCCAGCTCCAGCCAAG CGTAATCTGGAACATCGTATGGGTACAGGTCTT CTTCGCTGATCAGCTTCTGTTCCAT
933.DDX6_Up_R_3xMyc+3xHA	AAGAGCCTTGCACAGCCCCAGCTCCAGCCAAG CGTAATCTGGAACATCGTATGGGTAAGCGTAAT CTGGAACATCGTATGGGTAAGCGTAATCTGGAA CATCGTATGGGTAAAGCTTAGAACCGGAACCG
934.DDX6_5'sg	GAAATTAATACGACTCACTATAGGCGTCTATTT TTCAGCGCTCAGTTTTAGAGCTAGAAATAGC
935.DDX6_Down_R	CAGCCTCCTCTGCCCCTTCCCTGCAGCCCACCA ATTTGAGAGACCTGTGC
936.DDX6_3'sg	GAAATTAATACGACTCACTATAGGGACATCGGC AGGTGTAGACGGTTTTAGAGCTAGAAATAGC
942.DDX6Check_Up_F	CGCTACACCCATTGCGCAGGTCG
943.DDX6Check_Down_R	GCAAGTCCTGAAACGGGGCC
973.DDX6Check_In_F	CGCTCCACCATGCCTTTGTC
974.DDX6Check_In_R	TCGGCTGTCCATCACTGTCC
975.DDX6Check_Down_R	GGCCCCGTTTCAGGACTTGC
978.DDX27Check_Gen_F	TTTCTCCCACACCAGTGCAG
979.DDX27Check_Gen_R	CCATCTACGTTGTCGGCAGT
1017.PGL2666edit_F	GCTAACCATGGTGCC
1018.PGL2666edit_R	CCAGAACCCAGGTCC
1019.PGL2666_HA_gsg_Hindiii_R	TGATGAAAGCTTACCAGAACCAGCGTAATCTGG AACATCGTATGGGTAAGCGTAATCTGGAACATC GTATGGGTAAGCGTAATCTGGAACATCGTATGG GTACATGCTTGACAAGTG
1029.LmxM.01.0800_Up_F	TGCTCCACACACTCTGCAAGCCATCCACCTTAA TACGACTCACTATAAAACTGGAAGGACGTCATG TTTGTATAATGCAGACCTGCTGC
1030.LmxM.01.0800_Dn_R	CTTGTGTACAGGTGCGTTACTCATGGTGCCCCA ATTTGAGAGACCTGTGC
1031.LmxM.01.0800_5'sgRNA	GAAATTAATACGACTCACTATAGGACGGGTCG GGGTGGCGGAAGGTTTTAGAGCTAGAAATAGC
1032.LmxM.01.0800_3'sgRNA	GAAATTAATACGACTCACTATAGGATCCAGCAA GGAGTAGGGCAGTTTTAGAGCTAGAAATAGC
1033.LmxM.11.0470_Up_F	ATACACAACACTCCCTTCTCGTTTTTTCCCTAAT ACGACTCACTATAAAACTGGAAGACACACATG TCCGTATAATGCAGACCTGCTGC
1034.LmxM.11.0470_Dn_R	CGCTTACACCACACACAACACACATAAAGGCC AATTTGAGAGACCTGTGC
1035.LmxM.11.0470_5'sgRNA	GAAATTAATACGACTCACTATAGGAATGACTTG GCGGGCGTAAAGTTTTAGAGCTAGAAATAGC
1036.LmxM.11.0470_3'sgRNA	GAAATTAATACGACTCACTATAGGCACAAGAA CGCTCTCCATCAGTTTTAGAGCTAGAAATAGC
1037.LmxM.15.0130_Up_F	TCTTTCCAGTTTCTCTTTTCCCGCTCTCCTTAAT ACGACTCACTATAAAACTGGAAGTGGTTGGCCC TTGTATAATGCAGACCTGCTGC
1038.LmxM.15.0130_Dn_R	CAGCCTCCTCTGCCCCTTCCCTGCAGCCCACCA ATTTGAGAGACCTGTGC
1039.LmxM.15.0130_5'sgRNA	GAAATTAATACGACTCACTATAGGCGTCTATTT TTCAGCGCTCAGTTTTAGAGCTAGAAATAGC
1040.LmxM.15.0130_3'sgRNA	GAAATTAATACGACTCACTATAGGGACATCGGC AGGTGTAGACGGTTTTAGAGCTAGAAATAGC
1041.LmxM.18.1420_Up_F	CCGCGCACCCCACCGCGAGACGATCAACCGTA ATACGACTCACTATAAAACTGGAAGGCTAATCT GGAAGTATAATGCAGACCTGCTGC
1042.LmxM.18.1420_Dn_R	CAGAGAACACAAGAATGCACAACAAGACCACC AATTTGAGAGACCTGTGC
1043.LmxM.18.1420_5'sgRNA	GAAATTAATACGACTCACTATAGGTTTTCTTCTT GGTTTTCTACGTTTTAGAGCTAGAAATAGC

1044.LmxM.18.1420_3'sgRNA	GAAATTAATACGACTCACTATAGGGGCTGCTGG CGGTCTCTCTCGTTTTAGAGCTAGAAATAGC
1045.LmxM.36.1635_Up_F	TCTCTCTGTAATCACTCCCGCGTTTCGCCTTAAT ACGACTCACTATAAAACTGGAAGCTATGCGCCA TTGTATAATGCAGACCTGCTGC
1046.LmxM.36.1635_Dn_R	CGCACTTCTACCGTTTCCTGCTGTGTCCCCCCAA TTTGAGAGACCTGTGC
1047.LmxM.36.1635_5'sgRNA	GAAATTAATACGACTCACTATAGGCGTGAGAA AAAGAGAGATTCGTTTTAGAGCTAGAAATAGC
1048.LmxM.36.1635_3'sgRNA	GAAATTAATACGACTCACTATAGGGTATCTGTA CTGAATGGGCGGTTTTAGAGCTAGAAATAGC
1049.LmxM.21.1552_Up_F	CTCACACGTTACGCCGTCCTTTTTTCTCCTTAAT ACGACTCACTATAAAACTGGAAGCGATTTCTGA AAGTATAATGCAGACCTGCTGC
1050.LmxM.21.1552_Dn_R	CACATGCGTTATCGGGCACCACCACATCCACCA ATTTGAGAGACCTGTGC
1051.LmxM.21.1552_5'sgRNA	GAAATTAATACGACTCACTATAGGGATGGCAC AAGTGAGAGGGAGTTTTAGAGCTAGAAATAGC
1052.LmxM.21.1552_3'sgRNA	GAAATTAATACGACTCACTATAGGCAAGACGA CATCTACGCCTAGTTTTAGAGCTAGAAATAGC
1053.LmxM.27.0130_Up_F	GCTTTCCCCGCAGAGTAAGATAATAATCCTTAA TACGACTCACTATAAAACTGGAAGGCGCAATCC GGGGTATAATGCAGACCTGCTGC
1054.LmxM.27.0130_Dn_R	TGAAGGAAGAAGGGCACGTACCGCTTTCCACC AATTTGAGAGACCTGTGC
1055.LmxM.27.0130_5'sgRNA	GAAATTAATACGACTCACTATAGGTAGAAGTCG AGTTAGTTAAGGTTTTAGAGCTAGAAATAGC
1056.LmxM.27.0130_3'sgRNA	GAAATTAATACGACTCACTATAGGATACGTAAG ATAAGAAGATAGTTTTAGAGCTAGAAATAGC
1057.LmxM.27.1300_Up_F	CCCTCCTCTCCCTCCCCCAACTCACCCCCGTAAT ACGACTCACTATAAAACTGGAAGCAGGGAGTG GAAGTATAATGCAGACCTGCTGC
1058.LmxM.27.1300_Dn_R	ACTTCCCTTCTCTGATCCTTTCTCCTTCCTCCAA TTTGAGAGACCTGTGC
1059.LmxM.27.1300_5'sgRNA	GAAATTAATACGACTCACTATAGGCCTTCACGG CTCTTTCAAGGGTTTTAGAGCTAGAAATAGC
1060.LmxM.27.1300_3'sgRNA	GAAATTAATACGACTCACTATAGGCGCACGCAC GAGAGAGTGAGGTTTTAGAGCTAGAAATAGC
1061.LmxM.27.1680_Up_F	CGTCCTCCTCGAAGTCCATCGGCGTGTCCGTAA TACGACTCACTATAAAACTGGAAGTGCCATACC AGGGTATAATGCAGACCTGCTGC
1062.LmxM.27.1680_Dn_R	TCTTTGCCACTTTCCCCTTCGTCGAGAGCTCCAA TTTGAGAGACCTGTGC
1063.LmxM.27.1680_5'sgRNA	GAAATTAATACGACTCACTATAGGGTGAGTTGA AGAGCAACGAAGTTTTAGAGCTAGAAATAGC
1064.LmxM.27.1680_3'sgRNA	GAAATTAATACGACTCACTATAGGCTGGGGTCA TCACACACATGGTTTTAGAGCTAGAAATAGC
1065.LmxM.28.0825_Up_F	TCACGGCCATCGCAACCCGCCTCCCCCTCCTAA TACGACTCACTATAAAACTGGAAGAAGAAGAT GGAAGTATAATGCAGACCTGCTGC
1066.LmxM.28.0825_Dn_R	CAGCTCGCGCCAAAGCCCTCTCACTCGCCCCCA ATTTGAGAGACCTGTGC
1067.LmxM.28.0825_5'sgRNA	GAAATTAATACGACTCACTATAGGCTCTCTCGC CTTTCGTTGCTGTTTTAGAGCTAGAAATAGC
1068.LmxM.28.0825_3'sgRNA	GAAATTAATACGACTCACTATAGGAGGTAAGTC AAGAAAAGAGCGTTTTAGAGCTAGAAATAGC
1069.LmxM.33.2580_Up_F	AACTGCCTTACTCTGCCCTGCATTTTTCCCTAAT ACGACTCACTATAAAACTGGAAGGTAGTATCCA TTGTATAATGCAGACCTGCTGC
1070.LmxM.33.2580_Dn_R	CACGCCAGGAAATAGCCAACCAACCCCCCTCC AATTTGAGAGACCTGTGC
1071.LmxM.33.2580_5'sgRNA	GAAATTAATACGACTCACTATAGGCCTGTTGCG AGGGTAAAAAGGTTTTAGAGCTAGAAATAGC
1072.LmxM.33.2580_3'sgRNA	GAAATTAATACGACTCACTATAGGAAAATAGC AGGAATCGGGGGGTTTTAGAGCTAGAAATAGC
1073.LmxM.34.2200_Up_F	TCTTCTCCCCGATCTTTTGTGCGTTTTCCGTAAT ACGACTCACTATAAAACTGGAAGTTCTTGATGA AAGTATAATGCAGACCTGCTGC
1074.LmxM.34.2200_Dn_R	TGTGCGTGGGTGCGTGGGTGGGTGGGTGGGCC AATTTGAGAGACCTGTGC

1075.LmxM.34.2200_5'sgRNA	GAAATTAATACGACTCACTATAGGAGGAGTTTT TTCTTTTTGTTGTTTTAGAGCTAGAAATAGC
1076.LmxM.34.2200_3'sgRNA	GAAATTAATACGACTCACTATAGGTGGGTGTGG GGGCGTGTGTGGTTTTAGAGCTAGAAATAGC
1077.LmxM.08_29.0680_Up_F	CTCTTACCCTCCCCCTTCTCTGCCCTGCCGTAAT ACGACTCACTATAAAACTGGAAGACGACATCC GTTGTATAATGCAGACCTGCTGC
1078.LmxM.08_29.0680_Dn_R	GTAGATTAGTTACCCCTATCTTTGCGGTGGCCA ATTTGAGAGACCTGTGC
1079.LmxM.08_29.0680_5'sgRNA	GAAATTAATACGACTCACTATAGGGAACTGAGT TCTTTGGCATGGTTTTAGAGCTAGAAATAGC
1080.LmxM.08_29.0680_3'sgRNA	GAAATTAATACGACTCACTATAGGTACGGAGA GGAAGGGGGCAGGTTTTAGAGCTAGAAATAGC
1143.E.P_OL6137_sgRNA_R	CCCCTCTTCTTCCCCACACG
1144.E.P_OL9369	TCGAGGAGGAGGTGAACCCA
1161.LmxM.18.1420chk_Up_F	CCTTGGGCACCACGTCAAAC
1162.LmxM.18.1420chk_In_F	CGCACGCACATACTCACACC
1163.LmxM.18.1420chk_In_R	GCGCCGATATTCTTGCCCAG
1164.LmxM.21.1552chk_Up_F	TCACGGCGCATTTTCACGTC
1165.LmxM.21.1552chk_In_F	GGTATCTCGCAGTCTGTCGCT
1166.LmxM.21.1552chk_In_R	GCAAGCGTACTTTGGCCAGC
1167.LmxM.27.0130chk_Up_F	GATTACGGAGCTCCTCGGCG
1168.LmxM.27.0130chk_In_F	CTGTTCCCCCATTCTCGCCC
1169.LmxM.27.0130chk_In_R	TACGTCATCGCCGTGCTGAA
1170.LmxM.33.2580chk_Up_F	CTGCTCCGTCACCTCGCTAT
1171.LmxM.33.2580chk_In_F	CACAGCATCGCACCCCTTTT
1172.LmxM.33.2580chk_In_R	CGTGGCTTTGGCTTCATCGAG
1173.LmxM.28.0825chk_Up_F	GAAGTTGTTGTTTTGACGGCTGC
1174.LmxM.28.0825chk_In_F	CCCAACCGACTCACACTCCC
1175.LmxM.28.0825chk_In_R	TGGACGACAACAGCATCCGT
1176.LmxM.27.1680chk_Up_F	CGAGGCCATGGACTGGAACA
1177.LmxM.27.1680chk_In_F	CCCTCTCCCCCTCCTTCACT
1178.LmxM.27.1680chk_In_R	CGGGTGCAGGTCCATCTTGT
1179.LmxM.01.0800chk_Up_F	CGCAGGAGTCTCCGATGTGG
1180.LmxM.01.0800chk_In_F	GCGACCGCAGGAACATCTCT
1181.LmxM.01.0800chk_In_R	CTTCCGCCTTCTCGTCTCGT
1182.LmxM.08_29.0680chk_Up_F	TCGCGGATGATCTCTGGCTG
1183.LmxM.08_29.0680chk_In_F	CACCCTGAAATTCGTCGATAACCA
1184.LmxM.08_29.0680chk_In_R	TGGGAGCGCAAGATCGACAA
1185.LmxM.11.0470chk_Up_F	TTCTATGACTGCCGGTGCGT
1186.LmxM.11.0470chk_In_F	CTCCGCCAAAGGAAGCCAAC
1187.LmxM.11.0470chk_In_R	GGGGACACGATTCCTGCCTT
1188.LmxM.27.1300chk_Up_F	GATGCGCTCTGTCGGGATCT
1189.LmxM.27.1300chk_In_F	GTCGGTTTCTTTCTCCTAACAGGC
1190.LmxM.27.1300chk_In_R	GGCAAGGTGCGCTACACAGA
1191.LmxM.34.2200chk_Up_F	CCACGGCCATAGGCACTACC
1192.LmxM.34.2200chk_In_F	GCTCGAGGACAAGCAAAGGC
1193.LmxM.34.2200chk_In_R	TGAGCACGATCACGGCTGAG
1194.LmxM.36.1635chk_Up_F	CGGAGGTGGCCCTTCTTACC
1195.LmxM.36.1635chk_In_F	CGCTCATCTGGAAGTCGC
1196.LmxM.36.1635chk_In_R	GCTCTTATCCGCGCTGTG
1220.DDX27seq_Long_F	ATGTTCCGTGTTTCCTCTGTTTCCC
1221.DDX27seq_Short_F	CGTATGGGTACATAAGCTTCGTTGCGAACTCAT CGCTGAAGTTGTTGTTTTGAC

1226.RPB16_Ct_3xHA_F_(1)	GCAACGAAGCTTATGTACCCATACG
1227.RPB16_Ct_3xHA_R_(1)	CCAATTTGAGAGACCTGTGC
1228.RPB16_Ct_3xHA_F_(2)	GCACAGGTCTCTCAAATTGGGTGCCAGGAAGA AAGTTAAAGACAAGAACC
1229.RPB16_Ct_3xHA_R_(2)	CGCAGAAACATGAACAGGAGCG
1230.RPB16_Ct_3xHA_F_(3)	TGCTCGTGGATCCGTCTACGCACTGCCCCACCT T
1231.RPB16_Ct_3xHA_R_(3)	GCTAACTGCATGCGGAGTGGTGCTGATCGAAGC AC
1232.NMT 3'UTR_F	CTCATGAACTCCTTAATGATTGC
1233.NMT 3'UTR_R	CTACAGGCTAGCGCGTCGACATGATTGAACAA GATGG
1234.mCherry_Nt_R	GCATCGCTCGAGCCTGCTGGATCCTCAGAAGAA CTCG
1235.Neo_Nhe1_F	CCCTCTCTCGCCTTTTGTTGCTTGCTCGTTTCAA GC
1236.Neo_XhoI_R	CAGATTACGCTTAATCAGGTTCTGGATCCGTCT ACGC
1246.RPB16ct3HA_CheckInt_F	GCGTAGACGGATCCAGAACCTGATTAAGCGTA ATCTG
1247.RPB16_Ct_3xHA_AddTAA_F	CCATTTCCTTCGCAGAGTTTC
1248.RPB16_Ct_3xHA_AddTAA_R	CCTCTTCTCTTTCGTTGACTCC
1263.LmxM.05.0850_Up_F	GGCAAAGTTTGGGATCTCC
1264.LmxM.05.0850_Dn_R	GGATCGGGCGAACTCCGCCTTCTTGCCGATAAG CTTACCAGAACC
1265.LmxM.05.0850_5'sgRNA	ACTTCCAGATGAGCGTCCGCGGCCTCGGATAAG CTTACCAGAACC
1266.LmxM.05.0850_3'sgRNA	GTCGGACGCGGTAGTCCAGTTAGCAGCGATAA GCTTACCAGAACC
1267.LmxM.08_29.2830_Up_F	ACCGACACCACCGCAGCGGTAGCAGACGATAA GCTTACCAGAACC
1268.LmxM.08_29.2830_Dn_R	GCCATCGAAGTCCGCAAGATCGCTGCTGATAAG CTTACCAGAACC
1269.LmxM.08_29.2830_5'sgRNA	AGGCGGCTGGCCAAAGTACGCTTGCTGGATAA GCTTACCAGAACC
1270.LmxM.08_29.2830_3'sgRNA	CAGGGCCTCGACAGCGTTCTGTTGAGAGATAAG CTTACCAGAACC
1271.LmxM.11.0600_Up_F	GTACCCTTTCTTTCCGTACGGACGCGAGATAAG CTTACCAGAACC
1272.LmxM.11.0600_Dn_R	GTATTCATCACGGGACGGACGCGGCGGGATAA GCTTACCAGAACC
1273.LmxM.11.0600_5'sgRNA	CGACCCGTGCGAGTGGCGCTCCTCCATGATAAG CTTACCAGAACC
1274.LmxM.11.0600_3'sgRNA	CGGCGTTCGCGAGTTCGCACCGGCAGGGATAA GCTTACCAGAACC
1275.LmxM.18.0590_Up_F	CGAAGAGGAGTTCGTGTTCGGCAGGTAGATAA GCTTACCAGAACC
1276.LmxM.18.0590_Dn_R	GCCGGGCGGCGGTCCTCTCAGATCCGTGATAAG CTTACCAGAACC
1277.LmxM.18.0590_5'sgRNA	GCTCTGGCTGTGAGTGGCTGCACGCTCGATAAG CTTACCAGAACC
1278.LmxM.18.0590_3'sgRNA	CCCACTGTTGGCGATATCCATGGGGAAGATAAG CTTACCAGAACC
1279.LmxM.36.0740_Up_F	TCGGCGGATCTTGTCGGTGAGATCGGAGATAAG CTTACCAGAACC
1280.LmxM.36.0740_Dn_R	ATCGTCGTAACGAGAGACCGCGTCCACGATAA GCTTACCAGAACC
1281.LmxM.36.0740_5'sgRNA	ACACCGCGCTCCCTTGCAGCAGCCACCGATAAG CTTACCAGAACC
1282.LmxM.36.0740_3'sgRNA	TGAGGTGATGAAAATGCACTTCTCGGGGATAA GCTTACCAGAACC
1283.LmxM.21.0540_Up_F	CGTACGCTGAAGGACGATCTTGTAGAAGATAA GCTTACCAGAACC
1284.LmxM.21.0540_Dn_R	GGCCTTGCCATGGGAATCCAAGCCGGAGATAA GCTTACCAGAACC

1285LmxM.21.0540_5'sgRNA	CACCCATACAGCTAGCGCCGTTACCGTGCTTAA TACGACTCACTATAAAACTGGAAGCTACACCCT AAAGTATAATGCAGACCTGCTGC
1286.LmxM.21.0540_3'sgRNA	CCTCTCTGCACATGTCCCCCGCCTCCTCCTCCAA TTTGAGAGACCTGTGC
1287.LmxM.29.3370_Up_F	GAAATTAATACGACTCACTATAGGGCTGCTCTT CTTGGACGCATGTTTTAGAGCTAGAAATAGC
1288.LmxM.29.3370_Dn_R	GAAATTAATACGACTCACTATAGGAACGTAGC AAGCGAAGGAGGGTTTTAGAGCTAGAAATAGC
1289.LmxM.29.3370_5'sgRNA	GCGTTGAATTTGCGCTATCGTTCTCCACCTTAAT ACGACTCACTATAAAACTGGAAGATCAATGTAT TTGTATAATGCAGACCTGCTGC
1290LmxM.29.3370_3'sgRNA	CCATCAGCCGCATAAGCATTCTCTCGTCATCCA ATTTGAGAGACCTGTGC
1291LmxM.30.2810_Up_F	GAAATTAATACGACTCACTATAGGTGTCGAGTT TTTCTAAGGTGGTTTTAGAGCTAGAAATAGC
1292.LmxM.30.2810_Dn_R	GAAATTAATACGACTCACTATAGGCAGCTTTGA TCGCTATGCTCGTTTTAGAGCTAGAAATAGC
1293.LmxM.30.2810_5'sgRNA	AACGTGCGTGTGCTCCGCATCACCCCTCCCTAA TACGACTCACTATAAAACTGGAAGTATGTAACT ACCGTATAATGCAGACCTGCTGC
1294.LmxM.30.2810_3'sgRNA	CCACTCTGCGCTCATAGCGGACAACTTCCTCCA ATTTGAGAGACCTGTGC
1295.LmxM.31.1280_Up_F	GAAATTAATACGACTCACTATAGGGCTTGTGTT CCTGTGATTGGGTTTTAGAGCTAGAAATAGC
1296.LmxM.31.1280_Dn_R	GAAATTAATACGACTCACTATAGGGTGCAACA GTGTCTTGCGAAGTTTTAGAGCTAGAAATAGC
1297.LmxM.31.1280_5'sgRNA	TTCTCCCTCCCTCTGCGCTTCACTTTTCCGTAAT ACGACTCACTATAAAACTGGAAGTAGTTTGTAT TTGTATAATGCAGACCTGCTGC
1298.LmxM.31.1280_3'sgRNA	CGCTTCGCCGCTCCGCACCCGTCCCTGCCGCCA ATTTGAGAGACCTGTGC
1299.LmxM.34.4950_Up_F	GAAATTAATACGACTCACTATAGGAAAAAACA GCGGTCGAGGAGGTTTTAGAGCTAGAAATAGC
1300.LmxM.34.4950_Dn_R	GAAATTAATACGACTCACTATAGGCTTCAGAGC GATGATGAGGAGTTTTAGAGCTAGAAATAGC
1301.LmxM.34.4950_5'sgRNA	CTCTTCCATCTTCCCTTTGCCCCTCGTCCATAAT ACGACTCACTATAAAACTGGAAGATACAAACT CCCGTATAATGCAGACCTGCTGC
1302.LmxM.34.4950_3'sgRNA	TTCAAAGCTGGCGAACGAAAAGAGATGCCCCC AATTTGAGAGACCTGTGC
1303.LmxM.21.1552chk_Up_F	GAAATTAATACGACTCACTATAGGAAAAAAAG AGTATGCAAATGGTTTTAGAGCTAGAAATAGC
1304.LmxM.28.0825chk_Up_F	GAAATTAATACGACTCACTATAGGTGAGTGCCT CTCAAATCTCGGTTTTAGAGCTAGAAATAGC
1305.LmxM.27.1300chk_Up_F	CTGCCGTTCACGCGCCTCACGTGCTCCCCTTAA TACGACTCACTATAAAACTGGAAGGAGCCCAT AATTGTATAATGCAGACCTGCTGC
1306.LmxM.11.0470_Nt3HA_R	CCAGCAGATCCCCCTGCCGAGAAAGCAGCACC AATTTGAGAGACCTGTGC
1307.LmxM.15.0130_Nt3HA_R	GAAATTAATACGACTCACTATAGGCCAACGGTG CTCAGTGGGGTGTTTTAGAGCTAGAAATAGC
1308.LmxM.18.1420_Nt3HA_R	GAAATTAATACGACTCACTATAGGAGCACGGC CACGCATGCTCGGTTTTAGAGCTAGAAATAGC
1309.LmxM.36.1635_Nt3HA_R	TCCCTCCAGCTCTTCCCCCCCCCCCAACCGTAAT ACGACTCACTATAAAACTGGAAGCGCGTTTCTT AAGTATAATGCAGACCTGCTGC
1310.LmxM.21.1552_Nt3HA_R	GAATCGAATCACTTCCCACGCAAGCCGCCGCCA ATTTGAGAGACCTGTGC
1311.LmxM.27.0130_Nt3HA_R	GAAATTAATACGACTCACTATAGGAGACACTG ACGCGGCTCGTGGTTTTAGAGCTAGAAATAGC
1312.LmxM.27.1300_Nt3HA_R	GAAATTAATACGACTCACTATAGGGCACGTGAC CTTGTCAGGAAGTTTTAGAGCTAGAAATAGC
1313.LmxM.27.1680_Nt3HA_R	CTTTCGGCTGATACAGAAGGAAAGCCTCAATAA TACGACTCACTATAAAACTGGAAGCCTATTACG TCCGTATAATGCAGACCTGCTGC
1314.LmxM.33.2580_Nt3HA_R	TGGTCCCGGTGAAGGCGGCAAGCGGCGCTCCC AATTTGAGAGACCTGTGC
1315.LmxM.34.2200_Nt3HA_R	GAAATTAATACGACTCACTATAGGTAAAAAAA GACCGCTGTCTGGTTTTAGAGCTAGAAATAGC

1316.LmxM.05.0850_Nt3HA_R	GAAATTAATACGACTCACTATAGGACGCTCGCC CTTTATAAGTGGTTTTAGAGCTAGAAATAGC
1317.LmxM.08_29.2830_Nt3HA_R	AACCGCAGCACCACATAAGGTAGAGCGCTTTA ATACGACTCACTATAAAACTGGAAGGGCGAAG TCCTTGTATAATGCAGACCTGCTGC
1318.LmxM.11.0600_Nt3HA_R	TCCGATGCTGTGCTTCACTGTTCGTGCTTGCCAA TTTGAGAGACCTGTGC
1319.LmxM.18.0590_Nt3HA_R	GAAATTAATACGACTCACTATAGGGCTCGTTCA CACCTATTCACGTTTTAGAGCTAGAAATAGC
1320.LmxM.36.0740_Nt3HA_R	GAAATTAATACGACTCACTATAGGCTCAAGAA AACACGAACAAGGTTTTAGAGCTAGAAATAGC
1321.LmxM.21.0540_Nt3HA_R	CTTTTGACGTCATACTTTCTCGC
1322.LmxM.28.2620_Nt3HA_R	GACTTCGTCCAGGACTACG
1323.LmxM.29.3370_Nt3HA_R	CTTGAAGTCCTGTTCTTTGACG
1324.LmxM.30.2810_Nt3HA_R	GTTCCACCTCCCTCATTTCC
1325.LmxM.31.1280_Nt3HA_R	CTCTTCGCAGATGATGCAG
1326.LmxM.34.4950_Nt3HA_R	CTGGCGATCGAAGATGATG
1350.LmxM.17.0550_Up_F	CGTACTGCAGTAAGTACAGATATACG
1351.LmxM.17.0550_Dn_R	GGTGATTGAGTACATATGCGTC
1352.LmxM.17.0550_5'sgRNA	CGAAGAGAGTGCTAACGCC
1353.LmxM.17.0550_3'sgRNA	GGTTTTACTCACTGCGGTTTG
1354.LmxM.36.5845_Up_F	CCTACGGTAACACCGTCTC
1355.LmxM.36.5845_Dn_R	GGAGTAAAAGAAGCCACTGC
1356.LmxM.36.5845_5'sgRNA	GTGCATTTTCATCACCTCAGT
1357.LmxM.36.5845_3'sgRNA	CAGTAGCCCTTCAATCGCAG
1358.LmxM.36.6770_Up_F	GTCTCTGGCGTCGTTACTG
1359.LmxM.36.6770_Dn_R	CTGAATCGCAGTCTTCTCACC
1360.LmxM.36.6770_5'sgRNA	CCATTGTATCCTTCTTGATGGAGC
1361.LmxM.36.6770_3'sgRNA	CATGTCACTTCCGTAGTTTTCTATC
1362.LmxM.22.0060_Up_F	CTCTCCGCCTTCCTCTTTCTGCTCGCGCCTTAAT ACGACTCACTATAAAACTGGAAGGGACATGAC CGGGTATAATGCAGACCTGCTGC
1363.LmxM.22.0060_Dn_R	CTCGTGTCCGTGATGATTATGTCGCTGCCGCCA ATTTGAGAGACCTGTGC
1364.LmxM.22.0060_5'sgRNA	GAAATTAATACGACTCACTATAGGAATAGCGA CGAGGCTACGAGGTTTTAGAGCTAGAAATAGC
1365.LmxM.22.0060_3'sgRNA	GAAATTAATACGACTCACTATAGGTGCGTGAAT TCCGTGGCCTAGTTTTAGAGCTAGAAATAGC
1366.LmxM.22.1500_Up_F	TCGCCCTGCTGTCTTACTTAGCAGATTCCCTAAT ACGACTCACTATAAAACTGGAAGGGGCTTCAG GCCGTATAATGCAGACCTGCTGC
1367.LmxM.22.1500_Dn_R	GCCTTTCTCTCACCCCCTCGCGATGCGTCGCCA ATTTGAGAGACCTGTGC
1368.LmxM.22.1500_5'sgRNA	GAAATTAATACGACTCACTATAGGACTGCGCTG TAGCGCTGCAGGTTTTAGAGCTAGAAATAGC
1369.LmxM.22.1500_3'sgRNA	GAAATTAATACGACTCACTATAGGGTGTCTCGG TGGATGCTGTAGTTTTAGAGCTAGAAATAGC
1370.LmxM.24.1570_Up_F	GCGCGCACACACACAACGTGTGCGTTGCGTTAA TACGACTCACTATAAAACTGGAAGACATACGG CCGGGTATAATGCAGACCTGCTGC
1371.LmxM.24.1570_Dn_R	TTTACTTAAGCACCCGCACGCACACGTAAACCA ATTTGAGAGACCTGTGC
1372.LmxM.24.1570_5'sgRNA	GAAATTAATACGACTCACTATAGGTGACTTCGT TAGCGCTACATGTTTTAGAGCTAGAAATAGC
1373.LmxM.24.1570_3'sgRNA	GAAATTAATACGACTCACTATAGGGGAGAGCC CGAAGTGGTCGTGTTTTAGAGCTAGAAATAGC
1374.LmxM.30.0250_Up_F	AACAGGCGGAGGGCGGAAGACACAGAACCATA ATACGACTCACTATAAAACTGGAAGCATACACC ATCCGTATAATGCAGACCTGCTGC
1375.LmxM.30.0250_Dn_R	ACGGCCGACACAAACACGCACTAAACCCTTCC AATTTGAGAGACCTGTGC
1376.LmxM.30.0250_5'sgRNA	GAAATTAATACGACTCACTATAGGATGCAAAG GATTGCTTCCTGGTTTTAGAGCTAGAAATAGC

1377.LmxM.30.0250_3'sgRNA	GAAATTAATACGACTCACTATAGGCATAGGTAC ACTTCCTTTGTGTTTTAGAGCTAGAAATAGC
1378.LmxM.31.3490_Up_F	CATCAAAACAGCTTCACTAAAGAAAGGCGGTA ATACGACTCACTATAAAACTGGAAGGGATTGGT TCGGGTATAATGCAGACCTGCTGC
1379.LmxM.31.3490_Dn_R	AGGCGTTGTGAAGGGTCCGTGGACATACCCCCA ATTTGAGAGACCTGTGC
1380.LmxM.31.3490_5'sgRNA	GAAATTAATACGACTCACTATAGGAAGCCAAG TGCCAGCAACTGGTTTTAGAGCTAGAAATAGC
1381.LmxM.31.3490_3'sgRNA	GAAATTAATACGACTCACTATAGGATTGTGGCT CGTGTGGTGGAGTTTTAGAGCTAGAAATAGC
1382.LmxM.33.4560_Up_F	GCCTCCGCACACACAGGCACGCACACACCCTA ATACGACTCACTATAAAACTGGAAGGGCACGT CAGCCGTATAATGCAGACCTGCTGC
1383.LmxM.33.4560_Dn_R	TACCCCGACGCTACGCTTAAGACTTTAGGACCA ATTTGAGAGACCTGTGC
1384.LmxM.33.4560_5'sgRNA	GAAATTAATACGACTCACTATAGGGTGTGCGTG TGCGCAGGCGTGGTTTTAGAGCTAGAAATAGC
1385.LmxM.33.4560_3'sgRNA	GAAATTAATACGACTCACTATAGGGCAGAGAT ATTGAAAGAGGAGTTTTAGAGCTAGAAATAGC
1386.LmxM.05.0850chk_Up_F	TCCCCCTCTCGTTTTGTAGAGGTTTGCCTCTAAT ACGACTCACTATAAAACTGGAAGACTGGAATTC GGGTATAATGCAGACCTGCTGC
1387.LmxM.05.0850chk_In_F	CGCGGCGTTATTCGACTTGGCGTTTTCCCTCCA ATTTGAGAGACCTGTGC
1388.LmxM.05.0850chk_In_R	GAAATTAATACGACTCACTATAGGAGTAGCTAC ACACAGGAGGAGTTTTAGAGCTAGAAATAGC
1389.LmxM.08_29.2830chk_Up_F	GAAATTAATACGACTCACTATAGGCTTCTCTCT CGTCGTTCGAGGTTTTAGAGCTAGAAATAGC
1390.LmxM.08_29.2830chk_In_F	TTTGTTGCAGGGGGCCTGCTGAACACTCCCTAA TACGACTCACTATAAAACTGGAAGAAGTCGAG AGCCGTATAATGCAGACCTGCTGC
1391.LmxM.08_29.2830chk_In_R	AACACACACACACAGAAGAAAAAAGTCCCACC AATTTGAGAGACCTGTGC
1392.LmxM.11.0600chk_Up_F	GAAATTAATACGACTCACTATAGGACACTCGCA CGTAAACGTACGTTTTAGAGCTAGAAATAGC
1393.LmxM.11.0600chk_In_F	GAAATTAATACGACTCACTATAGGACAGAGAC GCCCGTCCTCGCGTTTTAGAGCTAGAAATAGC
1394.LmxM.11.0600chk_In_R	CTTCCCCAGCGCACTGCACGCATCTCACCATAA TACGACTCACTATAAAACTGGAAGGGCATCGAT CTTGTATAATGCAGACCTGCTGC
1395.LmxM.18.0590chk_Up_F	GTGTGTGTGTGTGTGTGTGTGTGTGCATCCA ATTTGAGAGACCTGTGC
1396.LmxM.18.0590chk_In_F	GAAATTAATACGACTCACTATAGGTGTATGCGT GTGTGCGTGTGGTTTTAGAGCTAGAAATAGC
1397.LmxM.18.0590chk_In_R	GAAATTAATACGACTCACTATAGGGGAACGTG CCTGTTTAGGAAGTTTTAGAGCTAGAAATAGC
1398.LmxM.30.2810chk_In_F	CGGACCTCTTTCTCTTTTTCGTATCGTTTTTAAT ACGACTCACTATAAAACTGGAAGGGCAGTCTCT GGGTATAATGCAGACCTGCTGC
1399.LmxM.30.2810chk_In_R	CCCCCCTTTCAAGGGCTACGCCTCGGGCCCCCA ATTTGAGAGACCTGTGC
1400.LmxM.31.1280chk_Up_F	GAAATTAATACGACTCACTATAGGCTTCAGTGT TATTGCTGTCGGTTTTAGAGCTAGAAATAGC
1401.LmxM.31.1280chk_In_F	GAAATTAATACGACTCACTATAGGATGTGCTCG AGAGTGATGGCGTTTTAGAGCTAGAAATAGC
1402.LmxM.31.1280chk_In_R	GCGAAGAGCGAGCACGAACTTCAAAGGAGGTA ATACGACTCACTATAAAACTGGAAGCCTGCATC GCAAGTATAATGCAGACCTGCTGC
1403.LmxM.30.2810chk_Up_F	CGCACAGGCACAGGCGCAGACACGCGCATACC AATTTGAGAGACCTGTGC
1408.LmxM.04.1170_Up_F	GAAATTAATACGACTCACTATAGGCGGAGTTTT TTTTTAAGGAAGTTTTAGAGCTAGAAATAGC
1409.LmxM.04.1170_Dn_R	GAAATTAATACGACTCACTATAGGGATCAACA ACACGGTTGACAGTTTTAGAGCTAGAAATAGC
1410.LmxM.04.1170_5'sgRNA	CGTAGCAAGACGAACCGTTGG
1411.LmxM.04.1170_3'sgRNA	GGGTACAACATTGAGGTAGCAACG
1412.LmxM.05.0360_Up_F	CGTTGAGCGTTCCGTCATCG

1413.LmxM.05.0360_Dn_R	CGGAGTAGTCGATGAAGCTTGCC
1414.LmxM.05.0360_5'sgRNA	CCTTCCCACGTTAGCTCTCTCC
1415.LmxM.05.0360_3'sgRNA	CAGCATTCTCGCAGTCACAGC
1416.LmxM.14.1140_Up_F	GGTTCGTCGGATTCGACTGATG
1417.LmxM.14.1140_Dn_R	GGAGGGAGTACGAGCGAAGAG
1418.LmxM.14.1140_5'sgRNA	CGCATGATCCTGATGGGTCC
1419.LmxM.14.1140_3'sgRNA	CGTTATCCTTCGCGGTGACAC
1420.LmxM.36.1850_Up_F	CACTCGTCAGCATGACGCAC
1421.LmxM.36.1850_Dn_R	CTTTGCTCGTCGCAGTGATGTC
1422.LmxM.36.1850_5'sgRNA	CAGCGAGAACTCTGCCGTCTC
1423.LmxM.36.1850_3'sgRNA	CTCCGGAGTGGCTCTATCGC
1424.LmxM.36.5820_Up_F	GATGTGGAGTCTCGCCGTAACG
1425.LmxM.36.5820_Dn_R	GGGAACGTGGCAGAAAAGAGC
1426.LmxM.36.5820_5'sgRNA	GCACGTTCCTTTCTTGCCTGC
1427.LmxM.36.5820_3'sgRNA	GCTTTCAGAGCAGTGCACGTAC
1428.LmxM.25.1080_Up_F	CACCACTGCTTGTAGGCAATCC
1429.LmxM.25.1080_Dn_R	GCGCTGTAAACAACTCGGTGG
1430.LmxM.25.1080_5'sgRNA	GCTGTTTACCCGCCCCTCAG
1431.LmxM.25.1080_3'sgRNA	GTTCTGGTCCTTCTGCAGCAGC
1432.LmxM.26.1530_Up_F	CTGTCAACGTCACGGACAGC
1433.LmxM.26.1530_Dn_R	GTGTACGAGTGCCCCAGTTG
1434.LmxM.26.1530_5'sgRNA	GTAGCAGATGAGCGTCGTCGAC
1435.LmxM.26.1530_3'sgRNA	CGTCTGTCTTGGCGCATGTG
1436.LmxM.31.0950_Up_F	GGAGAATAAGGATGGCGCTGC
1437.LmxM.31.0950_Dn_R	CCAGTTTGCTGACAGCGTCATC
1438.LmxM.31.0950_5'sgRNA	CATCGTGATGGCGTCTGTGG
1439.LmxM.31.0950_3'sgRNA	CATCACGACAGTGTTCGTCACC
1440.LmxM.34.0370_Up_F	GCACGTTGTCAACGAAGGCAAC
1441.LmxM.34.0370_Dn_R	GGTGTATCCAAGTCTTTGTCACTGC
1442.LmxM.34.0370_5'sgRNA	CAACAGGGCTATCTAGCTGCTC
1443.LmxM.34.0370_3'sgRNA	CCCTTGATACCCATCGATGGTC
1444.LmxM.34.2270_Up_F	CGTTCGTCTCTTGTGCGGTAC
1445.LmxM.34.2270_Dn_R	CTTTGCCGACCTCACTTGCTTTG
1446.LmxM.34.2270_5'sgRNA	GTCGAAACAGTGAGACGGCC
1447.LmxM.34.2270_3'sgRNA	CTTCCTCCTCCCAAGACGCC
1448.LmxM.34.3200_Up_F	GCTGACACCTCCCACACAC
1449.LmxM.34.3200_Dn_R	CGTGCTGAGCACATACAAATAGACC
1450.LmxM.34.3200_5'sgRNA	CCCGCCACTTCTTTGTTCCC
1451.LmxM.34.3200_3'sgRNA	GTACCGGAGGAGTTCACCGAC
1456.PuroR_R	CGTGCGCATCCCGTACAATC
1457.LmxM.34.3200chk_Up_F	TTCAAAGAGCCACATTTTTTCCAAAAACCGTAA TACGACTCACTATAAAACTGGAAGGGCCGGCT AGATGTATAATGCAGACCTGCTGC
1458.Lmxm.34.3200chk_In_F	TGCGATTTTCAAATTTGAGTTTTCTCTGTGCCAA TTTGAGAGACCTGTGC
1459.LmxM.34.3200chk_In_R	GAAATTAATACGACTCACTATAGGTTGGTGTAA AAAAAAGGTAGGTTTTAGAGCTAGAAATAGC
1460.LmxM.26.1530chk_Up_F	GAAATTAATACGACTCACTATAGGATTTGCACC ACCCACGCAGCGTTTTAGAGCTAGAAATAGC
1461.LmxM.26.1530chk_In_F	GCACACACATACACACGACACACTGTTTCTTAA TACGACTCACTATAAAACTGGAAGAAGGAAGA GATAGTATAATGCAGACCTGCTGC
1462.LmxM.26.1530chk_In_R	TCACCGACTGCCGGCAGGGCGCTGCCGCCGCCA ATTTGAGAGACCTGTGC

1463.LmxM.31.0950chk_Up_F	GAAATTAATACGACTCACTATAGGCACACAACT ACTTCTATCAAGTTTTAGAGCTAGAAATAGC
1464.LmxM.31.0950chk_In_F	GAAATTAATACGACTCACTATAGGCCACGCTCC TCTCGTTCCGGGTTTTAGAGCTAGAAATAGC
1465.LmxM.31.0950chk_In_R	TATATATATATTTCCCCTCTCCTTCCCCCCTAAT ACGACTCACTATAAAACTGGAAGGGACGGCAT CTAGTATAATGCAGACCTGCTGC
1466.LmxM.14.1140chk_Up_F	CACCTTGAACGAATTTCGATGAATGTGCCTCCA ATTTGAGAGACCTGTGC
1467.LmxM.14.1140chk_In_F	GAAATTAATACGACTCACTATAGGAGGCACGG AGAGGGGTTGGTGTTTTAGAGCTAGAAATAGC
1468.LmxM.14.1140chk_In_R	GAAATTAATACGACTCACTATAGGAGAATGAA AACACCAGAAAAGTTTTAGAGCTAGAAATAGC
1469.LmxM.34.0370chk_Up_F	TCTGAAGACTCTGTGTTTCCTAGGGTCCCTTAAT ACGACTCACTATAAAACTGGAAGAAGGAAGGA GCGGTATAATGCAGACCTGCTGC
1470.LmxM.34.0370chk_In_F	GAAGGAGACCGCTGCGGCACGTTGTCACCGCC AATTTGAGAGACCTGTGC
1471.LmxM.34.0370chk_In_R	GAAATTAATACGACTCACTATAGGAGAATCACC TGAACTGAAAGGTTTTAGAGCTAGAAATAGC
1472.LmxM.05.0360chk_Up_F	GAAATTAATACGACTCACTATAGGGAGATACA CTAGGGCCGTGAGTTTTAGAGCTAGAAATAGC
1473.LmxM.05.0360chk_In_F	GGTTGCAGCGGCTCGTCTTTCTGAATCCCGTAA TACGACTCACTATAAAACTGGAAGGGACGGCG CTCGGTATAATGCAGACCTGCTGC
1474.LmxM.05.0360chk_In_R	ACGCCAACGCACTCACAGGGACGCGCACCACC AATTTGAGAGACCTGTGC
1475.LmxM.36.5820chk_Up_F	GAAATTAATACGACTCACTATAGGTCAGAAAC GGGGCAGAGAAAGTTTTAGAGCTAGAAATAGC
1476.LmxM.36.5820chk_In_F	GAAATTAATACGACTCACTATAGGGACCTTCAC CGCATCAGCTGGTTTTAGAGCTAGAAATAGC
1477.LmxM.36.5820chk_In_R	TGTGCGTGTACTTGAGTGCTGCTTCTTTCGTAAT ACGACTCACTATAAAACTGGAAGCCGGACTCTC GAGTATAATGCAGACCTGCTGC
1478.LmxM.22.0060chk_Up_F	CACCGCCCTCGCCTACACACACCATCACCACCA ATTTGAGAGACCTGTGC
1479.LmxM.22.0060chk_In_F	GAAATTAATACGACTCACTATAGGTGTCTCCCT CTTGAAGAAGAGTTTTAGAGCTAGAAATAGC
1480.LmxM.22.0060chk_In_R	GAAATTAATACGACTCACTATAGGGAACGATGT GTGCGTGACAGGTTTTAGAGCTAGAAATAGC
1481.LmxM.30.0250chk_Up_F	GCTCGTCTTTTCGGCTCGTAAGAGCTGCCTTAA TACGACTCACTATAAAACTGGAAGGGCCGGCC GAGCGTATAATGCAGACCTGCTGC
1482.LmxM.30.0250chk_In_F	GGAGATGGAGTGAACCGACAACCGCCGATGCC AATTTGAGAGACCTGTGC
1483.LmxM.30.0250chk_In_R	GAAATTAATACGACTCACTATAGGAGCGTCAGC GAACCTGCGAGGTTTTAGAGCTAGAAATAGC
1484.LmxM.17.0550chk_Up_F	GAAATTAATACGACTCACTATAGGAGACAAGC GGGATTAGGAGGGTTTTAGAGCTAGAAATAGC
1485.LmxM.17.0550chk_In_F	CCGGAAGTCGTTTTGCGTGATTCTCTTCGCTAAT ACGACTCACTATAAAACTGGAAGCCGGACTGA GCTGTATAATGCAGACCTGCTGC
1486.LmxM.17.0550chk_In_R	ATTCCTGAGGTCGTAGCGTCTTCGTTAGTACCA ATTTGAGAGACCTGTGC
1487.LmxM.22.1500chk_Up_F	GAAATTAATACGACTCACTATAGGGAACACAG CATTAGCCGCAAGTTTTAGAGCTAGAAATAGC
1488.LmxM.22.1500chk_In_F	GAAATTAATACGACTCACTATAGGGACTTTTAC ACGGAGCGCAGGTTTTAGAGCTAGAAATAGC
1489.LmxM.22.1500chk_In_R	TCACCGGTTTCTCCCGTGCTTCTTTCTCCCTAAT ACGACTCACTATAAAACTGGAAGGGACGCGGC AGTGTATAATGCAGACCTGCTGC
1490.LmxM.24.1570chk_Up_F	CTTTTGGCACTCGGGGGAGCAGTAAAACGTCCA ATTTGAGAGACCTGTGC
1491.LmxM.24.1570chk_In_F	GAAATTAATACGACTCACTATAGGTGCACGCAG GAATACAACGAGTTTTAGAGCTAGAAATAGC
1492.LmxM.24.1570chk_In_R	GAAATTAATACGACTCACTATAGGCGTCGCAGC CGCTGCACTTGGTTTTAGAGCTAGAAATAGC
1493.LmxM.36.6770chk_Up_F	TTTCGCGTTCGGCTTTATTCTCTTCTTCCCTAAT ACGACTCACTATAAAACTGGAAGCCGGAGAAG TAAGTATAATGCAGACCTGCTGC

1494.LmxM.36.6770chk_In_F	ATTCGATAAGCACAGCAAGAGAGCTTCCCGCC AATTTGAGAGACCTGTGC
1495.LmxM.36.6770chk_In_R	GAAATTAATACGACTCACTATAGGGTGTGTCGC TGTGCGTGGAGGTTTTAGAGCTAGAAATAGC
1496.LmxM.31.3490chk_Up_F	GAAATTAATACGACTCACTATAGGGGGGCAGC TACAACAACTTAGTTTTAGAGCTAGAAATAGC
1497.LmxM.31.3490chk_In_F	CCACCCTCCGAGCAACACTGTCGCTTACCATAA TACGACTCACTATAAAACTGGAAGGGCCGCGC GTCAGTATAATGCAGACCTGCTGC
1498.LmxM.31.3490chk_In_R	ACATACGAATAGAGAGAGCGACATGTATCGCC AATTTGAGAGACCTGTGC
1516.LmxM.13.0450_Up_F	GAAATTAATACGACTCACTATAGGCGTCTTGCA ATGCGGATGTTGTTTTAGAGCTAGAAATAGC
1517.LmxM.13.0450_Dn_R	GAAATTAATACGACTCACTATAGGGCTAGCCTT GTCTGCACTCAGTTTTAGAGCTAGAAATAGC
1518.LmxM.13.0450_5'sgRNA	ACACAGAGCGTCACCCTTTGCTACACTCCTTAA TACGACTCACTATAAAACTGGAAGCCGGAGAT CATTGTATAATGCAGACCTGCTGC
1519.LmxM.13.0450_3'sgRNA	TCGCTCGAGCAAAGCGCACGTGATCGCACGCC AATTTGAGAGACCTGTGC
1520.LmxM.19.0190_Up_F	GAAATTAATACGACTCACTATAGGAGTTGTTCG TTAGTGGGTCGGTTTTAGAGCTAGAAATAGC
1521.LmxM.19.0190_Dn_R	GAAATTAATACGACTCACTATAGGCATTCAACG CTGCTAAATATGTTTTAGAGCTAGAAATAGC
1522.LmxM.19.0190_5'sgRNA	CCCTCTTTCTCTCCACTCGCGCACCCTCCTTAAT ACGACTCACTATAAAACTGGAAGGGACGTACG CTTGTATAATGCAGACCTGCTGC
1523.LmxM.19.0190_3'sgRNA	AAAAAAAAAAACGGAGGCCATTCCTGCACGCC AATTTGAGAGACCTGTGC
1524.LmxM.23.0730_Up_F	GAAATTAATACGACTCACTATAGGAGAGAAGA GGTTCGTTCGACGTTTTAGAGCTAGAAATAGC
1525.LmxM.23.0730_Dn_R	GAAATTAATACGACTCACTATAGGAGCAGTAA CCGTGTCACCTCGTTTTAGAGCTAGAAATAGC
1526.LmxM.23.0730_5'sgRNA	GAAATTAATACGACTCACTATAGGTCTGCGTTG TGTGCGCTTTGTGTTTTAGAGCTAGAAATAGC
1527.LmxM.23.0730_3'sgRNA	CGCGCTTTGTGGTATGTTAACTGTCTCTGCGCC AATTTGAGAGACCTGTGC
1528.LmxM.25.0290_Up_F	GAGGTGCATGCAGCCAGAAC
1529.LmxM.25.0290_Dn_R	CGATTGTTCGTGTGAGCTCGAAC
1530.LmxM.25.0290_5'sgRNA	CTCATCATCGGCCTCACCTGC
1531.LmxM.25.0290_3'sgRNA	CGAGGGAGATGCCCATGAGG
1532.LmxM.25.2360_Up_F	CTTCAACTTCCAGCAGCGTGG
1533.LmxM.25.2360_Dn_R	GGAGTAGGACGAGTCTTGCCC
1534.LmxM.25.2360_5'sgRNA	CTTTCGTCTTCCTTTGCGTGGG
1535.LmxM.25.2360_3'sgRNA	GCAACAACTCGACGAAAGAGTTTTCTC
1536.LmxM.27.2100_Up_F	CCTCGCTGTAGTACAGAGCG
1537.LmxM.27.2100_Dn_R	GGTGCTGTGACTGCTCTCTTG
1538.LmxM.27.2100_5'sgRNA	GTTTGTGTACGTGAACGGCGAC
1539.LmxM.27.2100_3'sgRNA	CGGTTCCGACAACATATGGTAGATC
1540.LmxM.29.1110_Up_F	CTGCTTAGTGGTGCGTGGC
1541.LmxM.29.1110_Dn_R	CAGCAAGCTTCCTCAGTCAGATC
1542.LmxM.29.1110_5'sgRNA	CAGCTAACAAGAGGAGGCTGC
1543.LmxM.29.1110_3'sgRNA	GTCTGCCTGTCTTCCGTGTG
1544.LmxM.29.2200_Up_F	GCGACGGAGGATGACATCTTCG
1545.LmxM.29.2200_Dn_R	CAGCTTTAACGCCTTCTCGTACG
1546.LmxM.29.2200_5'sgRNA	CCATACACACATACTCGTGTACACG
1547.LmxM.29.2200_3'sgRNA	GCTGAGTACACGCTGTCGC
1548.LmxM.30.1650_Up_F	CTCGTAGGCGATCCAGTTCACG
1549.LmxM.30.1650_Dn_R	GCCTGAATGGAAGGTTTTGCG
1550.LmxM.30.1650_5'sgRNA	GACTTCCTCAAAAACACCGCGC

1551.LmxM.30.1650_3'sgRNA	GCCTCAATCGTTTTGCTGCACC
1552.LmxM.31.1750_Up_F	CGTGGCTCATCTCTTCGATCTGC
1553.LmxM.31.1750_Dn_R	GTGACGACGATGAACCGGTG
1554.LmxM.31.1750_5'sgRNA	CGTACAGGGCCGTCACAAC
1555.LmxM.31.1750_3'sgRNA	GTTTACGTCGACAAGACGCGG
1556.LmxM.32.1150_Up_F	GTTCACCGTCGATTTCATCGAGC
1557.LmxM.32.1150_Dn_R	GCAGGCTTCTCCTCCTCTTCC
1558.LmxM.32.1150_5'sgRNA	GCTTCTACGCGCCACAAGTG
1559.LmxM.32.1150_3'sgRNA	GTGCTCCACCGTCGAATGC
1560.LmxM.05.0140_Up_F	CTGCGCGAGTGGAAAGCC
1561.LmxM.05.0140_Dn_R	CGTCGCACTCCCATTCTCTC
1562.LmxM.05.0140_5'sgRNA	GCGGTATCGGTAACATCAACCCC
1563.LmxM.05.0140_3'sgRNA	GCTGGGGCTCATCATGTTGC
1564.LmxM.29.3090_Up_F	TGCTCGACTCATACCTTCGC
1565.LmxM.29.3090_Dn_R	GAGAGATGCGGATGGAGTGG
1566.LmxM.29.3090_5'sgRNA	GAAGAGCAGTTTGCCGTTCG
1567.LmxM.29.3090_3'sgRNA	CACATTCCGCTCTGTCCACT
1581.LmxM.30.1650_3'sgRNA(2)	CGCATCTCTCGAAACGGGTA
1582.LmxM.30.1650_Dn_R(2)	CACACACGCACGCACACACACACATACCCCTA ATACGACTCACTATAAAACTGGAAGGGCTGGTC ATACGTATAATGCAGACCTGCTGC
1583.LmxM.13.0450Chk_Up_F	CTCCCCTTCCCCTGCCCAACACCCCCCCCTCCA ATTTGAGAGACCTGTGC
1584.LmxM.13.0450Chk_In_F	GAAATTAATACGACTCACTATAGGGACTTTCAC GCGTCCGTGTGCGTTTTAGAGCTAGAAATAGC
1585.LmxM.13.0450Chk_In_R	GAAATTAATACGACTCACTATAGGAAGAGAGA TGGTGTGGCGAAGTTTTAGAGCTAGAAATAGC
1586.LmxM.19.0190Chk_Up_F	CGACGCCCACGTGCAGCGTCTCTCCGACCCTAA TACGACTCACTATAAAACTGGAAGGGCCAACA CTGCGTATAATGCAGACCTGCTGC
1587.LmxM.19.0190Chk_In_F	ACCTCTCACACACGCACAGTAATGCCGCCCCCA ATTTGAGAGACCTGTGC
1588.LmxM.19.0190Chk_In_R	GAAATTAATACGACTCACTATAGGTGAGTGTGC GCGTGTAGAGTGTTTTAGAGCTAGAAATAGC
1589.LmxM.23.0730Chk_Up_F	GAAATTAATACGACTCACTATAGGTAACTGATA AAGCTGCGGGTGTTTTAGAGCTAGAAATAGC
1590.LmxM.23.0730Chk_In_F	CTTCTGCGCGAACGCGAACACAGAGCGCCTTAA TACGACTCACTATAAAACTGGAAGCCGGGTTCC TTCGTATAATGCAGACCTGCTGC
1591.LmxM.23.0730Chk_In_R	GGATGACATACACACACACGCACGCCCGCC AATTTGAGAGACCTGTGC
1592.LmxM.25.0290Chk_Up_F	GAAATTAATACGACTCACTATAGGTGTTGCTGT TCTCCGCTCGAGTTTTAGAGCTAGAAATAGC
1593.LmxM.25.0290Chk_In_F	GAAATTAATACGACTCACTATAGGAACAGGCG GCGGAATGAGCTGTTTTAGAGCTAGAAATAGC
1594.LmxM.25.0290Chk_In_R	GTGTGTGTGTCTGTGCACCCTCTAGCGCCATAA TACGACTCACTATAAAACTGGAAGGGACAACT GACGGTATAATGCAGACCTGCTGC
1595.LmxM.25.2360Chk_Up_F	TCGCCAATATTACTAGACAGAAAACCACCACCA ATTTGAGAGACCTGTGC
1596.LmxM.25.2360Chk_In_F	GAAATTAATACGACTCACTATAGGAGGAGCAA TGAAAAGACAGAAGTTTTAGAGCTAGAAATAG C
1597.LmxM.25.2360Chk_In_R	GAAATTAATACGACTCACTATAGGCGGCGATG GGACTTTGCATGGTTTTAGAGCTAGAAATAGC
1598.LmxM.27.2100Chk_Up_F	GGTGCATCGCTGCACACCCACGCACACCCATAA TACGACTCACTATAAAACTGGAAGCCGGGTTAA GGAGTATAATGCAGACCTGCTGC
1599.LmxM.27.2100Chk_In_F	CACACGGGGGGAGCGGTGTTATTGCGTGTGCCA ATTTGAGAGACCTGTGC
1600.LmxM.27.2100Chk_In_R	GAAATTAATACGACTCACTATAGGCGAATTCCC TTCGTTGTGTGGTTTTAGAGCTAGAAATAGC

1601.LmxM.29.2200Chk_Up_F	GAAATTAATACGACTCACTATAGGGAAAGAAA CGACAGAAGGCAGTTTTAGAGCTAGAAATAGC
1602.LmxM.29.2200Chk_In_F	GCAGATTTCCGCTTTCCCAAGTCTTCCCCCTAAT ACGACTCACTATAAAACTGGAAGGGACAACGT CATGTATAATGCAGACCTGCTGC
1603.LmxM.29.2200Chk_In_R	CATTTCCTATTATCTGCTTCTTTTATGAGACCAA TTTGAGAGACCTGTGC
1604.LmxM.31.1750Chk_Up_F	GAAATTAATACGACTCACTATAGGAAAAGAAT GAGAGAGAGCGCGTTTTAGAGCTAGAAATAGC
1605.LmxM.31.1750Chk_In_F	GAAATTAATACGACTCACTATAGGCGACTACAT TGGTCGGAGTCGTTTTAGAGCTAGAAATAGC
1606.LmxM.31.1750Chk_In_R	GACAGACACCCCGTAGCAGATACGCAGCCATA ATACGACTCACTATAAAACTGGAAGGGACATG CACACGTATAATGCAGACCTGCTGC
1607.LmxM.05.0140Chk_Up_F	CAACAACGGGGAGTACGACAGCCCTCACCCCC AATTTGAGAGACCTGTGC
1608.LmxM.05.0140Chk_In_F	GAAATTAATACGACTCACTATAGGAGGGCCCTT GTCGATGTGTGGTTTTAGAGCTAGAAATAGC
1609.LmxM.05.0140Chk_In_R	GAAATTAATACGACTCACTATAGGCTCCTCTGT GCTACCGCTTTGTTTTAGAGCTAGAAATAGC
1610.LmxM.29.3090Chk_Up_F	ACACACTACCTCTTCTTTCTCCGCTCTGATTAAT ACGACTCACTATAAAACTGGAAGCCGGGAACC AAAGTATAATGCAGACCTGCTGC
1611.LmxM.29.3090Chk_In_F	GCCTACGAGCAGCGTCCCTCCTTCCTTACGCCA ATTTGAGAGACCTGTGC
1612.LmxM.29.3090Chk_In_R	GAAATTAATACGACTCACTATAGGCCTACACAC GTTTGCTAAGCGTTTTAGAGCTAGAAATAGC
1613.LmxM.29.1110Chk_Up_F	GAAATTAATACGACTCACTATAGGCTGCCCTCA TCAAAGCTAAAGTTTTAGAGCTAGAAATAGC
1614.LmxM.29.1110Chk_In_F	ATCACCTCGGCCTTCGAGTAATAAAAACCATAA TACGACTCACTATAAAACTGGAAGGGACATGG TGTGGTATAATGCAGACCTGCTGC
1615.LmxM.29.1110Chk_In_R	CACGCCCTCTGCCGAAGCATCCGGGCACCGCCA ATTTGAGAGACCTGTGC
1616.LmxM.32.1150Chk_Up_F	GAAATTAATACGACTCACTATAGGTGAAGAGG TGTGAAGGCTGTGTTTTAGAGCTAGAAATAGC
1617.LmxM.32.1150Chk_In_F	GAAATTAATACGACTCACTATAGGGGCATCTCC TTCGTGATGCAGTTTTAGAGCTAGAAATAGC
1618.LmxM.32.1150Chk_In_R	CTTTCTCTTACCCTCCCGACACACACACACTAA TACGACTCACTATAAAACTGGAAGAAGGGTTG GCCCGTATAATGCAGACCTGCTGC
1619.LmxM.30.1650Chk_Up_F	TCTCCTGCACACCCCGCATGTGCAGGCCCACCA ATTTGAGAGACCTGTGC
1620.LmxM.30.1650Chk_In_F	GAAATTAATACGACTCACTATAGGACACCGCG AGGTAGCAGAGAGTTTTAGAGCTAGAAATAGC
1621.LmxM.30.1650Chk_In_R	GAAATTAATACGACTCACTATAGGGAAGCAGA CAGCGGAAGAGCAGTTTTAGAGCTAGAAATAG C
1622.HALmxM.30.1650_Seq_Up_F	CACGTAACGGGGAGCGTGAGTGAAACGCAGTA ATACGACTCACTATAAAACTGGAAGGGACACA ACGTTGTATAATGCAGACCTGCTGC
1623.HALmxM.30.1650_Seq_Dn_R	AAAGAGGGAAGAAGGCTACCAGGACTGCCGCC AATTTGAGAGACCTGTGC
1628.LmxM.1030_Up_F	GAAATTAATACGACTCACTATAGGTAGCACAG GCAACCGTCTATGTTTTAGAGCTAGAAATAGC
1629.LmxM.1030_Dn_R	GAAATTAATACGACTCACTATAGGTGATGTCCG TCGTCGACGCGGTTTTAGAGCTAGAAATAGC
1630.LmxM.1030_5'sgRNA	TACACAGGCGCCCGGCACTACAAAGGACCGTA ATACGACTCACTATAAAACTGGAAGAAGGGTT CCGGGGTATAATGCAGACCTGCTGC
1631.LmxM.1030_3'sgRNA	AAGGAGCAGCCGGAGCCCTCACAGCACACACC AATTTGAGAGACCTGTGC
1632.LmxM.18.0220_Up_F	GAAATTAATACGACTCACTATAGGGCAGAGAT CGTGCTTGTGCTGTTTTAGAGCTAGAAATAGC
1633.LmxM.18.0220_Dn_R	GAAATTAATACGACTCACTATAGGGATCATTAA GGGAGGCGATGGTTTTAGAGCTAGAAATAGC
1634.LmxM.18.0220_5'sgRNA	GCATCTGCATGCTGTTCAACC
1635.LmxM.18.0220_3'sgRNA	CCACAATCCGAACCGTGCTG
1636.LmxM.19.0295_Up_F	GTCAGTGAGACCCAGCGTTG

1637.LmxM.19.0295_Dn_R	GCAGCAAGGCACGTAACAGAATG
1638.LmxM.19.0295_5'sgRNA	CCAAAGAGCGCCGTCAAGAAG
1639.LmxM.19.0295_3'sgRNA	GCTGCGGCTCTTCGAACTC
1640.LmxM.19.0790_Up_F	GCTACAGCCAACACGAACATCC
1641.LmxM.19.0790_Dn_R	GCTTTGGGTGCTACATCGGC
1642.LmxM.19.0790_5'sgRNA	GCGTACTACCGTCGCTGTTCG
1643.LmxM.19.0790_3'sgRNA	GGTACAGAAGAAGCTTGAATTGACAGACAC
1644.LmxM.36.0050_Up_F	GCCGTTTTATGGTCGGACGTAC
1645.LmxM.36.0050_Dn_R	CGAGGGCCATTAACGTCGATAACAC
1646.LmxM.36.0050_5'sgRNA	CGTTGGAGCGCCAAAGTCC
1647.LmxM.36.0050_3'sgRNA	CATCTTGTCCCTTCAGCCGAG
1648.LmxM.36.1620_Up_F	GCTCTGCTTGGAGCAGAGTGG
1649.LmxM.36.1620_Dn_R	CCAGTAGTGCCATCACCAACTTC
1650.LmxM.36.1620_5'sgRNA	GTTGACAGCCCTCTACGCGC
1651.LmxM.36.1620_3'sgRNA	CAGGCCACCGTGACAATATCG
1652.LmxM.25.0520_Up_F	GCAGATGAAGTCGACGTCCG
1653.LmxM.25.0520_Dn_R	CTGCTGCACAACATCACCCC
1654.LmxM.25.0520_5'sgRNA	GAGCCACAGTTGTAGCACGAC
1655.LmxM.25.0520_3'sgRNA	GCTTACCTTCCTCCATCCTCTCC
1656.LmxM.31.0400_Up_F	CGATCGCTACTCCTGTCTCGG
1657.LmxM.31.0400_Dn_R	GGATACATGCCTTGCGCATAGC
1658.LmxM.31.0400_5'sgRNA	GCTCTGCTGCCGAACATGTTTAG
1659.LmxM.31.0400_3'sgRNA	GCAACTTCAACCGCAGCTGC
1660.LmxM.31.0750_Up_F	CGTTCAGCACACACGGAATGC
1661.LmxM.31.0750_Dn_R	GCATGGTGGATCTCCCCTACG
1662.LmxM.31.0750_5'sgRNA	CGTGAAGAACTGGGGCACTG
1663.LmxM.31.0750_3'sgRNA	GGCTTGGGGCTGGTCTTAGC
1664.LmxM.31.3390_Up_F	CCATCACCCTCTTGCCCTCTC
1665.LmxM.31.3390_Dn_R	GCCTGTACACCATGCACTTCTCC
1666.LmxM.31.3390_5'sgRNA	CCTGAGCAGTTCAGCACAACG
1667.LmxM.31.3390_3'sgRNA	CCGAGCGTGTGCGTATTGG
1668.LmxM.32.0260_Up_F	CCATCTGCACTTTCTTTGCCATGG
1669.LmxM.32.0260_Dn_R	GCTTCTGAGAGTGCTGTGATTGTAGC
1670.LmxM.32.0260_5'sgRNA	GGACAGAAGACGTGTGAAGGGAG
1671.LmxM.32.0260_3'sgRNA	GCACCATCCGGTACACGTTG
1672.LmxM.33.4550_Up_F	GTCCCCATACGCCACCACC
1673.LmxM.33.4550_Dn_R	CACCGCCAAGGAAGAGTCATCG
1674.LmxM.33.4550_5'sgRNA	CGTGTACCCTCCGTGGGAG
1675.LmxM.33.4550_3'sgRNA	CGCAGAAACGGATAGAGGAAGAGG
1676.LmxM.34.2270chk_Up_F	GGCCGACGGGATGGAGGTGGGACGCGAGATAA GCTTACCAGAACC
1677.LmxM.34.2270chk_In_F	AATGCGGGAAACAGAGGAAACACGGAAGATAA GCTTACCAGAACC
1678.LmxM.34.2270chk_In_R	CGACCCGTGCGAGTGGCGCTCCTCGATAAGCTT ACCAGAACC
1679.LmxM.10.1030Chk_Up_F	AAAGCTGCCTCGGCCTCTTCCAGGCGGGATAAG CTTACCAGAACC
1680.LmxM.10.1030Chk_In_F	CGACGTCTGCACCAGGCGGCCGTCGATAAGCTT ACCAGAACC
1681.LmxM.10.1030Chk_In_R	GCCGAGCCGCCGAGGAACAAAGCGGAGGATAA GCTTACCAGAACC

1682.LmxM.18.0220Chk_Up_F	CTCGCGCGCCTTGCGCAGAAAGTTCTCGATAAG CTTACCAGAACC
1683.LmxM.18.0220Chk_In_F	GCGCTCATTCTCGACGTTGGCATCCGCGATAAG CTTACCAGAACC
1684.LmxM.18.0220Chk_In_R	ATCGCCTTTGTATTGTCCACCGGGATGGATAAG CTTACCAGAACC
1685.LmxM.19.0295Chk_Up_F	CGTCTGGGTCGGGGTGTGTGTGGGCTGGATAAG CTTACCAGAACC
1686.LmxM.19.0295Chk_In_F	CTTAATCTGTGGACCGAAGCGGCGAGGGATAA GCTTACCAGAACC
1687.LmxM.19.0295Chk_In_R	CCACCGCAGCAGGGGGCACTTGCGAAGGATAA GCTTACCAGAACC
1688.LmxM.19.0790Chk_Up_F	ACCACCCGTTTCGATGCCTTTGGATTCGATAAG CTTACCAGAACC
1689.LmxM.19.0790Chk_In_F	TTGACCGGGCACGAACTCCTGGCCGTAGATAAG CTTACCAGAACC
1690.LmxM.19.0790Chk_In_R	CTGGATGCCGAGTCGGTGAAACTCGGCGATAA GCTTACCAGAACC
1691.LmxM.36.0050Chk_Up_F	CAGCACGCGACAGGAGAAGTTTGACAGGATAA GCTTACCAGAACC
1692.LmxM.36.0050Chk_In_F	CTTGGGGAGCTTGCGCCGCGCTGACGAGATAA GCTTACCAGAACC
1693.LmxM.36.0050Chk_In_R	CACCAAGGCCACCATCGATCTCGAGAAGATAA GCTTACCAGAACC
1694.LmxM.36.1620Chk_Up_F	TCCGCGCCCCTTCTTCTGTTTGGCGGGGATAAG CTTACCAGAACC
1695.LmxM.36.1620Chk_In_F	CGCGCTCAGATTGGTGGTCGCTGTGTCGATAAG CTTACCAGAACC
1696.LmxM.36.1620Chk_In_R	ACTCTCCACCGCGTACACGACGTAGGAGATAA GCTTACCAGAACC
1697.LmxM.25.0520Chk_Up_F	CAGCTGAGCCTTCCAGTTCGAGTCGGAGATAAG CTTACCAGAACC
1698.LmxM.25.0520Chk_In_F	CTCAGCACGGTTCGGATTGTGGCCGCTGATAAG CTTACCAGAACC
1699.LmxM.25.0520Chk_In_R	GTACTGATTGAGGTCCTGTTCCGTTTTGATAAG CTTACCAGAACC
1700.LmxM.31.0400Chk_Up_F	GGCCTTCTTGACGGCGCTCTTTGGCGGGATAAG CTTACCAGAACC
1701.LmxM.31.0400Chk_In_F	AAAGCGGGTCTGCATGGGCGTGCCTGAGATAA GCTTACCAGAACC
1702.LmxM.31.0400Chk_In_R	CCGACCATAAAACGGCGACATGGCGCTGATAA GCTTACCAGAACC
1703.LmxM.31.0750Chk_Up_F	CGTTTTTTCCAGCTTCCACACCTCGGAGATAAG CTTACCAGAACC
1704.LmxM.31.0750Chk_In_F	GAGGGCTGTCAACTTCTCCTCTGACGAGATAAG CTTACCAGAACC
1705.LmxM.31.0750Chk_In_R	ACCACACTTGTAGCACGTGATAGCGGAGATAA GCTTACCAGAACC
1706.LmxM.31.3390Chk_Up_F	GGCGGCACTCCCGTTCTGCGCGAATGAGATAAG CTTACCAGAACC
1707.LmxM.31.3390Chk_In_F	CGGTTGAAGTTGCACCTGATTCTTATAGATAAG CTTACCAGAACC
1708.LmxM.31.3390Chk_In_R	GGCGGGCTTAGCGGCGGCCTTGGCGGGGATAA GCTTACCAGAACC
1709.LmxM.29.3370Chk_Up_F	GCCCGAGTGGCCCTCGCGTCCGACGTCGATAAG CTTACCAGAACC
1710.LmxM.29.3370Chk_In_F	GATGGTGCAAACGCCCTCCCCCACCACGATAAG CTTACCAGAACC
1711.LmxM.29.3370Chk_In_R	CCACGGAGGGTACACGGCCGGCGCCTCGATAA GCTTACCAGAACC
1712.LmxM.32.0260Chk_Up_F	AATCGCGTTTCCGCCGTTTTCAGCTGAGATAAG CTTACCAGAACC
1713.LmxM.32.0260Chk_In_F	CTGGCCCGTTGGAATAGCTCCCCTCGGGATAAG CTTACCAGAACC
1714.LmxM.32.0260Chk_In_R	GAACGTGTCGGACACGGGTGAGGTGCTGATAA GCTTACCAGAACC
1715.LmxM.33.4550Chk_Up_F	ACTCGCTGATACGGCGTCGCTGGAGATAAGCTT ACCAGAACC
1716.LmxM.33.4550Chk_In_F	CTCTTGCTTCTTCGCATTCGTGCGGCCGATAAG CTTACCAGAACC

1717.LmxM.33.4550Chk_In_R	CCCGGTCACGTAGACGGTGCAGGCACTGATAA GCTTACCAGAACC
1718.LmxM.01.0800_Nt3HA_R	CTTGCCCGGTGGAATCGGAGGCGCACTGATAA GCTTACCAGAACC
1719.LmxM.28.0825_Nt3HA_R	CTCGTAGGACGGAATGCGAGCAAGCTGGATAA GCTTACCAGAACC
1720.LmxM.34.2200_Nt3HA_R	GGGGGAGGAGTCGCGTGACTCGTCAGTGATAA GCTTACCAGAACC
1721.LmxM.08_29.0680_Nt3HA_R	TTTGAGGAAGTCCAGCTTCATTCCAGGGATAAG CTTACCAGAACC
1722.LmxM.17.0550_Nt3HA_R	GTGATCGTTCCAGCTCTGTTCAGAGTAGATAAG CTTACCAGAACC
1723.LmxM.36.5845_Nt3HA_R	CGCGTTCGCAACCTGCTGTGCGGCAGCGATAAG CTTACCAGAACC
1724.LmxM.36.6770_Nt3HA_R	TCGTCGGCAGCGTCAGATGTGTATAAGAGACA GAGATGATGATTACTAATACGACTCACTATAAA ACTGGAAG
1725.LmxM.22.0060_Nt3HA_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA GTCTCTCTCTCGTTTTCATCCG
1726.LmxM.22.1500_Nt3HA_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA CTCGTTTTCATCCGGCAG
1727.LmxM.24.1570_Nt3HA_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA GGCAGCAGGTCTGCATTATAC
1728.LmxM.30.0250_Nt3HA_R	GGTTTAAAACACACACAAACGCACCAG
1729.LmxM.31.3490_Nt3HA_R	CAGAGCATCTACTCCAACGTCAACAC
1730.LmxM.33.4560_Nt3HA_R	CTTGCTGCCAGCGTTGATGTTC
1731.LmxM.04.1170_Nt3HA_R	AAGCCCAAGGGCATTCGCCCGATTCGCCGATAC CCATACGATGTTCCAGATTAC
1732.LmxM.05.0360_Nt3HA_R	GTCGTCCTTCACGGGACTTTCGTTTTGCCACCA ATTTGAGAGACCTGTGC
1733.LmxM.14.1140_Nt3HA_R	AAGTCTCGCCATAAGCGCATGAAGAAGCACTA CCCATACGATGTTCCAGATTAC
1734.LmxM.36.1850_Nt3HA_R	CAGCCTCCTCTGCCCCTTCCCTGCAGCCCACCA ATTTGAGAGACCTGTGCG
1735.LmxM.36.5820_Nt3HA_R	ATGATGGGGGGACAGCCGCCTTCGAACCAGTA CCCATACGATGTTCCAGATTAC
1736.LmxM.25.1080_Nt3HA_R	CCAGCAGATCCCCCTGCCGAGAAAGCAGCACC AATTTGAGAGACCTGTGC
1737.LmxM.26.1530_Nt3HA_R	GCACCTCGCTCAAAGAAGTCGTATGTCCATTAC CCATACGATGTTCCAGATTAC
1738.LmxM.31.0950_Nt3HA_R	GAAGGAGACCGCTGCGGCACGTTGTCACCGCC AATTTGAGAGACCTGTGC
1739.LmxM.34.0370_Nt3HA_R	CTAGCCTGCATGCGGAGTGG
1740.LmxM.34.2270_Nt3HA_R	CCACTCCGCATGCAGGCTAG
1741.LmxM.34.3200_Nt3HA_R	GCGTCGATGTTACCGACACAGC
1742.LmxM.10.1030_Nt3HA_R	CAGCAGTGCTGGGGAATAGGAG
1743.LmxM.18.0220_Nt3HA_R	CAGCAACGGAAACAGTTCCAGCC
1744.LmxM.19.0295_Nt3HA_R	CATGGAATTGGGCACGGAAAGG
1745.LmxM.19.0790_Nt3HA_R	GGCACATTCCAAAGTGTGCGACG
1746.LmxM.36.0050_Nt3HA_R	CCTGTTAGGCGTGCTGGAGG
1747.LmxM.36.1620_Nt3HA_R	CGACGGTAAGCAGCGCAACA
1748.LmxM.25.0520_Nt3HA_R	GCAAGAGAGCACCAGTTAGCTCC
1749.LmxM.31.0400_Nt3HA_R	CACCATGGTTAGGCTCCCGG
1750.LmxM.31.0750_Nt3HA_R	CCGGGAGCCTAACCATGGTG
1751.LmxM.31.3390_Nt3HA_R	GTCTCGAGACGCTGACGAGC
1752.LmxM.32.0260_Nt3HA_R	CCTCATAGAGGTGCGCTGTCG
1753.LmxM.33.4550_Nt3HA_R	CCCACTCACACATTTGTGGCG
1754.LmxM.13.0450_Nt3HA_R	GGTAGTCTTCCATCAGCGGCAG
1755.LmxM.19.0190_Nt3HA_R	CCACACATTCCCAGCAGAGCC

1756.LmxM.23.0730_Nt3HA_R	CGCGTAGTCCTGGACGAAGTC
1757.LmxM.25.0290_Nt3HA_R	CTACACCCATTGCGCAGGTC
1758.LmxM.25.2360_Nt3HA_R	CCAAGAGCTCATCGGACGCC
1759.LmxM.27.2100_Nt3HA_R	AAGCCCAAGGGCATTCGCCCGATTCGCCGAGG ATCAGGTTCTGGTACCGGC
1760.LmxM.29.1110_Nt3HA_R	AAGTCTCGCCATAAGCGCATGAAGAAGCACGG ATCAGGTTCTGGTACCGGC
1761.LmxM.29.2200_Nt3HA_R	ATGATGGGGGGACAGCCGCCTTCGAACCAGGG ATCAGGTTCTGGTACCGGC
1762.LmxM.30.1650_Nt3HA_R	GCACCTCGCTCAAAGAAGTCGTATGTCCATGGA TCAGGTTCTGGTACCGGC
1763.LmxM.31.1750_Nt3HA_R	GATCAGTACGCTAGCGGAATTCATGCCTTTGTC TCAAG
1764.LmxM.32.1150_Nt3HA_R	TAGCAGTACCTCGAGCTAAATGGGCACCATGGT TAGC
1765.LmxM.29.3090_Nt3HA_R	CTACAGGGATCCACCGGTGCTAGCCTGCAGGGT ACCATGATTGAACAAGATGGATTGCACGC
1766.OL8684	CAGACTTCTAGACTCGAGCCTGCTCGATCATCA GAAGAACTCG
1767.OL8685_Alt	GTCAGCCCCGGGCGGCGCAGTGTTTCTTTCCAG
1768.OL8685_Alt2	CTCGAGCCTGCAGGCCCTTCTCAGCCTCCTCTG C
1769.OL8685	GTCAGCCCCGGGCCTACGACTGGATCCACGAAC G
1794.LmxM.18.0590Chk_Up_F	CTCGAGCCTGCAGGGGGTGGGGTTTCTAGTCGT CC
1795.LmxM.18.0590Chk_In_F	GTCAGCCCCGGGCACACGCCGACCTCTGAAGA C
1796.LmxM.18.0590Chk_In_R	CTCGAGCCTGCAGGCAGTCGCACACACCATCCG C
1897.LmxM.05.0850_Ct3H_F	GTCAGCCCCGGGGACTGTTCGCGAATCCCTGC
1898.LmxM.05.0850_Ct3H_R	CTCGAGCCTGCAGGGTGAGACGCACCCAGCAG ATC
1899.LmxM.15.0130_Ct3H_F	GACCAGATCGATCCGCATGCAGGCTAGGGCAC AG
1900.LmxM.15.0130_Ct3H_R	AGCACGATCGATGGTACCACTAGTCTCAGAAG AACTCGTCAAGAAGGCGATAGAAGG
1901.LmxM.24.1570_Ct3H_F	GACTGAGCTAGCCTGCAGACTAGTCATGATTGA ACAAGATGGATTGCACGCAGG
1902.LmxM.24.1570_Ct3H_R	CAGACTTCTAGACCAATTTGAGAGACCTGTGC
1903.LmxM.25.0290_Ct3H_F	AGTCTAACTAGTCCATGGTGCCCATTTAGTTGG C
1904.LmxM.25.0290_Ct3H_R	GTCTGACCCGGGGGGTGGGGTTTCTAGTCGTCC
1946.Ct3HAchk1R	GTCTGACCCGGGCCTTCTCAGCCTCCTCTGCC
1947.Ct3HAchk2F	GTCTGACCCGGGGAGACGCACCCAGCAG
1948.LmxM.24.1570_Ct3HAchk1F	GTCTGACCCGGGGCACAGAGAAGGAGACCGCT G
1949.LmxM.24.1570_Ct3HAchk2R	GCTCAGACTAGTGCTAGCCACTAGAGCTTATTT TATGGC
1950.LmxM.25.0290_Ct3HAchk1F	GTCAGCCCCGGGGTTGTTCTCTCGTTGCGGAGC
1951.LmxM.25.0290_Ct3HAchk2R	AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT TAA CTT GCT ATT TCT AGC TCT AAA AC
1952.LmxM.05.0850_Ct3HAchk1F	GCA GCA GGT CTG CAT TAT AC
1953.LmxM.05.0850_Ct3HAchk2R	GGATCCACGAACGTCACCCACAGTAAGCCCTA ATACGACTCACTATAAAACTGGAAGGATATCTA ATTTGTATAATGCAGACCTGCTGC
1954.LmxM.15.0130_Ct3HAchk1F	GTCGTCCTTCACGGGACTTTCGTTTTGCCACCA ATTTGAGAGACCTGTGC
1955.LmxM.15.0130_Ct3HAchk2R	GAAATTAATACGACTCACTATAGGCGCAAACTG CTGAATAGCAGGTTTTAGAGCTAGAAATAGC
1956.Nt3HAchk1R	GAAATTAATACGACTCACTATAGGAGTGGGAA GAAGAAAGGCGTGTTTTAGAGCTAGAAATAGC

1957.Nt3HAchk2F	TTCGGCGAATTTCACTTTAGTGGTTTGCCTTAAT ACGACTCACTATAAAACTGGAAGATCGATAGT AAAGTATAATGCAGACCTGCTGC
1958.LmxM.24.1570_Nt3HAchk1F	TGCGTTGCATCACCCGACCAATGGCACCCGCCA ATTTGAGAGACCTGTGC
1959.LmxM.24.1570_Nt3HAchk2R	GAAATTAATACGACTCACTATAGGAAACCGCA AAAAAAATCACAGTTTTAGAGCTAGAAATAGC
1960.LmxM.25.0290_Nt3HAchk1F	GAAATTAATACGACTCACTATAGGAGTCTGTCT GTGCACGTGTGGTTTTAGAGCTAGAAATAGC
1961.LmxM.25.0290_Nt3HAchk2R	CCACTTTTGCCACATCGCTTTGTCTTACCCTAAT ACGACTCACTATAAAACTGGAAGTGAGGACTA ATTGTATAATGCAGACCTGCTGC
1962.LmxM.05.0850_Nt3HAchk1F	ATTCACCACCACCGTCATCACCGCTCACCACCA ATTTGAGAGACCTGTGC
1963.LmxM.05.0850_Nt3HAchk2R	GAAATTAATACGACTCACTATAGGCTGAGCAGT GAAGGAGCGCTGTTTTAGAGCTAGAAATAGC
1964.LmxM.15.0130_Nt3HAchk1F	GAAATTAATACGACTCACTATAGGGTGCGCACA AGTTACTGTGAGTTTTAGAGCTAGAAATAGC
1965.LmxM.15.0130_Nt3HAchk2R	CCTCCTCTACTGCAAGCGGGGTTTCTTCCCTAAT ACGACTCACTATAAAACTGGAAGAATCAGGCTT AAGTATAATGCAGACCTGCTGC
1966.LmxM.05.0850_CtDD3H_F	ATGAGACGAAAGGGCGAGCGAGACCAAAACCC AATTTGAGAGACCTGTGC
1967.LmxM.15.0130_CtDD3H_F	GAAATTAATACGACTCACTATAGGAGAGGCCC AGCCCATACAAAGTTTTAGAGCTAGAAATAGC
1968.LmxM.24.1570_CtDD3H_F	GAAATTAATACGACTCACTATAGGAAGAGGGC CTCGTAAATTGTGTTTTAGAGCTAGAAATAGC
1969.LmxM.25.0290_CtDD3H_F	AGGGTAACTGTGGTGCTTGATCAGACAAACTAA TACGACTCACTATAAAACTGGAAGCAGAAACT ATTTGTATAATGCAGACCTGCTGC
1970.BlastAmp_F	GGTTCTTTCCTTTGAGTCGCTTTTTTTCCACCAA TTTGAGAGACCTGTGC
1971.BlastAmp_R	GAAATTAATACGACTCACTATAGGAACGTACA AAAAAACTTCAAGTTTTAGAGCTAGAAATAGC
2016.BamHI_NeoR_F	GAAATTAATACGACTCACTATAGGTGCACGATG GCAAGAGAGGAGTTTTAGAGCTAGAAATAGC
2017.XbaI_NeoR_R	ATCCCTCCCCCCTCCCCAGAGAAACTTCCTTAA TACGACTCACTATAAAACTGGAAGTGCTGTTCT ACCGTATAATGCAGACCTGCTGC
2018.XmaI_LmxM.15.0130Ct3HA_F	TCCAGAAGTCGGAGGCAGCGCACACTCGCGCC AATTTGAGAGACCTGTGC
2019.SbfI_LmxM.15.0130Ct3HA_R	GAAATTAATACGACTCACTATAGGCGTAATGCG GGGAAAGCGCAGTTTTAGAGCTAGAAATAGC
2020.XmaI_LmxM.05.0850Ct3HA_F	GAAATTAATACGACTCACTATAGGTGCGTGTGT TTGGGACCTATGTTTTAGAGCTAGAAATAGC
2021.SbfI_LmxM.05.0850Ct3HA_R	GTGCGTGGCAGAGAAACGAGGAGGTGACCCTA ATACGACTCACTATAAAACTGGAAGCATCATTC TACCGTATAATGCAGACCTGCTGC
2022.XmaI_LmxM.25.0290Ct3HA_F	AGTGAGATACGCTCTGCCCCCTTCTCTCCTCCA ATTTGAGAGACCTGTGC
2023.SbfI_LmxM.25.0290Ct3HA_R	GAAATTAATACGACTCACTATAGGGAAAGTGC AGATGGGGCGTTGTTTTAGAGCTAGAAATAGC
2024.XmaI_LmxM.24.1570Ct3HA_F	GAAATTAATACGACTCACTATAGGGCGTCTATT GGCAGACACCAGTTTTAGAGCTAGAAATAGC
2025.SbfI_LmxM.24.1570Ct3HA_R	CCCAACTCCCTTCTCCGCATACGATCCCTCTAAT ACGACTCACTATAAAACTGGAAGCGACCTGTAT TTGTATAATGCAGACCTGCTGC
2026.ClaI_NeoR_pSSU_Nt_F	CATCGGCCGCAGCGTCGTACACACCTTGCACCA ATTTGAGAGACCTGTGC
2027.ClaI_NeoR_pSSU_Nt_R	GAAATTAATACGACTCACTATAGGTTCTCACCC CGTCGACCCAAGTTTTAGAGCTAGAAATAGC
2028.NheI_NeoR_pSSU_Ct_F	GAAATTAATACGACTCACTATAGGTATGCTGGC TTACCTTCCTGGTTTTAGAGCTAGAAATAGC
2029.XbaI_NeoR_pSSU_Ct_R	AGGCTGCGCAGCGGACACCGAGACGCACCTTA ATACGACTCACTATAAAACTGGAAGGCGTGAA CTCCCGTATAATGCAGACCTGCTGC
2030.SpeI_pSSU_Nt_F	GTGGGACCACCTTGCAAGGAACGTGTGCACCC AATTTGAGAGACCTGTGC
2031.XmaI_LmxM.05.0850Nt3HA_R	GAAATTAATACGACTCACTATAGGATGGTTGTG TGCGCAACAGTGTTTTAGAGCTAGAAATAGC

2032.XmaI_LmxM.15.0130Nt3HA_R	GAAATTAATACGACTCACTATAGGTAAAGGAG GCTGCTGCTGCCGTTTTAGAGCTAGAAATAGC
2033.XmaI_LmxM.24.1570Nt3HA_R	TTCGTCCCTGTGTCCGCGTTCGCGTATCCTTAAT ACGACTCACTATAAAACTGGAAGGCTGGTGTAT TTGTATAATGCAGACCTGCTGC
2034.XmaI_LmxM.25.0290Nt3HA_R	GAGTCACTGGAAACACCTAACGCGCCACCACC AATTTGAGAGACCTGTGC
2035.SpeI_pSSU_Ct_R	GAAATTAATACGACTCACTATAGGTAATCTCCC CCCCCTCGTGCGTTTTAGAGCTAGAAATAGC
2036.XmaI_LmxM.24.1570Ct3HA_F	GAAATTAATACGACTCACTATAGGTCTATAACG GCGAGAGTGGAGTTTTAGAGCTAGAAATAGC

Table 7.3. (Supplementary). InterPro codes for potential RBDs. Codes corresponding to individual domains or classes of domains from RNA-binding proteins were compiled from the literature as well as the search term 'RNA-binding' within the InterPro database. Column one lists all compiled domains whereas column two lists a subset that was linked to proteins involved with the basal machinery of translation and splicing regulation.

RNA-binding domain InterPro codes	'Basal machinery' subset
IPR011545	IPR006195
IPR001313	IPR009022
IPR011016	IPR004038
IPR000571	IPR002942
IPR018539	IPR004364
IPR001878	IPR021131
IPR021133	IPR002314
IPR013699	IPR001222
IPR025714	IPR006134
IPR012340	IPR007120
IPR018835	IPR016082
IPR024261	IPR007083
IPR001890	IPR022669
IPR000208	IPR013155
IPR007098	IPR007641
IPR007097	IPR020783
IPR006196	IPR020784
IPR007099	IPR016180
IPR007094	IPR022666
IPR007096	IPR020598
IPR008744	IPR013823
IPR018834	IPR029344
IPR002775	IPR003034
IPR019447	IPR002784
IPR018999	IPR023798
IPR022967	IPR020040
IPR002942	IPR001351
IPR033712	IPR012606
IPR004341	IPR001912
IPR028157	IPR007010
IPR031952	IPR015413
IPR014886	IPR012988
IPR006903	IPR013845
IPR015848	IPR013845
IPR025223	IPR013843
IPR014720	IPR018164
IPR029344	IPR001975
IPR034992	IPR005633
IPR034440	IPR018165
IPR034445	IPR005146
IPR034417	IPR007639
IPR034418	IPR013150
IPR034434	IPR021891
IPR034435	IPR019580
IPR034980	IPR019581

IPR034991	IPR019582
IPR034993	IPR015016
IPR025995	IPR019349
IPR034130	IPR007094
IPR034126	IPR006116
IPR007046	
IPR014789	
IPR019385	
IPR022755	
IPR011113	
IPR034423	
IPR034419	
IPR034420	
IPR034421	
IPR013699	
IPR034451	
IPR033763	
IPR021891	
IPR023558	
IPR032828	
IPR020124	
IPR034264	
IPR034397	
IPR034398	
IPR005580	
IPR022666	
IPR007010	
IPR027391	
IPR034396	
IPR007120	
IPR034504	
IPR034588	
IPR034652	
IPR015245	
IPR031766	
IPR034869	
IPR034359	
IPR034773	
IPR034148	
IPR000504	
IPR034415	
IPR035236	
IPR034228	
IPR034903	
IPR034351	
IPR034858	
IPR034633	
IPR034167	
IPR034240	

IPR034500	
IPR034220	
IPR034221	
IPR034237	
IPR034798	
IPR034376	
IPR034653	
IPR034152	
IPR034501	
IPR034502	
IPR034787	
IPR033744	
IPR034206	
IPR034207	
IPR034208	
IPR034203	
IPR021975	
IPR034143	
IPR034854	
IPR034855	
IPR034856	
IPR034594	
IPR034591	
IPR034215	
IPR034900	
IPR034924	
IPR034925	
IPR034927	
IPR034177	
IPR034200	
IPR034784	
IPR034194	
IPR034510	
IPR034859	
IPR034195	
IPR034365	
IPR007083	
IPR034374	
IPR003954	
IPR034898	
IPR034897	
IPR034268	
IPR034467	
IPR034201	
IPR034363	
IPR034123	
IPR034125	
IPR034406	
IPR033110	

IPR034326	
IPR034394	
IPR034256	
IPR034492	
IPR034494	
IPR034496	
IPR034915	
IPR034166	
IPR006590	
IPR034983	
IPR034985	
IPR034986	
IPR034987	
IPR034356	
IPR034846	
IPR034645	
IPR034470	
IPR034472	
IPR034473	
IPR007641	
IPR034795	
IPR034796	
IPR034797	
IPR034475	
IPR034535	
IPR034536	
IPR035000	
IPR035001	
IPR034996	
IPR019582	
IPR034910	
IPR034887	
IPR018995	
IPR034894	
IPR034506	
IPR034507	
IPR035002	
IPR034843	
IPR034209	
IPR034211	
IPR034212	
IPR034552	
IPR034558	
IPR034357	
IPR034914	
IPR034538	
IPR034539	
IPR034918	
IPR034999	

IPR021083	
IPR034844	
IPR034361	
IPR034362	
IPR034424	
IPR034425	
IPR034426	
IPR034919	
IPR027351	
IPR034353	
IPR034509	
IPR034639	
IPR034842	
IPR034837	
IPR034862	
IPR034863	
IPR034520	
IPR034988	
IPR034217	
IPR034218	
IPR034969	
IPR034970	
IPR034522	
IPR034523	
IPR001604	
IPR034513	
IPR034516	
IPR034788	
IPR034799	
IPR034800	
IPR034971	
IPR034186	
IPR034823	
IPR034823	
IPR034825	
IPR035005	
IPR035008	
IPR034458	
IPR034525	
IPR034278	
IPR034630	
IPR034631	
IPR034878	
IPR007201	
IPR000607	
IPR014014	
IPR007638	
IPR007639	
IPR029262	

IPR034847	
IPR034816	
IPR034453	
IPR034454	
IPR034817	
IPR034979	
IPR034928	
IPR034930	
IPR018885	
IPR034792	
IPR034793	
IPR034794	
IPR034486	
IPR009715	
IPR034845	
IPR034803	
IPR005376	
IPR034404	
IPR034352	
IPR014881	
IPR021985	
IPR034830	
IPR034827	
IPR034832	
IPR034880	
IPR034977	
IPR009067	
IPR034199	
IPR034198	
IPR034517	
IPR006861	
IPR013197	
IPR032226	
IPR027352	
IPR027355	
IPR034196	
IPR033122	
IPR034131	
IPR034134	
IPR034544	
IPR034548	
IPR034181	
IPR034562	
IPR034564	
IPR034409	
IPR034411	
IPR034410	
IPR034414	
IPR034407	

IPR034648	
IPR035467	
IPR034168	
IPR034083	
IPR032292	
IPR020598	
IPR005146	
IPR005121	
IPR031164	
IPR002059	
IPR015096	
IPR015318	
IPR004191	
IPR019580	
IPR003166	
IPR019581	
IPR031988	
IPR013155	
IPR022642	
IPR019471	
IPR032335	
IPR001339	
IPR001222	
IPR033133	
IPR015225	
IPR015240	
IPR001878	
IPR000690	
IPR034076	
IPR018515	
IPR003604	
IPR013643	
IPR011760	
IPR012975	
IPR013158	
IPR028564	
IPR025761	
IPR025768	
IPR013584	
IPR023798	
IPR000897	
IPR004125	
IPR037517	
IPR012996	
IPR020460	
IPR013159	
IPR013029	
IPR013317	
IPR002694	

IPR006564	
IPR019401	
IPR014898	
IPR009349	
IPR015408	
IPR034079	
IPR034081	
IPR034068	
IPR001374	
IPR002999	
IPR031336	
IPR019814	
IPR019815	
IPR035625	
IPR005039	
IPR022023	
IPR013846	
IPR015016	
IPR002521	
IPR002522	
IPR031310	
IPR001912	
IPR031309	
IPR003851	
IPR010666	
IPR020040	
IPR005568	
IPR003034	
IPR008932	
IPR013087	
IPR002501	
IPR001876	
IPR004149	
IPR004198	
IPR004181	
IPR015088	
IPR025762	
IPR002693	
IPR005120	
IPR008395	
IPR037503	
IPR013822	
IPR034071	
IPR034077	
IPR001628	
IPR006134	
IPR007222	
IPR007527	
IPR002857	

IPR000967	
IPR007529	
IPR001607	
IPR021005	
IPR024766	
IPR011124	
IPR001351	
IPR012313	
IPR000571	
IPR006145	
IPR002653	
IPR018957	
IPR012934	
IPR006580	
IPR001293	
IPR005735	
IPR007684	
IPR006510	
IPR008899	
IPR010507	
IPR011723	
IPR017455	
IPR012935	
IPR011125	
IPR034082	
IPR002853	
IPR013137	
IPR022800	
IPR011129	
IPR019349	
IPR013823	
IPR012606	
IPR020784	
IPR012988	
IPR021131	
IPR020783	
IPR005633	
IPR002784	
IPR001975	
IPR004038	
IPR022669	
IPR013845	
IPR013843	
IPR000197	
IPR008913	
IPR002893	
IPR001510	
IPR020859	
IPR017921	

IPR006612	
IPR013085	
IPR020057	
IPR024657	
IPR033647	
IPR033648	
IPR035778	
IPR024161	
IPR000253	
IPR007900	
IPR000483	
IPR002483	
IPR006295	
IPR032570	
IPR007588	
IPR019406	
IPR004457	
IPR006642	
IPR013010	
IPR016180	
IPR006895	
IPR007872	
IPR000306	
IPR010603	
IPR000953	
IPR037406	
IPR023780	
IPR009071	
IPR019787	
IPR001965	
IPR006572	
IPR000962	
IPR015273	
IPR022776	
IPR026000	
IPR020838	
IPR011023	
IPR004364	
IPR006195	
IPR002314	
IPR018164	
IPR018165	
IPR015413	

Table 7.4. (Supplementary). All RBP knockout attempts. All attempts to delete *L.mexicana* RBPs in this study are listed below. If one or more attempts successfully removed all CDS copies, as confirmed by PCR, they are listed as 'KO' in green. Attempts labelled 'Remaining CDS' (yellow) resulted in parasites with only replacement of some RBP copies, with remaining CDS copies detected by PCR. Those labelled 'No clones' produced no viable parasites post transfection and are most likely to be essential in promastigotes. Some secondary knockout attempts were carried out by Natalia Teles as part of an independent study.

Gene ID	Knockout attempts	Result
LmxM.15.0130	3	KO
LmxM.18.1420	1	KO
LmxM.27.1300	1	Remaining CDS
LmxM.08_29.0680	1	Remaining CDS
LmxM.34.2200	2	No clones
LmxM.11.0470	1	Remaining CDS
LmxM.01.0800	1	Remaining CDS
LmxM.36.1635	1	KO
LmxM.27.1680	1	Remaining CDS
LmxM.28.0825	1	Remaining CDS
LmxM.27.0130	1	KO
LmxM.33.2580	1	KO
LmxM.21.1552	2	Remaining CDS
LmxM.34.4950	3	KO
LmxM.36.0740	2	No clones
LmxM.18.0590	1	KO
LmxM.08_29.2830	1	Remaining CDS
LmxM.21.0540	2	No clones
LmxM.29.3370	1	KO
LmxM.05.0850	1	No clones
LmxM.31.1280	1	Remaining CDS
LmxM.11.0600	2	Remaining CDS
LmxM.30.2810	1	Remaining CDS
LmxM.22.0060	2	KO
LmxM.30.0250	1	KO
LmxM.31.3490	3	No clones
LmxM.36.6770	1	KO
LmxM.24.1570	1	KO
LmxM.36.5845	2	No clones
LmxM.22.1500	1	KO
LmxM.17.0550	1	KO
LmxM.33.4560	1	No clones
LmxM.34.3200	1	No clones
LmxM.26.1530	1	KO
LmxM.25.1080	2	No clones
LmxM.04.1170	2	No clones
LmxM.31.0950	1	KO
LmxM.34.0370	2	No clones
LmxM.34.2270	1	KO
LmxM.14.1140	1	KO
LmxM.36.5820	1	KO
LmxM.05.0360	1	No clones
LmxM.36.1850	1	No clones
LmxM.13.0450	1	KO

LmxM.19.0190	1	KO
LmxM.23.0730	1	KO
LmxM.25.0290	1	KO
LmxM.25.2360	1	No clones
LmxM.27.2100	1	No clones
LmxM.29.1110	1	No clones
LmxM.29.2200	1	KO
LmxM.30.1650	1	KO
LmxM.29.3090	1	No clones
LmxM.31.1750	1	No clones
LmxM.32.1150	1	KO
LmxM.32.0260	1	No clones
LmxM.10.1030	2	Remaining CDS
LmxM.18.0220	2	Remaining CDS
LmxM.19.0295	2	KO
LmxM.19.0790	3	No clones
LmxM.36.0050	2	Remaining CDS
LmxM.36.1620	2	Remaining CDS
LmxM.25.0520	2	No clones
LmxM.31.0400	2	Remaining CDS
LmxM.31.0750	2	No clones
LmxM.31.3390	2	KO
LmxM.33.4550	2	KO

Table 7.5 (Supplementary).RBP null mutant confirmation PCR band sizes. Band sizes for all PCRs presented in Figure 4.7.

RBP KO	Integration PCR(bp)	CDS PCR(bp)
LmxM.27.0730	1006	510
LmxM.32.1150	1057	518
LmxM.29.2200	1116	450
LmxM.19.0190	988	429
LmxM.13.0450	879	338
LmxM.25.0290	1107	514
LmxM.31.0950	1158	592
LmxM.26.1530	1181	529
LmxM.14.1140	1171	604
LmxM.22.1500	816	510
LmxM.18.1420	1235	534
LmxM.30.1650	1190	524
LmxM.31.3390	1070	544
LmxM.19.0295	1073	522
LmxM.05.0850	671	348
LmxM.34.2270	1021	493
LmxM.36.5820	1053	468
LmxM.33.2580	948	324
LmxM.17.0550	1234	560
LmxM.24.1570	1090	567
LmxM.30.0250	1092	530
LmxM.36.1635	861	271
LmxM.27.0130	805	302
LmxM.15.0130	1290	694
LmxM.33.4550	1114	531
LmxM.29.3370	1182	564
LmxM.36.6770	1070	469
LmxM.18.0590	1100	519

Table 7.6 (Supplementary). Methylation status in the absence of PRMT7. Of the RBPs that were selected for knockout in this study, the methylation of the following were affected by the deletion of the PRMT7 gene (Ferreira *et al.*, 2020).

Hypomethylated		Hypermethylated	
Gene ID	Product Description	Gene ID	Product Description
LmxM.32.0260	RGG2	LmxM.05.0850	zinc-finger, C2HC-type, putative
LmxM.18.1420	PUF2	LmxM.27.1300	ZC3H41
LmxM.05.0850	zinc-finger, C2HC-type, putative	LmxM.27.1680	hypothetical protein, conserved
LmxM.32.1150	PUF6	LmxM.04.1170	DRBD3/4
LmxM.27.0130	ZFP3		
LmxM.28.0825	RBP16		
LmxM.33.2580	Alba3		
LmxM.25.0290	RBP43		
LmxM.27.1300	KH domain containing protein, putative		
LmxM.19.0190	multiprotein-bridging factor 1, putative		
LmxM.27.1680	hypothetical protein, conserved		
LmxM.34.2200	DRBD2		

8 References

Abbafati, C. *et al.* (2020) 'Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019', *The Lancet*, 396(10258), pp. 1204–1222. doi: 10.1016/S0140-6736(20)30925-9.

Adhya, S. *et al.* (1997) 'Role of an RNA-binding protein in import of tRNA into *Leishmania* mitochondria', *Journal of Biological Chemistry*, 272(34), pp. 21396–21402. doi: 10.1074/jbc.272.34.21396.

Agarwala, R. *et al.* (2016) 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, 44(D1), pp. D7–D19. doi: 10.1093/nar/gkv1290.

Alberts, B. *et al.* (2014) *Molecular Biology of the Cell*. 6th edn. Garland Science, Taylor and Francis Group, 2015.

Alborzi, A. *et al.* (2008) 'Case report: Isolation of Leishmania tropica from a patient with visceral leishmaniasis and disseminated cutaneous leishmaniasis, southern Iran', *American Journal of Tropical Medicine and Hygiene*, 79(3), pp. 435–437. doi: 10.4269/ajtmh.2008.79.435.

Alcantara, M. V. *et al.* (2018) 'Knockout of the CCCH zinc finger protein TcZC3H31 blocks *Trypanosoma cruzi* differentiation into the infective metacyclic form', *Molecular and Biochemical Parasitology*. Elsevier, 221(October 2017), pp. 1–9. doi: 10.1016/j.molbiopara.2018.01.006.

Alcolea, P. J. *et al.* (2010) 'Temperature increase prevails over acidification in gene expression modulation of amastigote differentiation in *Leishmania infantum*', *BMC Genomics*, 11(1). doi: 10.1186/1471-2164-11-31.

Alcolea, P. J. *et al.* (2019) 'An insight into the constitutive proteome throughout *Leishmania donovani* promastigote growth and differentiation', *International Microbiology*. International Microbiology, 22(1), pp. 143–154. doi: 10.1007/s10123-018-00036-2.

Alcolea, P. J., Alonso, A. and Larraga, V. (2011) 'Proteome profiling of *Leishmania infantum* promastigotes', *Journal of Eukaryotic Microbiology*, 58(4), pp. 352–358. doi: 10.1111/j.1550-7408.2011.00549.x.

Alexander, J. and Vickerman, K. (1975) 'Fusion of host cell secondary lysosomes with the parasitophorous vacuoles of *Leishmania mexicana*-infected macrophages', *The Journal of Protozoology*, 22(4), pp. 502–508. doi: 10.1111/j.1550-7408.1975.tb05219.x.

Allocco, J. J. *et al.* (2006) 'Inhibitors of casein kinase 1 block the growth of *Leishmania major* promastigotes in vitro', *International Journal for Parasitology*, 36(12), pp. 1249–1259. doi: 10.1016/j.ijpara.2006.06.013.

Almagro Armenteros, J. J. *et al.* (2019) 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*. Springer US, 37(4), pp. 420–423. doi: 10.1038/s41587-019-0036-z.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Alvar, J. *et al.* (2008) 'The relationship between leishmaniasis and AIDS: The second 10 years', *Clinical Microbiology Reviews*, 21(2), pp. 334–359. doi: 10.1128/CMR.00061-07.

Alvar, J. *et al.* (2012) 'Leishmaniasis worldwide and global estimates of its incidence', *PLoS ONE*, 7(5). doi: 10.1371/journal.pone.0035671.

Ammerman, M. L. *et al.* (2010) 'TbRGG2 facilitates kinetoplastid RNA editing initiation and progression past intrinsic pause sites.', *Rna*, 16(11), pp. 2239–51. doi: 10.1261/rna.2285510.

Ammerman, M. L., Fisk, J. C. and Read, L. K. (2008) 'gRNA/pre-mRNA annealing and RNA chaperone activities of RBP16', *RNA*, 14(6), pp. 1069–1080. doi: 10.1261/rna.982908.

Amorim, J. C. *et al.* (2017) 'Quantitative proteome and phosphoproteome analyses highlight the adherent population during *Trypanosoma cruzi* metacyclogenesis', *Scientific Reports*. Springer US, 7(1), pp. 1–12. doi: 10.1038/s41598-017-10292-3.

Andrade-Narvaez, F. J. *et al.* (2005) 'The histopathology of cutaneous leishmaniasis due to *Leishmania (Leishmania) mexicana* in the Yucatan Peninsula, Mexico', *Revista do Instituto de Medicina Tropical de Sao Paulo*, 47(4), pp. 191–194. doi: 10.1590/S0036-46652005000400003.

Aphasizhev, R. *et al.* (2003) 'A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components', *Rna*, 9(1), pp. 62–76. doi: 10.1261/rna.2134303.

Apweiler, R. *et al.* (2001) 'The InterPro database, an integrated documentation resource for protein families, domains and functional sites', *Nucleic Acids Research*, 29(1), pp. 37–40. doi: 10.1093/nar/29.1.37.

Archer, S. K. *et al.* (2009) '*Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle', *PLoS Pathogens*, 5(8). doi: 10.1371/journal.ppat.1000565.

Armenteros, J. J. A. *et al.* (2019) 'Detecting sequence signals in targeting peptides using deep learning', *Life Science Alliance*, 2(5), pp. 1–14. doi: 10.26508/lsa.201900429.

Armijos, R. X. *et al.* (2004) 'Comparison of the effectiveness of two topical paromomycin treatments versus meglumine antimoniate for New World cutaneous leishmaniasis', *Acta Tropica*, 91(2), pp. 153–160. doi: 10.1016/j.actatropica.2004.03.009.

Aschale, Y. *et al.* (2019) 'Malaria-visceral leishmaniasis co-infection and associated factors among migrant laborers in West Armachiho district, North West Ethiopia: Community based cross-sectional study', *BMC Infectious Diseases*. BMC Infectious Diseases, 19(1), pp. 1–7. doi: 10.1186/s12879-019-3865-y.

Ashburner, M. *et al.* (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25–29. doi: 10.1038/75556.

Aslett, M. *et al.* (2009) 'TriTrypDB: A functional genomic resource for the Trypanosomatidae', *Nucleic Acids Research*, 38(SUPPL.1), pp. 457–462. doi: 10.1093/nar/gkp851.

Atayde, V. D. *et al.* (2013) 'The structure and repertoire of small interfering RNAs in *Leishmania (Viannia) braziliensis* reveal diversification in the trypanosomatid RNAi pathway', *Molecular Microbiology*, 87(3), pp. 580–593. doi: 10.1111/mmi.12117.

Azizi, H., Dumas, C. and Papadopoulou, B. (2017) 'The Pumilio-domain protein PUF6 contributes to SIDER2 retroposon-mediated mRNA decay in *Leishmania*.', *RNA (New York, N.Y.)*, (418). doi: 10.1261/rna.062950.117.

Baker, N. *et al.* (2021) 'Systematic functional analysis of *Leishmania* protein kinases identifies regulators of differentiation or survival', *Nature Communications*. Springer US, 12(1). doi: 10.1038/s41467-021-21360-8.

Balagopal, V. and Parker, R. (2009) 'Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs', *Current Opinion in Cell Biology*, 21(3), pp. 403–408. doi: 10.1016/j.ceb.2009.03.005.

Balaña-Fouce, R. *et al.* (2014) 'Trypanosomatids topoisomerase re-visited. New structural findings and role in drug discovery', *International Journal for Parasitology: Drugs and Drug Resistance*, 4(3), pp. 326–337. doi: 10.1016/j.ijpddr.2014.07.006.

Balasegaram, M. *et al.* (2012) 'Liposomal amphotericin B as a treatment for human leishmaniasis', *Expert Opinion on Emerging Drugs*, 17(4), pp. 493–510. doi: 10.1517/14728214.2012.748036.

Baltz, A. G. *et al.* (2012) 'The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts', *Molecular Cell*, 46(5), pp. 674–690. doi: 10.1016/j.molcel.2012.05.021.

Barak, E. *et al.* (2005) 'Differentiation of *Leishmania donovani* in host-free system: Analysis of signal perception and response', *Molecular and Biochemical Parasitology*, 141(1), pp. 99–108. doi: 10.1016/j.molbiopara.2005.02.004.

Bari, A. U. *et al.* (2012) 'Mucocutaneous leishmaniasis in central Punjab and Azad Kashmir regions of Pakistan', *Journal of Pakistan Association of Dermatologists*, 22(3), pp. 191–196.

Barja, P. P. *et al.* (2017) 'Haplotype selection as an adaptive mechanism in the protozoan pathogen *Leishmania donovani*', *Nature Ecology and Evolution*. Springer US, 1(12), pp. 1961–1969. doi: 10.1038/s41559-017-0361-x.

Barquilla, A. *et al.* (2012) 'Third target of rapamycin complex negatively regulates development of quiescence in *Trypanosoma brucei*', *Proceedings of the National Academy of Sciences*, 109(36), pp. 14399–14404. doi: 10.1073/pnas.1210465109.

Barquilla, A., Crespo, J. L. and Navarro, M. (2008) 'Rapamycin inhibits trypanosome cell growth by preventing TOR complex 2 formation', *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), pp. 14579–14584. doi: 10.1073/pnas.0802668105.

Bassett, A. R., Kong, L. and Liu, J.-L. (2015) 'A Genome-Wide CRISPR Library for High-Throughput Genetic Screening in Drosophila Cells', *Journal of Genetics and Genomics*. Elsevier Limited and Science Press, 42(6), pp. 301–309. doi: 10.1016/j.jgg.2015.03.011.

Basu, J. M. *et al.* (2006) 'Sodium antimony gluconate induces generation of reactive oxygen species and nitric oxide via phosphoinositide 3-kinase and mitogen-activated protein kinase activation in *Leishmania donovani*-infected macrophages', *Antimicrobial Agents and Chemotherapy*, 50(5), pp. 1788–1797. doi: 10.1128/AAC.50.5.1788-1797.2006.

Bates, E. J. (2000) 'Poly(A)-binding protein I of *Leishmania*: functional analysis and localisation in trypanosomatid parasites', *Nucleic Acids Research*, 28(5), pp. 1211–1220. doi: 10.1093/nar/28.5.1211.

Bates, P. A. *et al.* (1992) 'Axenic cultivation and characterization of Leishmania *mexicana* amastigote-like forms', *Parasitology*. University of York, 105(2), pp. 193–202. doi: 10.1017/S0031182000074102.

Bates, P. A. (2008) '*Leishmania* sand fly interaction: progress and challenges', *Current Opinion in Microbiology*, 11(4), pp. 340–344. doi: 10.1016/j.mib.2008.06.003.

Beckmann, B. M. *et al.* (2015) 'The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs', *Nature Communications*, 6. doi: 10.1038/ncomms10127.

Beneke, T. *et al.* (2017) 'A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids', *Royal Society Open Science*, 4(5), p. 170095. doi: 10.1098/rsos.170095.

Beneke, T. *et al.* (2019) 'Genetic dissection of a *Leishmania* flagellar proteome demonstrates requirement for directional motility in sand fly infections', *PLoS Pathogens*, 15(6), pp. 1–31. doi: 10.1371/journal.ppat.1007828.

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006) 'Adaptive linear step-up procedures that control the false discovery rate', *Biometrika*, 93(3), pp. 491–507. doi: 10.1093/biomet/93.3.491.

Bennis, I. *et al.* (2017) 'Psychosocial impact of scars due to cutaneous leishmaniasis on high school students in Errachidia province, Morocco', *Infectious Diseases of Poverty*. Infectious Diseases of Poverty, 6(1), pp. 1–8. doi: 10.1186/s40249-017-0267-5.

Bentel, M. *et al.* (2003) 'Developmentally induced changes of the proteome in the protozoan parasite *Leishmania donovani*', *Proteomics*, 3(9), pp. 1811–1829. doi: 10.1002/pmic.200300462.

Benz, C. *et al.* (2011) 'The *Trypanosoma brucei* zinc finger protein ZC3H18 is involved in differentiation', *Molecular and Biochemical Parasitology*. Elsevier B.V., 177(2), pp. 148–151. doi: 10.1016/j.molbiopara.2011.02.007.

Berman, J. D. *et al.* (1998) 'Efficacy and safety of liposomal amphotericin B (AmBisome) for visceral leishmaniasis in endemic developing countries', *Bulletin of the World Health Organization*, 76(1), pp. 25–32.

Berman, J. D., Gallalee, J. V. and Hansen, B. D. (1987) '*Leishmania mexicana*: Uptake of sodium stibogluconate (Pentostam) and pentamidine by parasite and macrophages', *Experimental Parasitology*, 64(1), pp. 127–131. doi: 10.1016/0014-4894(87)90018-X.

Berman, J. D., Waddel, D. and Hanson, B. D. (1985) 'Biochemical mechanisms of the antileishmanial activity of sodium stibogluconate', *Antimicrobial Agents and Chemotherapy*, 27(6), pp. 916–920. doi: 10.1128/AAC.27.6.916.

Bertomeu, T. *et al.* (2018) 'A high-resolution genome-wide CRISPR/Cas9 viability screen reveals structural features and contextual diversity of the human cell-essential proteome', *Molecular and Cellular Biology*, 38(1), pp. 1–24. doi: 10.1128/MCB.00302-17.

Beverley, S. M. *et al.* (1984) 'Unstable DNA amplifications in methotrexate resistant *Leishmania* consist of extrachromosomal circles which relocalize during stabilization', *Cell*, 38(2), pp. 431–439. doi: 10.1016/0092-8674(84)90498-7.

Beverley, S. M. (1991) 'Gene amplification in *Leishmania*', *Annual Review of Microbiology*, 45(1), pp. 417–444. doi: 10.1146/annurev.mi.45.100191.002221.

Beverley, S. M. and Clayton, C. E. (1993) 'Transfection of *Leishmania* and *Trypanosoma brucei* by Electroporation', in *Protocols in Molecular Parasitology*. New Jersey: Humana Press, pp. 333–348. doi: 10.1385/0-89603-239-6:333.

Bhandari, D. *et al.* (2011) 'Ubiquitination of mRNA cycling sequence binding protein from *Leishmania donovani* (LdCSBP) modulates the RNA endonuclease activity of its Smr domain', *FEBS Letters*. Federation of European Biochemical Societies, 585(5), pp. 809–813. doi: 10.1016/j.febslet.2011.02.007.

Bhandari, D. and Saha, P. (2007) 'mRNA cycling sequence binding protein from *Leishmania donovani* (LdCSBP) is covalently modified by ubiquitination', *FEMS Microbiology Letters*, 273(2), pp. 206–213. doi: 10.1111/j.1574-6968.2007.00789.x.

Bijlmakers, M. J. *et al.* (2016) 'A C2HC zinc finger is essential for the RING-E2 interaction of the ubiquitin ligase RNF125', *Scientific Reports*, 6(June), pp. 1–14. doi: 10.1038/srep29232.

Blattner, J. and Clayton, C. E. (1995) 'The 3′-untranslated regions from the *Trypanosoma brucei* phosphoglycerate kinase-encoding genes mediate developmental regulation', *Gene*, 162(1), pp. 153–156. doi: 10.1016/0378-1119(95)00366-E.

Blum, M. *et al.* (2021) 'The InterPro protein families and domains database: 20 years on', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D344–D354. doi: 10.1093/nar/gkaa977.

van den Bogaart, E. *et al.* (2012) 'Prevalence, features and risk factors for malaria co-infections amongst visceral leishmaniasis patients from Amudat hospital, Uganda', *PLoS Neglected Tropical Diseases*, 6(4). doi: 10.1371/journal.pntd.0001617.

Bono, F. *et al.* (2006) 'The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA', *Cell*, 126(4), pp. 713–725. doi: 10.1016/j.cell.2006.08.006.

Boucher, N. *et al.* (2004) 'The ribosomal RNA gene promoter and adjacent cis-acting DNA sequences govern plasmid DNA partitioning and stable inheritance in the parasitic protozoan *Leishmania*', *Nucleic Acids Research*, 32(9), pp. 2925–2936. doi: 10.1093/nar/gkh617.

Branon, T. C. *et al.* (2018) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–887. doi: 10.1038/nbt.4201.

Brettmann, E. A. *et al.* (2016) 'Tilting the balance between RNA interference and replication eradicates *Leishmania* RNA virus 1 and mitigates the inflammatory response', *Proceedings of the*

National Academy of Sciences of the United States of America, 113(43), pp. 11998–12005. doi: 10.1073/pnas.1615085113.

Briggs, E. M. *et al.* (2021) 'Single-cell transcriptomic analysis of bloodstream *Trypanosoma brucei* reconstructs cell cycle progression and developmental quorum sensing', *Nature Communications*. Springer US, 12(1). doi: 10.1038/s41467-021-25607-2.

Brindha, J., Balamurali, M. M. and Chanda, K. (2021) 'An overview on the therapeutics of neglected infectious diseases—leishmaniasis and Chagas diseases', *Frontiers in Chemistry*, 9(March), pp. 1–19. doi: 10.3389/fchem.2021.622286.

Bringaud, F. *et al.* (2007) 'Members of a large retroposon family are determinants of post-transcriptional gene expression in *Leishmania*', *PLoS Pathogens*, 3(9), pp. 1291–1307. doi: 10.1371/journal.ppat.0030136.

Britto, C. *et al.* (1998) 'Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes', *Gene*, 222(1), pp. 107–117. doi: 10.1016/S0378-1119(98)00472-7.

Brochu, C. *et al.* (2003) 'Antimony uptake systems in the protozoan parasite *Leishmania* and accumulation differences in antimony-resistant parasites', *Antimicrobial agents and chemotherapy*, 47(10), pp. 3073–3079. doi: 10.1128/AAC.47.10.3073.

Buchan, J. R. and Parker, R. (2009) 'Eukaryotic Stress Granules: The Ins and Outs of Translation', *Molecular Cell*. Elsevier Ltd, 36(6), pp. 932–941. doi: 10.1016/j.molcel.2009.11.020.

Burza, S. *et al.* (2014) 'HIV and visceral leishmaniasis coinfection in Bihar, India: An underrecognized and underdiagnosed threat against elimination', *Clinical Infectious Diseases*, 59(4), pp. 552–555. doi: 10.1093/cid/ciu333.

Burza, S., Croft, S. L. and Boelaert, M. (2018) 'Leishmaniasis', *The Lancet*, 392(10151), pp. 951–970. doi: 10.1016/S0140-6736(18)31204-2.

Bussotti, G. *et al.* (2018) '*Leishmania* genome dynamics during environmental adaptation reveal strain-specific differences in gene copy', *mBio*, 9(6), pp. 1–18.

Cabañas, M. J., Vázquez, D. and Modolell, J. (1978) 'Dual interference of Hygromycin B with ribosomal translocation and with aminoacyl-tRNA recognition', *European Journal of Biochemistry*, 87(1), pp. 21–27. doi: 10.1111/j.1432-1033.1978.tb12347.x.

Carbon, S. *et al.* (2021) 'The Gene Ontology resource: Enriching a GOld mine', *Nucleic Acids Research*, 49(D1), pp. D325–D334. doi: 10.1093/nar/gkaa1113.

Carnes, J. *et al.* (2012) 'Mutational analysis of *Trypanosoma brucei* editosome proteins KREPB4 and KREPB5 reveals domains critical for function', *Rna*, 18(10), pp. 1897–1909. doi: 10.1261/rna.035048.112.

Carnielli, J. B. T. *et al.* (2014) ‘Proteomic analysis of the soluble proteomes of miltefosine-sensitive and -resistant *Leishmania infantum chagasi* isolates obtained from Brazilian patients with different treatment outcomes’, *Journal of Proteomics*. Elsevier B.V., 108, pp. 198–208. doi: 10.1016/j.jprot.2014.05.010.

Carnielli, J. B. T. *et al.* (2018) ‘A *Leishmania infantum* genetic marker associated with miltefosine treatment failure for visceral leishmaniasis’, *EBioMedicine*. The Authors, 36, pp. 83–91. doi: 10.1016/j.ebiom.2018.09.029.

Caro, F. *et al.* (2006) ‘*Trypanosoma cruzi*: Analysis of the complete PUF RNA-binding protein family’, *Experimental Parasitology*, 113(2), pp. 112–124. doi: 10.1016/j.exppara.2005.12.015.

Caruthers, J. M., Johnson, E. R. and McKay, D. B. (2000) ‘Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 97(24), pp. 13080–13085. doi: 10.1073/PNAS.97.24.13080.

Cassola, A. *et al.* (2015) ‘Association of UBP1 to ribonucleoprotein complexes is regulated by interaction with the trypanosome ortholog of the human multifunctional P32 protein’, *Molecular Microbiology*, 97(6), pp. 1079–1096. doi: 10.1111/mmi.13090.

Castello, A. *et al.* (2012) ‘Insights into RNA biology from an atlas of mammalian mRNA-binding proteins’, *Cell*. Elsevier Inc., 149(6), pp. 1393–1406. doi: 10.1016/j.cell.2012.04.031.

Castello, A. *et al.* (2017) ‘Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap’, *Nature Protocols*. Nature Publishing Group, 12(12), pp. 2447–2464. doi: 10.1038/nprot.2017.106.

Cavalier-Smith, T. (2016) ‘Higher classification and phylogeny of Euglenozoa’, *European Journal of Protistology*. Elsevier GmbH, 56, pp. 250–276. doi: 10.1016/j.ejop.2016.09.003.

Cecílio, P. *et al.* (2020) ‘Engineering a vector-based pan-*Leishmania* vaccine for humans: proof of principle’, *Scientific Reports*. Nature Publishing Group UK, 10(1), pp. 1–20. doi: 10.1038/s41598-020-75410-0.

Chakravarty, J. *et al.* (2011) ‘A clinical trial to evaluate the safety and immunogenicity of the LEISH-F1+MPL-SE vaccine for use in the prevention of visceral leishmaniasis’, *Vaccine*. Elsevier Ltd, 29(19), pp. 3531–3537. doi: 10.1016/j.vaccine.2011.02.096.

Chakravarty, J. and Sundar, S. (2019) ‘Current and emerging medications for the treatment of leishmaniasis’, *Expert Opinion on Pharmacotherapy*. Taylor & Francis, 20(10), pp. 1251–1265. doi: 10.1080/14656566.2019.1609940.

Chang, K.-P. and Dwyer, D. (1978) ‘*Leishmania donovani*. Hamster macrophage interactions in vitro: cell entry, intracellular survival, and multiplication of amastigotes’, *Journal of Experimental Medicine*, 147(2), pp. 515–530. doi: 10.1084/jem.147.2.515.

Chávez-Fumagalli, M. A. *et al.* (2019) '*In silico Leishmania* proteome mining applied to identify drug target potential to be used to treat against visceral and tegumentary leishmaniasis', *Journal of Molecular Graphics and Modelling*, 87, pp. 89–97. doi: 10.1016/j.jmgm.2018.11.014.

Chen, M. and Manley, J. L. (2009) 'Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(11), pp. 741–754. doi: 10.1038/nrm2777.

Chen, S. *et al.* (2015) 'Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis', *Cell*. Elsevier Inc., 160(6), pp. 1246–1260. doi: 10.1016/j.cell.2015.02.038.

Chen, Y. T. *et al.* (2008) 'Synthesis of macrocyclic trypanosomal cysteine protease inhibitors', *Bioorganic and Medicinal Chemistry Letters*. Elsevier Ltd, 18(22), pp. 5860–5863. doi: 10.1016/j.bmcl.2008.06.012.

Chowdhury, R. *et al.* (2014) 'How far are we from visceral leishmaniasis elimination in Bangladesh? An assessment of epidemiological surveillance data', *PLoS Neglected Tropical Diseases*, 8(8). doi: 10.1371/journal.pntd.0003020.

Cirillo, D., Agostini, F. and Tartaglia, G. G. (2013) 'Predictions of protein-RNA interactions', *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2), pp. 161–175. doi: 10.1002/wcms.1119.

Clayton, C. (2019) 'Regulation of gene expression in trypanosomatids: Living with polycistronic transcription', *Open Biology*, 9(6). doi: 10.1098/rsob.190072.

Clayton, C. E. (2002) 'Life without transcriptional control? From fly to man and back again', *EMBO Journal*, 21(8), pp. 1881–1888. doi: 10.1093/emboj/21.8.1881.

Clayton, C. E. (2016) 'Gene expression in Kinetoplastids', *Current Opinion in Microbiology*. Elsevier Ltd, 32, pp. 46–51. doi: 10.1016/j.mib.2016.04.018.

Coelho, A. C. *et al.* (2012) 'Multiple mutations in heterogeneous miltefosine-resistant *Leishmania major* population as determined by whole genome sequencing', *PLoS Neglected Tropical Diseases*, 6(2). doi: 10.1371/journal.pntd.0001512.

Coler, R. N. *et al.* (2015) 'From mouse to man: safety, immunogenicity and efficacy of a candidate leishmaniasis vaccine LEISH-F3+GLA-SE', *Clinical and Translational Immunology*. Nature Publishing Group, 4(4), pp. e35-13. doi: 10.1038/cti.2015.6.

Copeland, N. K. and Aronson, N. E. (2015) 'Leishmaniasis: Treatment updates and clinical practice guidelines review', *Current Opinion in Infectious Diseases*, 28(5), pp. 426–437. doi: 10.1097/QCO.0000000000000194.

Costa, F. C. *et al.* (2018) 'Expanding the toolbox for *Trypanosoma cruzi*: A parasite line incorporating a bioluminescence-fluorescence dual reporter and streamlined CRISPR/Cas9

functionality for rapid in vivo localisation and phenotyping', *PLoS Neglected Tropical Diseases*, 12(4), pp. 1–21. doi: 10.1371/journal.pntd.0006388.

da Costa, K. S. *et al.* (2017) 'Structural and evolutionary analyses of *Leishmania* Alba proteins', *Molecular and Biochemical Parasitology*. Elsevier, 217(August), pp. 23–31. doi: 10.1016/j.molbiopara.2017.08.006.

da Costa Lima, T. D. *et al.* (2010) 'Functional characterization of three *Leishmania* poly(A) binding protein homologues with distinct binding properties to RNA and protein partners', *Eukaryotic Cell*, 9(10), pp. 1484–1494. doi: 10.1128/EC.00148-10.

Coulson, R. M. R. and Smith, D. F. (1990) 'Isolation of genes showing increased or unique expression in the infective promastigotes of *Leishmania major*', *Molecular and Biochemical Parasitology*, 40(1), pp. 63–75. doi: 10.1016/0166-6851(90)90080-6.

Courret, N. *et al.* (2002) 'Biogenesis of *Leishmania*-harbouring parasitophorus vacuoles following phagocytosis of the metacyclic promastigote or amastigote stages of the parasites', *Journal of Cell Science*, 115(11), pp. 2303–2316. doi: 10.1242/jcs.115.11.2303.

Cox, F. E. G. E. G. (2002) 'History of human parasitology', *Clinical Microbiology Reviews*, 15(4), pp. 595–612. doi: 10.1128/CMR.15.4.595-612.2002.

Croan, D. G., Morrison, D. A. and Ellis, J. T. (1997) 'Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences', *Molecular and Biochemical Parasitology*, 89(2), pp. 149–159. doi: 10.1016/S0166-6851(97)00111-4.

Cuervo, P. *et al.* (2007) 'Proteome analysis of *Leishmania (Viannia) braziliensis* by two-dimensional gel electrophoresis and mass spectrometry', *Molecular and Biochemical Parasitology*, 154(1), pp. 6–21. doi: 10.1016/j.molbiopara.2007.03.013.

Curotto de Lafaille, M. A., Laban, A. and Wirth, D. F. (1992) 'Gene expression in *Leishmania*: Analysis of essential 5' DNA sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 89(7), pp. 2703–2707. doi: 10.1073/pnas.89.7.2703.

Dallagiovanna, B. *et al.* (2005) '*Trypanosoma cruzi*: Molecular characterization of TcPUF6, a Pumilio protein', *Experimental Parasitology*, 109(4), pp. 260–264. doi: 10.1016/j.exppara.2005.01.003.

Dallagiovanna, B. *et al.* (2008) 'Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*', *Journal of Biological Chemistry*. Â© 2008 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 283(13), pp. 8266–8273. doi: 10.1074/jbc.M703097200.

Damasceno, J. D. *et al.* (2020) 'Genome duplication in *Leishmania major* relies on persistent subtelomeric dna replication', *eLife*, 9, pp. 1–24. doi: 10.7554/ELIFE.58030.

Damianou, A. *et al.* (2020) 'Essential roles for deubiquitination in *Leishmania* life cycle progression', *PLoS Pathogens*, 16(6). doi: 10.1371/journal.ppat.1008455.

Daneshbod, Y. *et al.* (2011) 'Clinical, histopathologic, and cytologic diagnosis of mucosal leishmaniasis and literature review', *Archives of Pathology & Laboratory Medicine*, 135(4), pp. 478–482. doi: 10.5858/2010-0069-OA.1.

Das, A. *et al.* (2015) 'High throughput sequencing analysis of *Trypanosoma brucei* DRBD3/PTB1-bound mRNAs', *Molecular and Biochemical Parasitology*. Elsevier B.V., 199(1–2), pp. 1–4. doi: 10.1016/j.molbiopara.2015.02.003.

Dassoni, F. (2017) 'Cutaneous leishmaniasis and its clinical spectrum. An overview from the Old World', *Clinical Dermatology*, 5(1), p. 12. doi: 10.11138/cderm/2017.5.1.012.

Datta, A. K., Datta, R. and Sen, B. (2008) 'Antiparasitic chemotherapy: Tinkering with the purine salvage pathway', *Advances in Experimental Medicine and Biology*, 625, pp. 116–132. doi: 10.1007/978-0-387-77570-8_10.

David, C. V. and Craft, N. (2009) 'Cutaneous and Mucocutaneous Leishmaniasis', *Dermatologic Therapy*, 22, pp. 491–502. doi: 10.1111/j.1529-8019.2009.01272.x.

Dean, S. *et al.* (2015) 'A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids', *Open Biology*, 5(1). doi: 10.1098/rsob.140197.

Dean, S., Sunter, J. D. and Wheeler, R. J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource', *Trends in Parasitology*. Elsevier Ltd, 33(2), pp. 80–82. doi: 10.1016/j.pt.2016.10.009.

Depledge, D. P. *et al.* (2010) '*Leishmania*-specific surface antigens show sub-genus sequence variation and immune recognition', *PLoS Neglected Tropical Diseases*, 4(9). doi: 10.1371/journal.pntd.0000829.

Desjeux, P. *et al.* (2013) 'Report of the post Kala-Azar dermal leishmaniasis (PKDL) consortium meeting, New Delhi, India, 27-29 June 2012', *Parasites and Vectors*. Parasites & Vectors, 6(1), p. 1. doi: 10.1186/1756-3305-6-196.

Dever, T. E. (2002) 'Gene-specific regulation by general translation factors', *Cell*, 108(4), pp. 545–556. doi: 10.1016/S0092-8674(02)00642-6.

Diana, M. and Inchaustegui, P. (2015) 'Purification of specific mRNP via the nascent polypeptide The RNA Binding Proteins ZC3H22 and ZC3H38', (June).

Dias-Lopes, G. *et al.* (2021) 'Axenic amastigotes of *Leishmania* species as a suitable model for in vitro studies', *Acta Tropica*, 220(January). doi: 10.1016/j.actatropica.2021.105956.

Dillon, L. A. L. *et al.* (2015) 'Transcriptomic profiling of gene expression and RNA processing during *Leishmania major* differentiation', *Nucleic Acids Research*, 43(14), pp. 6799–6813. doi:

10.1093/nar/gkv656.

Diniz, J. A. *et al.* (2021) 'Protein methyltransferase 7 deficiency in *Leishmania major* increases neutrophil associated pathology in murine model', *PLoS Neglected Tropical Diseases*, 15(3), pp. 1–20. doi: 10.1371/journal.pntd.0009230.

Diro, E. *et al.* (2015) 'Atypical manifestations of visceral leishmaniasis in patients with HIV in north Ethiopia: A gap in guidelines for the management of opportunistic infections in resource poor settings', *The Lancet Infectious Diseases*, 15(1), pp. 122–129. doi: 10.1016/S1473-3099(14)70833-3.

DNDi (2021a) *CpG-D35 for cutaneous leishmaniasis*. Available at: https://dndi.org/research-development/portfolio/cpg-d35/ (Accessed: 30 December 2021).

DNDi (2021b) *DNDi Research Development Portfolio*. Available at: https://dndi.org/research-development/portfolio/ (Accessed: 30 December 2021).

DNDi (2021c) *DNDi website*. Available at: https://dndi.org/diseases/visceral-leishmaniasis/ (Accessed: 16 December 2021).

DNDi (2021d) *New CL combination therapies*. Available at: https://dndi.org/research-development/portfolio/new-cl-combos/ (Accessed: 30 December 2021).

Doehl, J. S. P. *et al.* (2017) '*Leishmania* HASP and SHERP genes are required for in vivo differentiation, parasite transmission and virulence attenuation in the host', *PLoS Pathogens*, 13(1), pp. 1–35. doi: 10.1371/journal.ppat.1006130.

Dostálová, A. and Volf, P. (2012) '*Leishmania* development in sand flies: Parasite-vector interactions overview', *Parasites and Vectors*, 5(1), pp. 1–12. doi: 10.1186/1756-3305-5-276.

Droll, D. *et al.* (2010) 'The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation', *FEBS Letters*. Federation of European Biochemical Societies, 584(6), pp. 1156–1162. doi: 10.1016/j.febslet.2010.02.018.

Droll, D. *et al.* (2013) 'Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein', *PLoS Pathogens*, 9(4). doi: 10.1371/journal.ppat.1003286.

Duncan, S. M. *et al.* (2016) 'Conditional gene deletion with DiCre demonstrates an essential role for CRK3 in *Leishmania mexicana* cell cycle regulation', 100(April), pp. 931–944. doi: 10.1111/mmi.13375.

Duncan, S. M., Jones, N. G. and Mottram, J. C. (2017) 'Recent advances in *Leishmania* reverse genetics: Manipulating a manipulative parasite', *Molecular and Biochemical Parasitology*. Elsevier, 216(June), pp. 30–38. doi: 10.1016/j.molbiopara.2017.06.005.

El-Sayed, N. M. *et al.* (2005) 'Comparative genomics of trypanosomatid parasitic protozoa.', *Science (New York, N.Y.)*, 309(5733), pp. 404–409. doi: 10.1126/science.1112181.

Eldin, C. *et al.* (2021) '"Chiclero's Ulcer" Due to *Leishmania mexicana* in Travelers Returning from Central America: A Case Report and Review of the Literature', *Pathogens*, 10(9). doi: 10.3390/pathogens10091112.

Erben, E. *et al.* (2021) 'Insights into the functions and RNA binding of *Trypanosoma brucei* ZC3H22, RBP9 and DRBD7', *Parasitology*, 148(10), pp. 1186–1195. doi: 10.1017/S0031182021000123.

Erben, E., Chakraborty, C. and Clayton, C. (2013) 'The CAF1-NOT complex of trypanosomes', *Frontiers in Genetics*, 4(JAN), pp. 2–5. doi: 10.3389/fgene.2013.00299.

Erben, E. D. *et al.* (2014) 'A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in *Trypanosoma brucei*', *PLoS Pathogens*, 10(6). doi: 10.1371/journal.ppat.1004178.

Erdmann, M. *et al.* (2006) 'Interacting protein kinases involved in the regulation of flagellar length', *Molecular Biology of the Cell*, 17(4), pp. 2035–2045. doi: 10.1091/mbc.e05-10-0976.

Espada, C. R. *et al.* (2021) 'In vitro susceptibility to miltefosine of *Leishmania infantum* (Syn. *l. chagasi*) isolates from different geographical areas in Brazil', *Microorganisms*, 9(6). doi: 10.3390/microorganisms9061228.

Estévez, A. M. (2008) 'The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes', *Nucleic Acids Research*, 36(14), pp. 4573–4586. doi: 10.1093/nar/gkn406.

Eswaran, J. and Knapp, S. (2010) 'Insights into protein kinase regulation and inhibition by large scale structural comparison', *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., 1804(3), pp. 429–432. doi: 10.1016/j.bbapap.2009.10.013.

Eva Kyriacou (2021) *Preliminary Analysis of CRISPR-Cas9 Knockout of RNA Binding Proteins in Leishmania mexicana*. University of York.

El Fakhry, Y., Ouellette, M. and Papadopoulou, B. (2002) 'A proteomic approach to identify developmentally regulated proteins in *Leishmania infantum*', *Proteomics*, 2(8), pp. 1007–1017. doi: 10.1002/1615-9861(200208)2:8<1007::AID-PROT1007>3.0.CO;2-G.

Fei, T. *et al.* (2017) 'Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing', *Proceedings of the National Academy of Sciences of the United States of America*, 114(26), pp. E5207–E5215. doi: 10.1073/pnas.1617467114.

Fernández-Moya, S. M. *et al.* (2012) 'Alterations in DRBD3 ribonucleoprotein complexes in response to stress in *Trypanosoma brucei*', *PLoS ONE*. Edited by Z. Li, 7(11), p. e48870. doi: 10.1371/journal.pone.0048870.

Ferreira, T. R. *et al.* (2014) 'Altered expression of an RBP-associated arginine methyltransferase 7

in *Leishmania major* affects parasite infection', 94(October), pp. 1085–1102. doi: 10.1111/mmi.12819.

Ferreira, T. R. *et al.* (2020) 'PRMT7 regulates RNA-binding capacity and protein stability in *Leishmania* parasites', *Nucleic Acids Research*. Oxford University Press, pp. 1–16. doi: 10.1093/nar/gkaa211.

Fiebig, M. *et al.* (2014) 'SLaP mapper: A webserver for identifying and quantifying spliced-leader addition and polyadenylation site usage in kinetoplastid genomes', *Molecular and Biochemical Parasitology*. Elsevier B.V., 196(2), pp. 71–74. doi: 10.1016/j.molbiopara.2014.07.012.

Field, J. *et al.* (1988) 'Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method', *Molecular and Cellular Biology*, 8(5), pp. 2159–2165. doi: 10.1128/mcb.8.5.2159-2165.1988.

Field, M. C. *et al.* (2004) 'New approaches to the microscopic imaging of *Trypanosoma brucei*', *Microscopy and Microanalysis*, 10(5), pp. 621–636. doi: 10.1017/S1431927604040942.

Findlay, R. C. *et al.* (2021) 'High-speed, three-dimensional imaging reveals chemotactic behaviour specific to human-infective *Leishmania* parasites', *eLife*, 10, pp. 1–11. doi: 10.7554/eLife.65051.

Finn, R. D. *et al.* (2017) 'InterPro in 2017-beyond protein family and domain annotations', *Nucleic Acids Research*, 45(D1), pp. D190–D199. doi: 10.1093/nar/gkw1107.

Fisk, J. C. *et al.* (2008) 'TbRGG2, an essential RNA editing accessory factor in two *Trypanosoma brucei* life cycle stages', *Journal of Biological Chemistry*. Â© 2008 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 283(34), pp. 23016–23025. doi: 10.1074/jbc.M801021200.

Fisk, J. C. *et al.* (2009) 'Distinct and overlapping functions of MRP1/2 and RBP16 in mitochondrial RNA metabolism.', *Molecular and cellular biology*, 29(19), pp. 5214–5225. doi: 10.1128/MCB.00520-09.

Fleck, K., Nitz, M. and Jeffers, V. (2021) '"Reading" a new chapter in protozoan parasite transcriptional regulation', *PLoS Pathogens*, 17(12), pp. 1–21. doi: 10.1371/journal.ppat.1010056.

Flinn, H. M. and Smith, D. F. (1992) 'Genomic organisation and expression of a differentially-regulated gene family from *Leishmania major*', *Nucleic Acids Research*, 20(4), pp. 755–762. doi: 10.1093/nar/20.4.755.

Flora, R. *et al.* (2014) 'Immunosuppression during *Leishmania donovani* infection: a potential target for the development of therapy', *Annals of parasitology*, 60(4), pp. 239–245.

Folgueira, C., Martínez-Bonet, M. and Requena, J. M. (2010) 'The *Leishmania infantum* PUF proteins are targets of the humoral response during visceral leishmaniasis.', *BMC research notes*, 3, p. 13. doi: 10.1186/1756-0500-3-13.

Foucher, A. L., Papadopoulou, B. and Ouellette, M. (2006) 'Prefractionation by digitonin extraction increases representation of the cytosolic and intracellular proteome of *Leishmania infantum*', *Journal of Proteome Research*, 5(7), pp. 1741–1750. doi: 10.1021/pr060081j.

Freedman, D. J. and Beverley, S. M. (1993) 'Two more independent selectable markers for stable transfection of *Leishmania*', *Molecular and Biochemical Parasitology*, 62(1), pp. 37–44. doi: 10.1016/0166-6851(93)90175-W.

Freire, E. R. *et al.* (2014) 'eIF4F-like complexes formed by cap-binding homolog TbEIF4E5 with TbEIF4G1 or TbEIF4G2 are implicated in post-transcriptional regulation in *Trypanosoma brucei*', *Rna*, 20(8), pp. 1272–1286. doi: 10.1261/rna.045534.114.

Ganguly, S. *et al.* (2015) 'PKDL—A silent parasite pool for transmission of leishmaniasis in Kala-azar endemic areas of Malda district, West Bengal, India', *PLoS Neglected Tropical Diseases*, 9(10), pp. 1–13. doi: 10.1371/journal.pntd.0004138.

Garami, A. and Ilg, T. (2001) 'The role of phosphomannose isomerase in *Leishmania mexicana* glycoconjugate synthesis and virulence', *Journal of Biological Chemistry*. Â© 2001 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 276(9), pp. 6566–6575. doi: 10.1074/jbc.M009226200.

Garapati, P. *et al.* (2018) 'Knowledge, stigma, health seeking behaviour and its determinants among patients with post kalaazar dermal leishmaniasis, Bihar, India', *PLoS ONE*, 13(9), pp. 1–13. doi: 10.1371/journal.pone.0203407.

Garvey, E. P. and Santi, D. V. (1986) 'Stable amplified DNA in drug-resistant *Leishmania* exists as extrachromosomal circles', *Science*, 233(4763), pp. 535–540. doi: 10.1126/science.3726545.

Gay, L. S., Wilson, M. E. and Donelson, J. E. (1996) 'The promoter for the ribosomal RNA genes of *Leishmania chagasi*', *Molecular and Biochemical Parasitology*, 77(2), pp. 193–200. doi: 10.1016/0166-6851(96)02594-7.

Genest, P. A. *et al.* (2007) 'Telomeric localization of the modified DNA base J in the genome of the protozoan parasite *Leishmania*', *Nucleic Acids Research*, 35(7), pp. 2116–2124. doi: 10.1093/nar/gkm050.

Genest, P. A. *et al.* (2015) 'Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing', *Nucleic Acids Research*, 43(4), pp. 2102–2115. doi: 10.1093/nar/gkv095.

Geoghegan, V. *et al.* (2021) 'Protein kinase signalling at the *Leishmania* kinetochore captured by XL-BioID', *BioXiv*. doi: 10.1101/2021.07.08.451598.

Gerstberger, S., Hafner, M. and Tuschl, T. (2014) 'A census of human RNA-binding proteins', *Nature Reviews Genetics*. Nature Publishing Group, 15(12), pp. 829–845. doi: 10.1038/nrg3813.

Ghorbani, M. and Farhoudi, R. (2018) 'Leishmaniasis in humans: Drug or vaccine therapy?', *Drug Design, Development and Therapy*, 12, pp. 25–40. doi: 10.2147/DDDT.S146521.

Ghosh, P. and Sowdhamini, R. (2016) 'Genome-wide survey of putative RNA-binding proteins encoded in the human proteome', *Mol. BioSyst.* Royal Society of Chemistry, 12(2), pp. 532–540. doi: 10.1039/C5MB00638D.

Gilbert, I. H. (2002) 'Inhibitors of dihydrofolate reductase in *Leishmania* and trypanosomes', *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1587(2–3), pp. 249–257. doi: 10.1016/S0925-4439(02)00088-1.

Giraud, E. *et al.* (2019) 'Quantifying *Leishmania* metacyclic promastigotes from individual sandfly bites reveals the efficiency of vector transmission', *Communications Biology*. Springer US, 2(1), pp. 25–28. doi: 10.1038/s42003-019-0323-8.

Giudice, G. *et al.* (2016) 'ATtRACT-a database of RNA-binding proteins and associated motifs', *Database*, 2016(2), pp. 1–9. doi: 10.1093/database/baw035.

Glisovic, T. *et al.* (2008) 'RNA-binding proteins and post-transcriptional gene regulation', *FEBS Letters*, 582(14), pp. 1977–1986. doi: 10.1016/j.febslet.2008.03.004.

Góngora, R. *et al.* (2003) 'Mapping the proteome of *Leishmania Viannia* parasites using two-dimensional polyacrylamide gel electrophoresis and associated technologies.', *Biomédica : revista del Instituto Nacional de Salud*, 23(2), pp. 153–160. doi: 10.7705/biomedica.v23i2.1207.

Goodwin, L. G. (1945) 'The chemotherapy of experimental leishmaniasis II. A dose response curve for the activity of sodium stibogluconate', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 39(2), pp. 133–145. doi: 10.1016/0035-9203(45)90004-6.

Goos, C. *et al.* (2019) 'Trypanosomes can initiate nuclear export co-transcriptionally', *Nucleic Acids Research*. Oxford University Press, 47(1), pp. 266–282. doi: 10.1093/nar/gky1136.

Gossage, S. M., Rogers, M. E. and Bates, P. A. (2003) 'Two separate growth phases during the development of *Leishmania* in sand flies: Implications for understanding the life cycle', *International Journal for Parasitology*, 33(10), pp. 1027–1034. doi: 10.1016/S0020-7519(03)00142-5.

Goto, Y. *et al.* (2006) 'Cloning, characterization, and serodiagnostic evaluation of *Leishmania infantum* tandem repeat proteins', *Infection and Immunity*, 74(7), pp. 3939–3945. doi: 10.1128/IAI.00101-06.

Goto, Y., Coler, R. N. and Reed, S. G. (2007) 'Bioinformatic identification of tandem repeat antigens of the *Leishmania* donovani complex', *Infection and Immunity*, 75(2), pp. 846–851. doi: 10.1128/IAI.01205-06.

Goyard, S. and Beverley, S. M. (2000) 'Blasticidin resistance: A new independent marker for

stable transfection of *Leishmania*', *Molecular and Biochemical Parasitology*, 108(2), pp. 249–252. doi: 10.1016/S0166-6851(00)00210-3.

Grace, T. D. C. (1962) 'Establishment of four strains of cells from insect tissues grown in vitro', *Nature*, 195(4843), pp. 788–789. doi: 10.1038/195788a0.

Grant, K. M. *et al.* (2004) 'Inhibitors of *Leishmania mexicana* CRK3 cyclin-dependent kinase: chemical library screen and antileishmanial activity', *Antimicrobial Agents and Chemotherapy*, 48(8), pp. 3033–3042. doi: 10.1128/AAC.48.8.3033-3042.2004.

Gray, W. H. and Trevan, J. W. (1931) 'Experiments on antimony compounds used in the treatment of bilharzia disease and kala-azar', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 25(3), pp. 147–152. doi: 10.1016/S0035-9203(31)90049-7.

Grondin, K., Roy, G. and Ouellette, M. (1996) 'Formation of extrachromosomal circular amplicons with direct or inverted duplications in drug-resistant *Leishmania tarentolae*', *Molecular and Cellular Biology*, 16(7), pp. 3587–3595. doi: 10.1128/mcb.16.7.3587.

Guerra, N. *et al.* (2011) 'Identification and functional characterization of a poly(A)-binding protein from *Leishmania infantum* (LiPABP)', *FEBS Letters*. Federation of European Biochemical Societies, 585(1), pp. 193–198. doi: 10.1016/j.febslet.2010.11.042.

Gupta, S. K. *et al.* (2013) 'Basal splicing factors regulate the stability of mature mRNAs in trypanosomes', *Journal of Biological Chemistry*, 288(7), pp. 4991–5006. doi: 10.1074/jbc.M112.416578.

Gupta, S. K. *et al.* (2014) 'Two splicing factors carrying serine-arginine motifs, TSR1 and TSR1IP, regulate splicing, mRNA stability, and rRNA processing in *Trypanosoma brucei*', *RNA Biology*, 11(6), pp. 715–731. doi: 10.4161/rna.29143.

Guy, R. A. and Belosevic, M. (1993) 'Comparison of receptors required for entry of *Leishmania* major amastigotes into macrophages', *Infection and Immunity*, 61(4), pp. 1553–1558. doi: 10.1128/iai.61.4.1553-1558.1993.

Hajjaran, H. *et al.* (2013) 'Unusual presentation of disseminated cutaneous leishmaniasis due to *Leishmania major*: Case reports of four Iranian patients', *Asian Pacific Journal of Tropical Medicine*. Hainan Medical College, 6(4), pp. 333–336. doi: 10.1016/S1995-7645(13)60066-8.

Hajjaran, H. *et al.* (2018) 'Comparison of the proteome profiling of Iranian isolates of *Leishmania tropica*, *L. major* and *L. infantum* by two-dimensional electrophoresis (2-DE) and mass-spectrometry.', *Iranian journal of parasitology*, 10(4), pp. 530–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26811718.

Haldar, A. K., Sen, P. and Roy, S. (2011) 'Use of Antimony in the Treatment of Leishmaniasis: Current Status and Future Directions', *Molecular Biology International*, 2011, pp. 1–23. doi: 10.4061/2011/571242.

Hall, M. J. R. *et al.* (2021) 'Micro-CT visualization of a promastigote secretory gel (PSG) and parasite plug in the digestive tract of the sand fly *Lutzomyia longipalpis* infected with *Leishmania mexicana*', *PLoS Neglected Tropical Diseases*, 15(8), pp. 1–17. doi: 10.1371/journal.pntd.0009682.

Hall, T. M. T. (2005) 'Multiple modes of RNA recognition by zinc finger proteins', *Current Opinion in Structural Biology*, 15(3 SPEC. ISS.), pp. 367–373. doi: 10.1016/j.sbi.2005.04.004.

Handler, M. Z. *et al.* (2015) 'Cutaneous and mucocutaneous leishmaniasis: Clinical perspectives', *Journal of the American Academy of Dermatology*. Elsevier Inc, 73(6), pp. 897–908. doi: 10.1016/j.jaad.2014.08.051.

Harmer, J. *et al.* (2018) 'Farming, slaving and enslavement: histories of endosymbioses during kinetoplastid evolution', *Parasitology*, 145(10), pp. 1311–1323. doi: 10.1017/s0031182018000781.

Harris, M. T., Mitchell, W. G. and Morris, J. C. (2014) 'Targeting protozoan parasite metabolism: glycolytic enzymes in the therapeutic crosshairs', *Current Medicinal Chemistry*, 21(15), pp. 1668–1678. doi: 10.2174/09298673113206660286.

Hashiguchi, Y. *et al.* (2016) 'Diffuse and disseminated cutaneous leishmaniasis: Clinical cases experienced in Ecuador and a brief review', *Tropical Medicine and Health*. Tropical Medicine and Health, 44(1), pp. 1–9. doi: 10.1186/s41182-016-0002-0.

Hayman, M. L. and Read, L. K. (1999) '*Trypanosoma brucei* RBP16 Is a Mitochondrial Y-box Family Protein with Guide RNA Binding Activity', *Journal of Biological Chemistry*, 274(17), pp. 12067–12074. doi: 10.1074/jbc.274.17.12067.

Hendriks, E. F. *et al.* (2001) 'A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form', *EMBO Journal*, 20(23), pp. 6700–6711. doi: 10.1093/emboj/20.23.6700.

Hendriks, E. F. and Matthews, K. R. (2005) 'Disruption of the developmental programme of *Trypanosoma brucei* by genetic ablation of TbZFP1, a differentiation-enriched CCCH protein', *Molecular Microbiology*, 57(3), pp. 706–716. doi: 10.1111/j.1365-2958.2005.04679.x.

Henn, G. A. de L. *et al.* (2018) 'Is Visceral Leishmaniasis the same in HIV-coinfected adults?', *Brazilian Journal of Infectious Diseases*, 22(2), pp. 92–98. doi: 10.1016/j.bjid.2018.03.001.

Hentze, M. W. *et al.* (2018) 'A brave new world of RNA-binding proteins', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. doi: 10.1038/nrm.2017.130.

Hitchcock, R. A. *et al.* (2007) 'The promoter and transcribed regions of the *Leishmania tarentolae* spliced leader RNA gene array are devoid of nucleosomes', *BMC Microbiology*, 7, pp. 1–14. doi: 10.1186/1471-2180-7-44.

Hoek, M., Zanders, T. and Cross, G. A. M. (2002) '*Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a *Pumilio* family protein', *Molecular and biochemical*

parasitology, 120(2), pp. 269–283. doi: 10.1016/S0166-6851(02)00009-9.

Iantorno, S. A. *et al.* (2017) 'Gene expression in *Leishmania* is regulated predominantly by gene dosage', *mBio*, 8(5). doi: 10.1128/mBio.01393-17.

Inbar, E. *et al.* (2017) 'The transcriptome of *Leishmania major* developmental stages in their natural sand fly vector', *mBio*. doi: 10.1128/mBio.00029-17.

Ismail, A. *et al.* (1999) 'Immunopathology of post kala-azar dermal leishmaniasis (PKDL): T-cell phenotypes and cytokine profile', *The Journal of Pathology*, 189(4), pp. 615–622. doi: 10.1002/(SICI)1096-9896(199912)189:4<615::AID-PATH466>3.0.CO;2-Z.

Ivens, A. C. *et al.* (2005) 'The genome of the kinetoplastid parasite, *Leishmania major*', *Science*, 309(5733), pp. 436–442. doi: 10.1126/science.1112680.

Jackson, A. P. (2010) 'The evolution of amastin surface glycoproteins in trypanosomatid parasites', *Molecular Biology and Evolution*, 27(1), pp. 33–45. doi: 10.1093/molbev/msp214.

Jansson, L. and Hedman, J. (2019) 'Challenging the proposed causes of the PCR plateau phase', *Biomolecular Detection and Quantification*. Elsevier, 17(February), p. 100082. doi: 10.1016/j.bdq.2019.100082.

Jensen, B. C. *et al.* (2021) 'Chromatin-associated protein complexes link DNA base J and transcription termination in *Leishmania*', *mSphere*, 6(1), pp. 1–18. doi: 10.1128/msphere.01204-20.

Jha, B. A. *et al.* (2014) 'Depletion of the trypanosome pumilio domain protein PUF2 or of some other essential proteins causes transcriptome changes related to coding region length', *Eukaryotic Cell*, 13(5), pp. 664–674. doi: 10.1128/EC.00018-14.

Jha, B. A. *et al.* (2015) 'The DRBD13 RNA binding protein is involved in the insect-stage differentiation process of *Trypanosoma brucei*', *FEBS Letters*, 589(15), pp. 1966–1974. doi: 10.1016/j.febslet.2015.05.036.

Jha, T. K. *et al.* (1999) 'Miltefosine, an oral agent, for the treatment of Indian visceral Leishmaniasis', *New England Journal of Medicine*, 341(24), pp. 1795–1800. doi: 10.1056/NEJM199912093412403.

Jha, T. K. (2006) 'Drug unresponsiveness & combination therapy for kala-azar', *Indian Journal of Medical Research*, 123(3), pp. 389–398.

Johnson, P. J., Kooter, J. M. and Borst, P. (1987) 'Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene', *Cell*, 51(2), pp. 273–281. doi: 10.1016/0092-8674(87)90154-1.

Jones, N. G. *et al.* (2018) 'Genetically validated drug targets in *Leishmania*: current knowledge and future prospects', *ACS Infectious Diseases*, 4(4), pp. 467–477. doi: 10.1021/acsinfecdis.7b00244.

Jones, N. G. *et al.* (2021) 'Bromodomain factor 5 is an essential transcriptional regulator of the *Leishmania* genome', *bioRxiv*, p. 2021.09.29.462384.

Jones, P. *et al.* (2014) 'InterProScan 5: Genome-scale protein function classification', *Bioinformatics*, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031.

Joshi, P. B. *et al.* (1995) 'The gene encoding streptothricin acetyltransferase (sat) as a selectable marker for *Leishmania* expression vectors', *Gene*, 156(1), pp. 145–149. doi: 10.1016/0378-1119(95)00042-5.

Joung, J. *et al.* (2017) 'Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening', *Nature Protocols*. Nature Publishing Group, 12(4), pp. 828–863. doi: 10.1038/nprot.2017.016.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*. Springer US, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Kamba, P. F. *et al.* (2018) *The 27 kDa Trypanosoma brucei pentatricopeptide repeat protein is a G-tract specific RNA binding protein*, *Scientific Reports*. doi: 10.1038/s41598-018-34377-9.

Kaufer, A. *et al.* (2017) 'The evolution of trypanosomatid taxonomy', *Parasites and Vectors*. Parasites & Vectors, 10(1), pp. 1–17. doi: 10.1186/s13071-017-2204-7.

Kaye, P. and Scott, P. (2011) 'Leishmaniasis: complexity at the host–pathogen interface', *Nature Reviews Microbiology*. Nature Publishing Group, 9(8), pp. 604–615. doi: 10.1038/nrmicro2608.

Kazerooni, P. A. *et al.* (2018) 'Under-ascertainment, under-reporting and timeliness of Iranian communicable disease surveillance system for zoonotic diseases', *Public Health*. Elsevier Ltd, 154, pp. 130–135. doi: 10.1016/j.puhe.2017.10.029.

Kellner, M. *et al.* (2015) 'DEAD-box helicase DDX27 regulates 3' end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex', *Experimental Cell Research*. Elsevier, 334(1), pp. 146–159. doi: 10.1016/j.yexcr.2015.03.017.

Khare, S. *et al.* (2016) 'Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness', *Nature*. Nature Publishing Group, 537(7619), pp. 229–233. doi: 10.1038/nature19339.

Kieny, M.-P. and Pécoul, B. (2020) 'DNDi annual report 2020'.

Killick-Kendrick, R. (1990) 'Phlebotomine vectors of the leishmaniases: a review', *Medical and Veterinary Entomology*, 4(1), pp. 1–24. doi: 10.1111/j.1365-2915.1990.tb00255.x.

Killick-Kendrick, R. (1999) 'The biology and control of Phlebotomine sand flies', *Clinics in Dermatology*, 17(3), pp. 279–289. doi: 10.1016/S0738-081X(99)00046-2.

Killick Kendrick, R., Molyneux, D. H. and Ashford, R. W. (1974) 'Leishmania in phlebotomid

sandflies. I. Modifications of the flagellum associated with attachment to the mid gut and oesophageal valve of the sandfly', *Proceedings of the Royal Society of London - Biological Sciences*, 187(1089), pp. 409–419. doi: 10.1098/rspb.1974.0085.

Kima, P. E. *et al.* (2000) 'Internalization of *Leishmania mexicana* complex amastigotes via the Fc receptor is required to sustain infection in murine cutaneous leishmaniasis', *Journal of Experimental Medicine*, 191(6), pp. 1063–1067. doi: 10.1084/jem.191.6.1063.

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), pp. 1087–1096. doi: 10.1016/j.ijpara.2007.04.007.

Kimutai, R. *et al.* (2017) 'Safety and effectiveness of sodium stibogluconate and paromomycin combination for the treatment of visceral leishmaniasis in eastern Africa: Results from a pharmacovigilance programme', *Clinical Drug Investigation*. Springer International Publishing, 37(3), pp. 259–272. doi: 10.1007/s40261-016-0481-0.

Klein, C. *et al.* (2015) 'Polysomes of *Trypanosoma brucei*: Association with initiation factors and RNA-binding proteins', *PLoS ONE*, 10(8), pp. 1–12. doi: 10.1371/journal.pone.0135973.

Kloehn, J. *et al.* (2015) 'Characterization of metabolically quiescent *Leishmania* parasites in murine lesions using heavy water labeling', *PLoS Pathogens*, 11(2), pp. 1–19. doi: 10.1371/journal.ppat.1004683.

Knight, J. D. R. *et al.* (2007) 'Conservation, variability and the modeling of active protein kinases', *PLoS ONE*, 2(10). doi: 10.1371/journal.pone.0000982.

Koike-Yusa, H. *et al.* (2014) 'Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library', *Nature Biotechnology*, 32(3), pp. 267–273. doi: 10.1038/nbt.2800.

Kolev, N. G. *et al.* (2012) 'Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein', *Science*, 338(6112), pp. 1352–1353. doi: 10.1126/science.1229641.

Kolev, N. G., Ullu, E. and Tschudi, C. (2014) 'The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*', *Cellular Microbiology*, 16(4), pp. 482–489. doi: 10.1111/cmi.12268.

König, J. *et al.* (2010) 'iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution', *Nature Structural & Molecular Biology*, 17(7), pp. 909–915. doi: 10.1038/nsmb.1838.

Koushika, S. P., Lisbin, M. J. and White, K. (1996) 'ELAV, a *Drosophila* neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform', *Current Biology*, 6(12), pp. 1634–1641. doi: 10.1016/S0960-9822(02)70787-2.

Kramer, S. and Carrington, M. (2011) ‘Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids’, *Trends in Parasitology*. Elsevier Ltd, 27(1), pp. 23–30. doi: 10.1016/j.pt.2010.06.011.

Kramer, S., Kimblin, N. C. and Carrington, M. (2010) ‘Genome-wide in silico screen for CCCH-type zinc finger proteins of *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*.’, *BMC genomics*, 11, p. 283. doi: 10.1186/1471-2164-11-283.

Kuhls, K. *et al.* (2011) ‘Comparative microsatellite typing of new world *Leishmania infantum* reveals low heterogeneity among populations and its recent old world origin’, *PLoS Neglected Tropical Diseases*, 5(6), pp. 1–16. doi: 10.1371/journal.pntd.0001155.

Kulshrestha, A. *et al.* (2014) ‘Comparative transcript expression analysis of miltefosine-sensitive and miltefosine-resistant *Leishmania donovani*’, *Parasitology Research*, 113(3), pp. 1171–1184. doi: 10.1007/s00436-014-3755-6.

Kumar, G. A. *et al.* (2019) ‘*Leishmania donovani* internalizes into host cells via caveolin-mediated endocytosis’, *Scientific Reports*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41598-019-49007-1.

Kumar, S., Stecher, G. and Tamura, K. (2016) ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets’, *Molecular biology and evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Kwon, S. C. *et al.* (2013) ‘The RNA-binding protein repertoire of embryonic stem cells’, *Nature Structural and Molecular Biology*. Nature Publishing Group, 20(9), pp. 1122–1130. doi: 10.1038/nsmb.2638.

Laemmli, U. K. (1970) ‘Cleavage of structural proteins during the assembly of the head of bacteriophage T4’, *Nature*, 227(5259), pp. 680–685. doi: 10.1038/227680a0.

Lai, G. N. *et al.* (2008) ‘Migratory dermal dendritic cells act as rapid sensors of protozoan parasites’, *PLoS Pathogens*, 4(11). doi: 10.1371/journal.ppat.1000222.

Lainson, R. and Shaw, J. J. (1987) ‘Evolution, classification and geographical distribution of *Leishmania*’, *Journal of Tropical Medicine and Hygine*, 59, pp. 421–425.

LeBowitz, J. H. *et al.* (1993) ‘Coupling of poly(A) site selection and trans-splicing in *Leishmania*.’, *Genes & Development*, 7(6), pp. 996–1007. doi: 10.1101/gad.7.6.996.

Lee, Y. and Rio, D. C. (2015) ‘Mechanisms and regulation of alternative Pre-mRNA splicing’, *Annual Review of Biochemistry*, 84, pp. 291–323. doi: 10.1146/annurev-biochem-060614-034316.

Levy, G. V. *et al.* (2015) ‘Depletion of the SR-related protein TbRRM1 leads to cell cycle arrest and apoptosis-like death in *Trypanosoma brucei*’, *PLoS ONE*, 10(8), pp. 1–20. doi: 10.1371/journal.pone.0136070.

Li, W. *et al.* (2012) ‘PSI-Search: Iterative HOE-reduced profile SSEARCH searching’,

Bioinformatics, 28(12), pp. 1650–1651. doi: 10.1093/bioinformatics/bts240.

Liang, X. *et al.* (2003) 'Trans and cis-splicing in Trypanosomatids: Mechanism, factors, and regulation', *Eukaryotic Cell*, 2(5), pp. 830–840. doi: 10.1128/EC.2.5.830-840.2003.

de Lima, G. B. *et al.* (2021) 'The translation initiation factor EIF4E5 from Leishmania: crystal structure and interacting partners', *RNA Biology*. Taylor & Francis, 00(00), pp. 1–17. doi: 10.1080/15476286.2021.1918919.

Lindoso, J. A. L. *et al.* (2016) 'Leishmaniasis–HIV coinfection: Current challenges', *HIV/AIDS - Research and Palliative Care*, 8, pp. 147–156. doi: 10.2147/HIV.S93789.

Ling, A. S., Trotter, J. R. and Hendriks, E. F. (2011) 'A zinc finger protein, TbZC3H20, stabilizes two developmentally regulated mRNAs in trypanosomes', *Journal of Biological Chemistry*, 286(23), pp. 20152–20162. doi: 10.1074/jbc.M110.139261.

Lira, C. B. B. *et al.* (2007) 'LaRbp38: A *Leishmania amazonensis* protein that binds nuclear and kinetoplast DNAs', *Biochemical and Biophysical Research Communications*, 358(3), pp. 854–860. doi: 10.1016/j.bbrc.2007.05.005.

Livi, C. M. *et al.* (2015) 'CatRAPID signature: Identification of ribonucleoproteins and RNA-binding regions', *Bioinformatics*, 32(5), pp. 773–775. doi: 10.1093/bioinformatics/btv629.

Llanos-Cuentas, A. *et al.* (2010) 'A clinical trial to evaluate the safety and immunogenicity of the LEISH-F1+MPL-SE vaccine when used in combination with sodium stibogluconate for the treatment of mucosal leishmaniasis', *Vaccine*. Elsevier Ltd, 28(46), pp. 7427–7435. doi: 10.1016/j.vaccine.2010.08.092.

Louradour, I. *et al.* (2022) 'Stress conditions promote *Leishmania* hybridization in vitro marked by expression of the ancestral gamete fusogen HAP2 as revealed by single-cell RNAseq', *eLife*, 11, pp. 1–23. doi: 10.7554/eLife.73488.

Love, D. C., Kane, M. M. and Mosser, D. M. (1998) '*Leishmania amazonensis*: The phagocytosis of amastigotes by macrophages', *Experimental Parasitology*, 88(3), pp. 161–171. doi: 10.1006/expr.1998.4232.

Low, G. C. (1919) 'Antimony in the treatment of American Leishmaniasis of the skin', *BMJ*, 1(3042), pp. 479–480. doi: 10.1136/bmj.1.3042.479.

van Luenen, H. G. A. M. *et al.* (2012) 'Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*', *Cell*, 150(5), pp. 909–921. doi: 10.1016/j.cell.2012.07.030.

Lueong, S. *et al.* (2016) 'Gene expression regulatory networks in *Trypanosoma brucei*: Insights into the role of the mRNA-binding proteome', *Molecular Microbiology*, 100(3), pp. 457–471. doi: 10.1111/mmi.13328.

Lukeš, J. *et al.* (2014) ‘Evolution of parasitism in kinetoplastid flagellates’, *Molecular and Biochemical Parasitology*, 195(2), pp. 115–122. doi: 10.1016/j.molbiopara.2014.05.007.

Luque-Ortega, J. R. and Rivas, L. (2007) ‘Miltefosine (hexadecylphosphocholine) inhibits cytochrome c oxidase in *Leishmania donovani* promastigotes’, *Antimicrobial Agents and Chemotherapy*, 51(4), pp. 1327–1332. doi: 10.1128/AAC.01415-06.

Luu, V. D. *et al.* (2006) ‘Functional analysis of *Trypanosoma brucei* PUF1’, *Molecular and Biochemical Parasitology*, 150(2), pp. 340–349. doi: 10.1016/j.molbiopara.2006.09.007.

Lye, L. F. *et al.* (2010) ‘Retention and Loss of RNA interference pathways in trypanosomatid protozoans’, *PLoS Pathogens*, 6(10). doi: 10.1371/journal.ppat.1001161.

Maclean, L. M. *et al.* (2012) ‘Trafficking and release of *Leishmania* metacyclic HASPB on macrophage invasion’, *Cellular Microbiology*, 14(5), pp. 740–761. doi: 10.1111/j.1462-5822.2012.01756.x.

Madeddu, G. *et al.* (2014) ‘Mucocutaneous leishmaniasis as presentation of HIV infection in Sardinia, insular Italy’, *Parasitology International*. Elsevier Ireland Ltd, 63(1), pp. 35–36. doi: 10.1016/j.parint.2013.10.002.

Mair, B. *et al.* (2019) ‘Essential gene profiles for human pluripotent stem cells identify uncharacterized genes and substrate dependencies’, *Cell Reports*. ElsevierCompany., 27(2), pp. 599-615.e12. doi: 10.1016/j.celrep.2019.02.041.

Mani, J. *et al.* (2011) ‘Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery’, *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0022463.

Marchat, L. A. *et al.* (2015) ‘DEAD/DExH-Box RNA helicases in selected human parasites’, *Korean Journal of Parasitology*, 53(5), pp. 583–595. doi: 10.3347/kjp.2015.53.5.583.

Marchese, D. *et al.* (2016) ‘Advances in the characterization of RNA-binding proteins’, *Wiley Interdisciplinary Reviews: RNA*, 7(6), pp. 793–810. doi: 10.1002/wrna.1378.

Marcili, A. *et al.* (2014) ‘Phylogenetic relationships of *Leishmania* species based on trypanosomatid barcode (SSU rDNA) and gGAPDH genes: Taxonomic revision of *Leishmania (L.) infantum chagasi* in South America’, *Infection, Genetics and Evolution*. Elsevier B.V., 25, pp. 44–51. doi: 10.1016/j.meegid.2014.04.001.

Marra, F., Chiappetta, M. C. and Vincenti, V. (2014) ‘Ear, nose and throat manifestations of mucocutaneous Leishmaniasis: A literature review’, *Acta Biomedica*, 85(1), pp. 3–7.

Martel, D. *et al.* (2017) ‘Characterisation of Casein Kinase 1.1 in *Leishmania donovani* Using the CRISPR Cas9 Toolkit’, *BioMed Research International*, 2017(i). doi: 10.1155/2017/4635605.

Matia-González, A. M., Laing, E. E. and Gerber, A. P. (2015) ‘Conserved mRNA-binding

proteomes in eukaryotic organisms', *Nature Structural and Molecular Biology*, 22(12), pp. 1027–1033. doi: 10.1038/nsmb.3128.

Matte, C., Arango Duque, G. and Descoteaux, A. (2021) '*Leishmania donovani* metacyclic promastigotes impair phagosome properties in inflammatory monocytes', *Infection and Immunity*. Edited by J. P. J. Saeij, 89(7). doi: 10.1128/IAI.00009-21.

Matthews, K. R., Tschudi, C. and Ullu, E. (1994) 'A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes', *Genes and Development*, 8(4), pp. 491–501. doi: 10.1101/gad.8.4.491.

McDermott, S. M. and Stuart, K. (2017) 'The essential functions of KREPB4 are developmentally distinct and required for endonuclease association with editosomes', *Rna*, 23(11), pp. 1672–1684. doi: 10.1261/rna.062786.117.

McDonald, L. *et al.* (2018) 'Non-linear hierarchy of the quorum sensing signalling pathway in bloodstream form African trypanosomes', *PLoS Pathogens*, 14(6), pp. 1–31. doi: 10.1371/journal.ppat.1007145.

McGwire, B. S. and Satoskar, A. R. (2014) 'Leishmaniasis: Clinical syndromes and treatment', *Qjm*, 107(1), pp. 7–14. doi: 10.1093/qjmed/hct116.

Membrive, N. A. *et al.* (2017) 'Disseminated cutaneous leishmaniasis caused by *Leishmania braziliensis* in Southern Brazil', *Revista do Instituto de Medicina Tropical de São Paulo*, 59(January), pp. 2–5. doi: 10.1590/s1678-9946201759037.

Miller, M. M. and Read, L. K. (2003) '*Trypanosoma brucei*: Functions of RBP16 cold shock and RGG domains in macromolecular interactions', *Experimental Parasitology*, 105(2), pp. 140–148. doi: 10.1016/j.exppara.2003.12.002.

Mishra, A. *et al.* (2021) 'Selective nuclear export of mRNAs is promoted by DRBD18 in *Trypanosoma brucei*', *Molecular Microbiology*, 116(3), pp. 827–840. doi: 10.1111/mmi.14773.

Mishra, M. *et al.* (1992) 'Amphotericin versus pentamidine in antimony-unresponsive kala-azar', *The Lancet*, 340(8830), pp. 1256–1257. doi: 10.1016/0140-6736(92)92952-C.

Mißlitz, A. *et al.* (2000) 'Targeted integration into a rRNA locus results in uniform and high level expression of transgenes in *Leishmania* amastigotes', *Molecular and Biochemical Parasitology*, 107(2), pp. 251–261. doi: 10.1016/S0166-6851(00)00195-X.

Mistry, J. *et al.* (2021) 'Pfam: The protein families database in 2021', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D412–D419. doi: 10.1093/nar/gkaa913.

Mitchell, S. F. *et al.* (2013) 'Global analysis of yeast mRNPs', *Nature Structural & Molecular Biology*, 20(1), pp. 127–133. doi: 10.1038/nsmb.2468.

Moafi, M. *et al.* (2019) '*Leishmania* vaccines entered in clinical trials: A review of literature',

International Journal of Preventive Medicine, 10(1), p. 95. doi: 10.4103/ijpvm.IJPVM_116_18.

Molyneux, D. H., Killick-Kendrick, R. and Ashford, R. W. (1975) 'Leishmania in phlebotomid sandflies III. The ultrastructure of *Leishmania mexicana amazonensis* in the midgut and pharynx of *Lutzomyia longipalpis*', *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 190(1100), pp. 341–357. doi: 10.1098/rspb.1975.0098.

Mondelaers, A. *et al.* (2016) 'Genomic and molecular characterization of Miltefosine resistance in *Leishmania infantum* strains with either natural or acquired resistance through experimental selection of intracellular amastigotes', *PLoS ONE*, 11(4), pp. 1–15. doi: 10.1371/journal.pone.0154101.

Monge-Maillo, B. and López-Vélez, R. (2013a) 'Therapeutic options for old world cutaneous leishmaniasis and new world cutaneous and mucocutaneous leishmaniasis', *Drugs*, 73(17), pp. 1889–1920. doi: 10.1007/s40265-013-0132-1.

Monge-Maillo, B. and López-Vélez, R. (2013b) 'Therapeutic options for visceral leishmaniasis', *Drugs*, 73(17), pp. 1863–1888. doi: 10.1007/s40265-013-0133-0.

Monge-Maillo, B., López-Vélez, R. and Saravolatz, L. D. (2015) 'Miltefosine for visceral and cutaneous leishmaniasis: Drug characteristics and evidence-based treatment recommendations', *Clinical Infectious Diseases*, 60(9), pp. 1398–1404. doi: 10.1093/cid/civ004.

Moreira, W. *et al.* (2014) 'Proteomic analysis of metacyclogenesis in *Leishmania infantum* wild-type and PTR1 null mutant', *EuPA Open Proteomics*. European Proteomics Association (EuPA), 4, pp. 171–183. doi: 10.1016/j.euprot.2014.07.003.

Morgan, J. F., Morton, H. J. and Parker, R. C. (1950) 'Nutrition of animal cells in tissue culture. I. Initial studies on a synthetic medium', *Experimental Biology and Medicine*, 73(1), pp. 1–8. doi: 10.3181/00379727-73-17557.

Mörking, P. A. *et al.* (2004) 'TcZFP1: A CCCH zinc finger protein of *Trypanosoma cruzi* that binds poly-C oligoribonucleotides in vitro', *Biochemical and Biophysical Research Communications*, 319(1), pp. 169–177. doi: 10.1016/j.bbrc.2004.04.162.

Mörking, P. A. *et al.* (2012) 'The zinc finger protein TcZFP2 binds target mRNAs enriched during *Trypanosoma cruzi* metacyclogenesis', *Memórias do Instituto Oswaldo Cruz*, 107(6), pp. 790–799. doi: 10.1590/S0074-02762012000600014.

Morrison, C. and Gannon, F. (1994) 'The impact of the PCR plateau phase on quantitative PCR', *BBA - Gene Structure and Expression*, 1219(2), pp. 493–498. doi: 10.1016/0167-4781(94)90076-0.

Mosleh, I. M. *et al.* (2008) 'Cutaneous leishmaniasis in the Jordanian side of the Jordan Valley: Severe under-reporting and consequences on public health management', *Tropical Medicine and International Health*, 13(6), pp. 855–860. doi: 10.1111/j.1365-3156.2008.02063.x.

Moss, C. X. *et al.* (2015) 'An essential signal peptide peptidase identified in an RNAi screen of serine peptidases of *Trypanosoma brucei*', *PLOS ONE*. Edited by M. M. Rodrigues, 10(3), p. e0123241. doi: 10.1371/journal.pone.0123241.

Mugo, E. and Clayton, C. (2017) 'Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*', *PLOS Pathogens*. Edited by L. Read, 13(8), p. e1006560. doi: 10.1371/journal.ppat.1006560.

Muhich, M. L. and Boothroyd, J. C. (1988) 'Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis', *Molecular and Cellular Biology*, 8(9), pp. 3837–3846. doi: 10.1128/mcb.8.9.3837-3846.1988.

Müller, K., Matuschewski, K. and Silvie, O. (2011) 'The Puf-family RNA-binding protein Puf2 controls sporozoite conversion to liver stages in the malaria parasite', *PLoS ONE*, 6(5), pp. 1–10. doi: 10.1371/journal.pone.0019860.

Müller, M. *et al.* (2010) 'Rapid decay of unstable *Leishmania* mRNAs bearing a conserved retroposon signature 3'-UTR motif is initiated by a site-specific endonucleolytic cleavage without prior deadenylation', *Nucleic Acids Research*, 38(17), pp. 5867–5883. doi: 10.1093/nar/gkq349.

Müller, M., Padmanabhan, P. K. and Papadopoulou, B. (2010) 'Selective inactivation of SIDER2 retroposon-mediated mRNA decay contributes to stage- and species-specific gene expression in *Leishmania*', *Molecular Microbiology*, 77(2), pp. 471–491. doi: 10.1111/j.1365-2958.2010.07226.x.

Naguleswaran, A. *et al.* (2015) '*Trypanosoma brucei* RRM1 is a nuclear RNA-binding protein and modulator of chromatin structure', *mBio*, 6(2), pp. 1–11. doi: 10.1128/mBio.00114-15.

Najafabadi, H. S. *et al.* (2013) 'Global identification of conserved post-transcriptional regulatory programs in trypanosomatids', *Nucleic Acids Research*, 41(18), pp. 8591–8600. doi: 10.1093/nar/gkt647.

Nandan, D. *et al.* (2017) 'Comprehensive identification of mRNABinding proteins of *Leishmania donovani* by interactome capture', *PLoS ONE*, 12(1), pp. 1–16. doi: 10.1371/journal.pone.0170068.

Názer, E., Verdún, R. E. and Sánchez, D. O. (2011) 'Nucleolar localization of RNA binding proteins induced by Actinomycin D and heat shock in *Trypanosoma cruzi*', *PLoS ONE*, 6(5). doi: 10.1371/journal.pone.0019920.

Negreira, G. H. *et al.* (2021) 'High throughput single-cell genome sequencing gives insights into the generation and evolution of mosaic aneuploidy in *Leishmania donovani*', *Nucleic Acids Research*. Oxford University Press, (15), p. 2021.05.11.443577. doi: 10.1093/nar/gkab1203.

Nett, I. R. E. *et al.* (2009) 'The phosphoproteome of bloodstream form *Trypanosoma brucei*, causative agent of African sleeping sickness', *Molecular and Cellular Proteomics*, 8(7), pp. 1527–1538. doi: 10.1074/mcp.M800556-MCP200.

Nielsen, H. and Engelbrecht, J. (1997) 'Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites Artificial neural networks have been used for many biological', *Protein Engineering*, 10(1), pp. 1–6. doi: 10.1142/S0129065797000537.

Nirujogi, R. S. *et al.* (2014) 'Moving from unsequenced to sequenced genome: Reanalysis of the proteome of *Leishmania donovani*', *Journal of Proteomics*. Elsevier B.V., 97, pp. 48–61. doi: 10.1016/j.jprot.2013.04.021.

Nocua, P. A. *et al.* (2017) '*Leishmania braziliensis* SCD6 and RBP42 proteins, two factors with RNA binding capacity', *Parasites and Vectors*. Parasites & Vectors, 10(1), pp. 1–14. doi: 10.1186/s13071-017-2557-y.

Nocua, P. A., Requena, J. M. and Puerta, C. J. (2021) 'Identification of the interactomes associated with SCD6 and RBP42 proteins in *Leishmania braziliensis*', *Journal of Proteomics*, 233(November 2020). doi: 10.1016/j.jprot.2020.104066.

Norris, A. D., Gracida, X. and Calarco, J. A. (2017) 'CRISPR-mediated genetic interaction profiling identifies RNA binding proteins controlling metazoan fitness', *eLife*, 6, pp. 1–18. doi: 10.7554/eLife.28129.

Nostrand, E. L. Van *et al.* (2017) 'A Large-Scale Binding and Functional Map of Human RNA Binding Proteins Correspondence and requests for materials should be addressed to Brenton Graveley (graveley@uchc.edu), Chris Burge (cburge@mit.edu), Xiang-dong Fu (xdfu@ucsd.edu)', *bioRxiv*, pp. 1–74. doi: 10.1101/179648.

Van Nostrand, E. L., Freese, P., *et al.* (2020) 'A large-scale binding and functional map of human RNA-binding proteins.', *Nature*. Springer US, 583(7818), pp. 711–719. doi: 10.1038/s41586-020-2077-3.

Van Nostrand, E. L., Pratt, G. A., *et al.* (2020) 'Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins', *Genome Biology*. Genome Biology, 21(1), pp. 1–26. doi: 10.1186/s13059-020-01982-9.

Oliveira, C. *et al.* (2017) 'RNA-binding proteins and their role in the regulation of gene expression in *Trypanosoma cruzi* and *Saccharomyces cerevisiae*', *Genetics and Molecular Biology*, 40(1), pp. 22–30. doi: 10.1590/1678-4685-GMB-2016-0258.

Oliveira, C. *et al.* (2021) 'Characterization of the RNA-binding protein TcSgn1 in *Trypanosoma cruzi*', *Microorganisms*, 9(5), pp. 1–12. doi: 10.3390/microorganisms9050986.

Opperdoes, F. R. and Szikora, J. P. (2006) 'In silico prediction of the glycosomal enzymes of *Leishmania major* and trypanosomes', *Molecular and Biochemical Parasitology*, 147(2), pp. 193–206. doi: 10.1016/j.molbiopara.2006.02.010.

Osman, M. *et al.* (2017) 'A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63-KH', *PLoS Neglected Tropical*

Diseases, 11(5), pp. 1–24. doi: 10.1371/journal.pntd.0005527.

Ouellette, M. *et al.* (1991) 'Direct and inverted DNA repeats associated with P-glycoprotein gene amplification in drug resistant *Leishmania*', *EMBO Journal*, 10(4), pp. 1009–1016. doi: 10.1002/j.1460-2075.1991.tb08035.x.

Oyama, R. K. *et al.* (2008) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.', *Plant systematics and evolution = Entwicklungsgeschichte und Systematik der Pflanzen*, 273(3–4), pp. 3389–3402. doi: 10.1093/oxfordjournals.molbev.a026201.

Ozgur, S. *et al.* (2015) 'Structure of a human 4E-T/DDX6/CNOT1 complex reveals the different interplay of DDX6-binding proteins with the CCR4-NOT complex', *Cell Reports*, 13(4), pp. 703–711. doi: 10.1016/j.celrep.2015.09.033.

Paape, D. *et al.* (2010) 'Gel free analysis of the proteome of intracellular *Leishmania mexicana*', *Molecular and Biochemical Parasitology*, 169(2), pp. 108–114. doi: 10.1016/j.molbiopara.2009.10.009.

De Pablos, L. M. *et al.* (2017) 'Characterization of RBP9 and RBP10, two developmentally regulated RNA-binding proteins in *Trypanosoma brucei*', *Open Biology*, 7(4). doi: 10.1098/rsob.160159.

De Pablos, L. M. *et al.* (2019) 'The mRNA-bound proteome of *Leishmania mexicana* : novel genetic insight into an ancient parasite', *Molecular & Cellular Proteomics*, (July), p. mcp.RA118.001307. doi: 10.1074/mcp.ra118.001307.

De Pablos, L. M., Ferreira, T. R. and Walrad, P. B. (2016) 'Developmental differentiation in Leishmania lifecycle progression : post-transcriptional control conducts the orchestra', *Current Opinion in Microbiology*. Elsevier Ltd, 34(Figure 1), pp. 82–89. doi: 10.1016/j.mib.2016.08.004.

Padmanabhan, P. K. *et al.* (2012) 'Apoptosis-like programmed cell death induces antisense ribosomal RNA (rRNA) fragmentation and rRNA degradation in *Leishmania*.', *Cell death and differentiation*. Nature Publishing Group, 19(12), pp. 1972–82. doi: 10.1038/cdd.2012.85.

Padmanabhan, P. K. *et al.* (2016) 'DDX3 DEAD-box RNA helicase plays a central role in mitochondrial protein quality control in *Leishmania*', *Cell Death and Disease*. Nature Publishing Group, 7(10), p. e2406. doi: 10.1038/cddis.2016.315.

de Paiva, R. M. C. *et al.* (2015) 'Amastin knockdown in *Leishmania braziliensis* affects parasite-macrophage interaction and results in impaired viability of intracellular amastigotes', *PLoS Pathogens*, 11(12), pp. 1–24. doi: 10.1371/journal.ppat.1005296.

Paris, C. *et al.* (2004) 'Miltefosine induces apoptosis-like death in *Leishmania donovani* promastigotes', *Antimicrobial Agents and Chemotherapy*, 48(3), pp. 852–859. doi: 10.1128/AAC.48.3.852-859.2004.

Parker, R. and Song, H. (2004) ‘The enzymes and control of eukaryotic mRNA turnover’, *Nature Structural and Molecular Biology*, 11(2), pp. 121–127. doi: 10.1038/nsmb724.

Parnas, O. *et al.* (2015) ‘A Genome-wide CRISPR screen in primary immune cells to dissect regulatory networks’, *Cell*. Elsevier Inc., 162(3), pp. 675–686. doi: 10.1016/j.cell.2015.06.059.

Parvez, S. *et al.* (2020) ‘Modified solid lipid nanoparticles encapsulated with Amphotericin B and Paromomycin: an effective oral combination against experimental murine visceral leishmaniasis’, *Scientific Reports*. Nature Publishing Group UK, 10(1), pp. 1–14. doi: 10.1038/s41598-020-69276-5.

Paterou, A. *et al.* (2009) ‘Identification and stage-specific association with the translational apparatus of TbZFP3 , a CCCH protein That promotes trypanosome life-cycle development’, 281(51), pp. 39002–39013. doi: 10.1074/jbc.M604280200.Identification.

Pelletier, M. *et al.* (2001) ‘Arginine methylation of a mitochondrial guide RNA binding protein from *Trypanosoma brucei*’, *Molecular and Biochemical Parasitology*, 118(1), pp. 49–59. doi: 10.1016/S0166-6851(01)00367-X.

Pelletier, M. and Read, L. K. (2003) ‘RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAS in *Trypanosoma brucei*’, *Rna*, 9(4), pp. 457–468. doi: 10.1261/rna.2160803.

Peng, D. *et al.* (2015) ‘CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*’, *mBio*, 6(1), pp. 1–11. doi: 10.1128/mBio.02097-14.

Pérez-Victoria, F. J. *et al.* (2006) ‘Phospholipid translocation and miltefosine potency require both *L. donovani* miltefosine transporter and the new protein LdRos3 in *Leishmania* parasites’, *Journal of Biological Chemistry*, 281(33), pp. 23766–23775. doi: 10.1074/jbc.M605214200.

Pérez-Victoria, F. J., Castanys, S. and Gamarro, F. (2003) ‘*Leishmania donovani* resistance to Miltefosine involves a defective inward translocation of the drug’, *Antimicrobial Agents and Chemotherapy*, 47(8), pp. 2397–2403. doi: 10.1128/AAC.47.8.2397-2403.2003.

Pescher, P. *et al.* (2011) ‘Quantitative proteome profiling informs on phenotypic traits that adapt *Leishmania donovani* for axenic and intracellular proliferation’, *Cellular Microbiology*, 13(7), pp. 978–991. doi: 10.1111/j.1462-5822.2011.01593.x.

Peters, C. *et al.* (1995) ‘The role of macrophage receptors in adhesion and uptake of *Leishmania mexicana* amastigotes’, *Journal of Cell Science*, 108(12), pp. 3715–3724. doi: 10.1242/jcs.108.12.3715.

Peters, N. C. and Sacks, D. L. (2009) ‘The impact of vector-mediated neutrophil recruitment on cutaneous leishmaniasis’, *Cellular Microbiology*, 11(9), pp. 1290–1296. doi: 10.1111/j.1462-5822.2009.01348.x.

Petrillo-Peixoto, M. L. and Beverley, S. M. (1988) ‘Amplified DNAs in laboratory stocks of *Leishmania tarentolae*: extrachromosomal circles structurally and functionally similar to the inverted-H-region amplification of methotrexate-resistant *Leishmania major*’, *Molecular and Cellular Biology*, 8(12), pp. 5188–5199. doi: 10.1128/mcb.8.12.5188-5199.1988.

Pimenta, P. F. P. *et al.* (1992) ‘Stage-specific adhesion of *Leishmania* promastigotes to the sandfly midgut’, *Science*, 256(5065), pp. 1812–1815. doi: 10.1126/science.1615326.

Pimenta, P. F. P. *et al.* (1994) ‘Evidence that the vectorial competence of phlebotomine sand flies for different species of *Leishmania* is controlled by structural polymorphisms in the surface lipophosphoglycan’, *Proceedings of the National Academy of Sciences of the United States of America*, 91(19), pp. 9155–9159. doi: 10.1073/pnas.91.19.9155.

Pinna, R. A. *et al.* (2016) ‘Malaria-cutaneous leishmaniasis co-infection: Influence on disease outcomes and immune response’, *Frontiers in Microbiology*, 7(JUN), pp. 1–15. doi: 10.3389/fmicb.2016.00982.

Pitale, D. M. *et al.* (2019) ‘*Leishmania donovani* induces autophagy in human blood–derived neutrophils’, *The Journal of Immunology*, 202(4), pp. 1163–1175. doi: 10.4049/jimmunol.1801053.

Podešvová, L., Huang, H. and Yurchenko, V. (2017) ‘Inducible protein stabilization system in *Leishmania mexicana*’, *Molecular and Biochemical Parasitology*, 214, pp. 62–64. doi: 10.1016/j.molbiopara.2017.03.008.

Ponte-Sucre, A. *et al.* (2017) ‘Drug resistance and treatment failure in leishmaniasis: A 21st century challenge’, *PLoS Neglected Tropical Diseases*, 11(12), pp. 1–24. doi: 10.1371/journal.pntd.0006052.

Price, H. P. *et al.* (2003) ‘Myristoyl-CoA:Protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites’, *Journal of Biological Chemistry*. Â© 2003 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 278(9), pp. 7206–7214. doi: 10.1074/jbc.M211391200.

Quijada, L. *et al.* (2002) ‘Expression of the human RNA-binding protein HuR in *Trypanosoma brucei* increases the abundance of mRNAs containing AU-rich regulatory elements’, *Nucleic Acids Research*, 30(20), pp. 4414–4424. doi: 10.1093/nar/gkf577.

Rachidi, N. *et al.* (2014) ‘Pharmacological assessment defines *Leishmania donovani* casein kinase 1 as a drug target and reveals important functions in parasite viability and intracellular infection’, *Antimicrobial Agents and Chemotherapy*, 58(3), pp. 1501–1515. doi: 10.1128/AAC.02022-13.

Radhakrishnan, A. *et al.* (2016) ‘The DEAD-Box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality’, *Cell*. Elsevier Inc., 167(1), pp. 122-132.e9. doi: 10.1016/j.cell.2016.08.053.

Rajasekaran, R. and Chen, Y. P. P. (2015) ‘Potential therapeutic targets and the role of technology

in developing novel antileishmanial drugs', *Drug Discovery Today*. Elsevier Ltd, 20(8), pp. 958–968. doi: 10.1016/j.drudis.2015.04.006.

Ralph, P. and Nakoinz, I. (1975) 'Phagocytosis and cytolysis by a macrophage tumour and its cloned cell line', *Nature*, 257(5525), pp. 393–394. doi: 10.1038/257393a0.

Ramírez, C. A. *et al.* (2013) 'Identification of proteins interacting with HSP70 mRNAs in *Leishmania braziliensis*', *Journal of Proteomics*. The Authors, 94, pp. 124–137. doi: 10.1016/j.jprot.2013.09.008.

Ranjan, A. *et al.* (2005) 'Risk factors for Indian kala-azar', *American Journal of Tropical Medicine and Hygiene*, 73(1), pp. 74–78. doi: 10.4269/ajtmh.2005.73.74.

Ready, P. D. (2010) 'Leishmaniasis emergence in Europe', *Eurosurveillance*, 15(10), pp. 29–39. doi: 10.2807/ese.15.10.19505-en.

Ready, P. D. (2013) 'Biology of phlebotomine sand flies as vectors of disease agents', *Annual Review of Entomology*, 58(1), pp. 227–250. doi: 10.1146/annurev-ento-120811-153557.

Ready, P. D. (2014) 'Epidemiology of visceral leishmaniasis', *Clinical Epidemiology*, 6(1), pp. 147–154. doi: 10.2147/CLEP.S44267.

Real, F. *et al.* (2014) 'Cell-to-cell transfer of *Leishmania amazonensis* amastigotes is mediated by immunomodulatory LAMP-rich parasitophorous extrusions', *Cellular Microbiology*, 16(10), pp. 1549–1564. doi: 10.1111/cmi.12311.

Reis-Cunha, J. L., Valdivia, H. O. and Bartholomeu, D. C. (2017) 'Gene and chromosomal copy number variations as an adaptive mechanism towards a parasitic lifestyle in trypanosomatids', *Current Genomics*, 19(2), pp. 87–97. doi: 10.2174/1389202918666170911161311.

de Rezende, E. *et al.* (2017) 'Quantitative proteomic analysis of amastigotes from *Leishmania (L.) amazonensis* LV79 and PH8 strains reveals molecular traits associated with the virulence phenotype', *PLoS Neglected Tropical Diseases*, 11(11), pp. 1–18. doi: 10.1371/journal.pntd.0006090.

Rico, E. *et al.* (2017) 'Genome-wide RNAi selection identifies a regulator of transmission stage-enriched gene families and cell-type differentiation in *Trypanosoma brucei*', *PLoS Pathogens*, 13(3), pp. 1–25. doi: 10.1371/journal.ppat.1006279.

Rico, E. *et al.* (2018) 'Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes', *Scientific Reports*, 8(1), pp. 1–10. doi: 10.1038/s41598-018-26303-w.

Ridley, D. S. (1980) 'A histological classification of cutaneous leishmaniasis and its geographical expression', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 74(4), pp. 515–521. doi: 10.1016/0035-9203(80)90069-3.

Rijal, S. *et al.* (2013) 'Increasing failure of miltefosine in the treatment of kala-azar in nepal and the potential role of parasite drug resistance, reinfection, or noncompliance', *Clinical Infectious Diseases*, 56(11), pp. 1530–1538. doi: 10.1093/cid/cit102.

Ritmeijer, K. *et al.* (2011) 'Limited effectiveness of high-dose liposomal amphotericin B (AmBisome) for treatment of visceral leishmaniasis in an ethiopian population with high HIV prevalence', *Clinical Infectious Diseases*, 53(12). doi: 10.1093/cid/cir674.

Rittig, M. G. *et al.* (1998) 'Coiling phagocytosis discriminates between different spirochetes and is enhanced by phorbol myristate acetate and granulocyte-macrophage colony- stimulating factor', *Infection and Immunity*, 66(2), pp. 627–635. doi: 10.1128/iai.66.2.627-635.1998.

Roatt, B. M. *et al.* (2020) 'Recent advances and new strategies on leishmaniasis treatment', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 104(21), pp. 8965–8977. doi: 10.1007/s00253-020-10856-w.

Rocha, N. A. *et al.* (2011) 'Visceral leishmaniasis in children: A cohort of 120 patients in a metropolitan city of Brazil', *Turkish Journal of Pediatrics*, 53(2), pp. 154–160.

Rochette, A. *et al.* (2005) 'Characterization and developmental gene regulation of a large gene family encoding amastin surface proteins in *Leishmania spp.*', *Molecular and Biochemical Parasitology*, 140(2), pp. 205–220. doi: 10.1016/j.molbiopara.2005.01.006.

Rodrigues V., J., Da Silva, J. S. and Campos-Neto, A. (1998) 'Transforming growth factor β and immunosuppression in experimental visceral leishmaniasis', *Infection and Immunity*, 66(3), pp. 1233–1236. doi: 10.1128/iai.66.3.1233-1236.1998.

Rogers, L. (1939) 'The antimony treatment of kala-azar', *Nature*, 144(3659), pp. 1003–1004. doi: 10.1038/1441003a0.

Rogers, M. B. *et al.* (2011) 'Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*', *Genome Res.*, 21(12), pp. 2129–2142. doi: 10.1101/gr.122945.111.Freely.

Rogers, M. E. *et al.* (2004) 'Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG', *Nature*, 430(6998), pp. 463–467. doi: 10.1038/nature02675.

Romagnoli, B. A. A. *et al.* (2020) 'RNA Binding Proteins and Gene Expression Regulation in *Trypanosoma cruzi*', *Frontiers in Cellular and Infection Microbiology*, 10(February), pp. 1–17. doi: 10.3389/fcimb.2020.00056.

Romaniuk, M. A. (2016) 'Regulation of RNA binding proteins in trypanosomatid protozoan parasites', *World Journal of Biological Chemistry*, 7(1), p. 146. doi: 10.4331/wjbc.v7.i1.146.

Rosa, M. E. A. and MacHado, P. R. L. (2011) 'Disseminated leishmaniasis: clinical, immunological, and therapeutic aspects', *Drug Development Research*, 72(6), pp. 437–441. doi:

10.1002/ddr.20450.

Rosen, B. P. (2002) ‘Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes’, *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 133(3), pp. 689–693. doi: 10.1016/S1095-6433(02)00201-5.

Rosenberg, M. *et al.* (2017) ‘Denaturing CLIP, dCLIP, pipeline identifies discrete RNA footprints on chromatin-associated proteins and reveals that CBX7 targets 3′ UTRs to regulate mRNA expression’, *Cell Systems*. Elsevier Inc., 5(4), pp. 368-385.e15. doi: 10.1016/j.cels.2017.09.014.

Roux, K. J. *et al.* (2018) ‘BioID: A screen for protein-protein interactions.’, *Current protocols in protein science*, 91(November 2013), pp. 19.23.1-19.23.15. doi: 10.1002/cpps.51.

Le Rutte, E. A. *et al.* (2016) ‘Feasibility of eliminating visceral leishmaniasis from the Indian subcontinent: Explorations with a set of deterministic age-structured transmission models’, *Parasites and Vectors*. Parasites & Vectors, 9(1), pp. 1–14. doi: 10.1186/s13071-016-1292-0.

Ruy, P. D. C. *et al.* (2019) ‘Comparative transcriptomics in *Leishmania braziliensis*: disclosing differential gene expression of coding and putative noncoding RNAs across developmental stages’, *RNA Biology*. Taylor & Francis, 16(5), pp. 639–660. doi: 10.1080/15476286.2019.1574161.

Sacks, D. L. and Perkins, P. V. (1984) ‘Identification of an infective stage of *Leishmania* promastigotes’, *Science*, 223(4643), pp. 1417–1419. doi: 10.1126/science.6701528.

Sádlová, J. *et al.* (2010) ‘The stage-regulated HASPB and SHERP proteins are essential for differentiation of the protozoan parasite *Leishmania major* in its sand fly vector, *Phlebotomus papatasi*’, *Cellular Microbiology*, 12(12), pp. 1765–1779. doi: 10.1111/j.1462-5822.2010.01507.x.

Salah, A. B. *et al.* (1995) ‘A randomized, placebo-controlled trial in Tunisia treating cutaneous leishmaniasis with paromomycin ointment’, *American Journal of Tropical Medicine and Hygiene*, 53(2), pp. 162–166. doi: 10.4269/ajtmh.1995.53.162.

Salam, N., Al-Shaqha, W. M. and Azzi, A. (2014) ‘Leishmaniasis in the Middle East: Incidence and epidemiology’, *PLoS Neglected Tropical Diseases*, 8(10), pp. 1–8. doi: 10.1371/journal.pntd.0003208.

Salinovich, O. and Montelaro, R. C. (1986) ‘Reversible staining and peptide mapping of proteins transferred to nitrocellulose after separation by sodium dodecylsulfate-polyacrylamide gel electrophoresis’, *Analytical Biochemistry*, 156(2), pp. 341–347. doi: 10.1016/0003-2697(86)90263-0.

Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning*. doi: 10.3724/sp.j.1141.2012.01075.

Sanchiz, Á. *et al.* (2020) ‘The experimental proteome of *Leishmania infantum* promastigote and its usefulness for improving gene annotations’, *Genes*, 11(9), pp. 1–20. doi: 10.3390/genes11091036.

Sanderson, L., Yardley, V. and Croft, S. L. (2014) ‘Activity of anti-cancer protein kinase inhibitors

against *Leishmania spp.*', *Journal of Antimicrobial Chemotherapy*, 69(7), pp. 1888–1891. doi: 10.1093/jac/dku069.

Santos, R. E. R. S. *et al.* (2017) 'A DiCre recombinase-based system for inducible expression in *Leishmania major*', *Molecular and Biochemical Parasitology*. Elsevier, 216(March), pp. 45–48. doi: 10.1016/j.molbiopara.2017.06.006.

Saravanan, P. *et al.* (2010) 'Mitogen-activated protein kinase 4 of *Leishmania* parasite as a therapeutic target', *European Journal of Medicinal Chemistry*. Elsevier Masson SAS, 45(12), pp. 5662–5670. doi: 10.1016/j.ejmech.2010.09.020.

Saxena, A. *et al.* (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), pp. 53–65. doi: 10.1016/j.molbiopara.2006.11.011.

Sbicego, S. *et al.* (2003) 'RBP38, a novel RNA-binding protein from trypanosomatid mitochondria, modulates RNA stability', *Eukaryotic Cell*, 2(3), pp. 560–568. doi: 10.1128/EC.2.3.560-568.2003.

Scherrer, T. *et al.* (2010) 'A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes', *PLoS ONE*, 5(11). doi: 10.1371/journal.pone.0015499.

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. doi: 10.1038/nmeth.2019.

Schlein, Y., Jacobson, R. L. and Messer, G. (1992) '*Leishmania* infections damage the feeding mechanism of the sandfly vector and implement parasite transmission by bite', *Proceedings of the National Academy of Sciences of the United States of America*, 89(20), pp. 9944–9948. doi: 10.1073/pnas.89.20.9944.

Schumann Burkard, G. *et al.* (2013) 'Nucleolar proteins regulate stage-specific gene expression and ribosomal RNA maturation in *Trypanosoma brucei*', *Molecular Microbiology*, 88(4), pp. 827–840. doi: 10.1111/mmi.12227.

Schütz, P. *et al.* (2010) 'Comparative structural analysis of human DEAD-Box RNA helicases', *PLoS ONE*, 5(9), pp. 1–11. doi: 10.1371/journal.pone.0012791.

Seifert, K. *et al.* (2007) 'Inactivation of the miltefosine transporter, LdMT, causes miltefosine resistance that is conferred to the amastigote stage of *Leishmania donovani* and persists *in vivo*', *International Journal of Antimicrobial Agents*, 30(3), pp. 229–235. doi: 10.1016/j.ijantimicag.2007.05.007.

Senior, A. W. *et al.* (2020) 'Improved protein structure prediction using potentials from deep learning', *Nature*. Springer US, 577(7792), pp. 706–710. doi: 10.1038/s41586-019-1923-7.

Serafim, T. D. *et al.* (2018) 'Sequential blood meals promote *Leishmania* replication and reverse

metacyclogenesis augmenting vector infectivity', *Nature Microbiology*. Springer US. doi: 10.1038/s41564-018-0125-7.

Serpeloni, M. *et al.* (2011) 'An essential nuclear protein in trypanosomes is a component of mrna transcription/export pathway', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0020730.

Shah, S. M. A. *et al.* (2012) 'Clinical and hematological presentation of visceral leishmaniasis in children in a tertiary care hospital', *Journal of Medical Sciences (Peshawar)*, 20(2), pp. 94–97.

Shalem, O. *et al.* (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science*, 343(6166), pp. 84–87. doi: 10.1126/science.1247005.

Shalev, M. *et al.* (2013) 'Identification of the molecular attributes required for aminoglycoside activity against *Leishmania*', *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), pp. 13333–13338. doi: 10.1073/pnas.1307365110.

Shapiro, S. Z. *et al.* (1984) 'Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes', *Acta Tropica*, 41(4), pp. 313–323. doi: 10.5169/seals-313308.

Sharma, N. *et al.* (2012) 'Evaluation of plumbagin and its derivative as potential modulators of redox thiol metabolism of *Leishmania* parasite', *Parasitology Research*, 110(1), pp. 341–348. doi: 10.1007/s00436-011-2498-x.

Sharma, U. and Singh, S. (2008) 'Insect vectors of *Leishmania*: Distribution, physiology and their control', *Journal of Vector Borne Diseases*, 45(4), pp. 255–272.

Shehzad, T. and Abbas, A. (2010) 'A case of mucocutaneous leishmaniasis from Interior Sindh', *Journal of Pakistan Association of Dermatologists*, 20(3), pp. 180–181.

Shimozako, H. J., Wu, J. and Massad, E. (2017) 'Mathematical modelling for Zoonotic Visceral Leishmaniasis dynamics: A new analysis considering updated parameters and notified human Brazilian data', *Infectious Disease Modelling*. Elsevier Ltd, 2(2), pp. 143–160. doi: 10.1016/j.idm.2017.03.002.

Shiravand, B. *et al.* (2018) 'Modeling spatial risk of zoonotic cutaneous leishmaniasis in Central Iran', *Acta Tropica*, 185(May), pp. 327–335. doi: 10.1016/j.actatropica.2018.06.015.

Shrivastava, R., Drory-Retwitzer, M. and Shapira, M. (2018) 'Nutritional stress targets LeishIF4E-3 to storage granules that contain RNA and ribosome components in Leishmania', *PLoS Neglected Tropical Diseases*, 13(3), pp. 1–30. doi: 10.1371/journal.pntd.0007237.

Sidik, S. M. *et al.* (2016) 'A Genome-wide CRISPR screen in *Toxoplasma* identifies essential apicomplexan genes', *Cell*. Elsevier Inc., 166(6), pp. 1423-1435.e12. doi: 10.1016/j.cell.2016.08.019.

da Silva, R. and Sacks, D. L. (1987) 'Metacyclogenesis is a major determinant of *Leishmania* promastigote virulence and attenuation', *Infection and Immunity*, 55(11), pp. 2802–2806. doi:

10.1128/iai.55.11.2802-2806.1987.

da Silva Vieira, T. *et al.* (2019) '*Leishmania braziliensis*: Strain-Specific modulation of phagosome maturation', *Frontiers in Cellular and Infection Microbiology*, 9(September), pp. 1–12. doi: 10.3389/fcimb.2019.00319.

Simpson, A. G. B., Stevens, J. R. and Lukeš, J. (2006) 'The evolution and diversity of kinetoplastid flagellates', *Trends in Parasitology*, 22(4), pp. 168–174. doi: 10.1016/j.pt.2006.02.006.

Singh, G. *et al.* (2015) 'The clothes make the mRNA: Past and present trends in mRNP fashion', *Annual Review of Biochemistry*, 84, pp. 325–354. doi: 10.1146/annurev-biochem-080111-092106.

Singh, R. *et al.* (2016) 'Universal minicircle sequence binding protein of *Leishmania donovani* regulates pathogenicity by controlling expression of cytochrome-b', *Cell & Bioscience*. BioMed Central, 6(1), p. 13. doi: 10.1186/s13578-016-0072-z.

Singh, S. P. *et al.* (2006) 'Serious underreporting of visceral leishmaniasis through passive case reporting in Bihar, India', *Tropical Medicine and International Health*, 11(6), pp. 899–905. doi: 10.1111/j.1365-3156.2006.01647.x.

Singh, V. P. *et al.* (2010) 'Short report: Estimation of under-reporting of visceral leishmaniasis cases in Bihar, India', *American Journal of Tropical Medicine and Hygiene*, 82(1), pp. 9–11. doi: 10.4269/ajtmh.2010.09-0235.

Sinha, R. *et al.* (2018) 'Genome plasticity in cultured *Leishmania donovani*: Comparison of early and late passages', *Frontiers in Microbiology*, 9(JUL), pp. 1–20. doi: 10.3389/fmicb.2018.01279.

Smith, M., Bringaud, F. and Papadopoulou, B. (2009) 'Organization and evolution of two SIDER retroposon subfamilies and their impact on the *Leishmania* genome', *BMC Genomics*, 10, pp. 1–14. doi: 10.1186/1471-2164-10-240.

Smith, O. P. *et al.* (1995) 'Visceral leishmaniasis: Rapid response to AmBisome treatment', *Archives of Disease in Childhood*, 73(2), pp. 157–159. doi: 10.1136/adc.73.2.157.

Sollelis, L. *et al.* (2015) 'First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites', *Cellular Microbiology*, 17(10), pp. 1405–1412. doi: 10.1111/cmi.12456.

Sonnhammer, E. L. L. *et al.* (1998) 'Pfam: Multiple sequence alignments and HMM-profiles of protein domains', *Nucleic Acids Research*, 26(1), pp. 320–322. doi: 10.1093/nar/26.1.320.

Sonnhammer, E. L. L., Eddy, S. R. and Durbin, R. (1997) 'Pfam: A comprehensive database of protein domain families based on seed alignments', *Proteins: Structure, Function and Genetics*, 28(3), pp. 405–420. doi: 10.1002/(SICI)1097-0134(199707)28:3<405::AID-PROT10>3.0.CO;2-L.

Späth, G. F. and Beverley, S. M. (2001) 'A lipophosphoglycan-independent method for isolation of infective *Leishmania* metacyclic promastigotes by density gradient centrifugation', *Experimental Parasitology*, 99(2), pp. 97–103. doi: 10.1006/expr.2001.4656.

Staneva, D. P. *et al.* (2021) 'A systematic analysis of Trypanosoma brucei chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination', *Genome Research*, 31(11), pp. 2138–2154. doi: 10.1101/gr.275368.121.

Stempliuk, V. D. A. and Floeter-Winter, L. M. (2002) 'Functional domains of the rDNA promoter display a differential recognition in *Leishmania*', *International Journal for Parasitology*, 32(4), pp. 437–447. doi: 10.1016/S0020-7519(01)00371-X.

Sterkers, Y. *et al.* (2011) 'FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in *Leishmania major*', *Cellular Microbiology*, 13(2), pp. 274–283. doi: 10.1111/j.1462-5822.2010.01534.x.

Stern, M. Z. *et al.* (2009) 'Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism', *Rna*, 15(4), pp. 648–665. doi: 10.1261/rna.1230209.

Stevens, J. R. (2008) 'Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes', *Parasite*, 15(3), pp. 226–232. doi: 10.1051/parasite/2008153226.

Subota, I. *et al.* (2011) 'ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation', *Molecular Biology of the Cell*, 22(22), pp. 4205–4219. doi: 10.1091/mbc.E11-06-0511.

Sundar, S. *et al.* (1998) 'Trial of oral miltefosine for visceral leishmaniasis', *Lancet*, 352(9143), pp. 1821–1823. doi: 10.1016/S0140-6736(98)04367-0.

Sysoev, V. O. *et al.* (2016) 'Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms12128.

Tasbihi, M. *et al.* (2019) 'Mitochondrial proteome profiling of *Leishmania tropica*', *Microbial Pathogenesis*. Elsevier Ltd, 133(April), p. 103542. doi: 10.1016/j.micpath.2019.103542.

Taylor, V. M. *et al.* (2010) '*Leishmania tarentolae*: Utility as an in vitro model for screening of antileishmanial agents', *Experimental Parasitology*. Elsevier Inc., 126(4), pp. 471–475. doi: 10.1016/j.exppara.2010.05.016.

Teixeira, S. M. R. *et al.* (1994) 'A differentially expressed gene family encoding "amastin," a surface protein of *Trypanosoma cruzi* amastigotes', *Journal of Biological Chemistry*, 269(32), pp. 20509–20516. doi: 10.1016/s0021-9258(17)32022-7.

Tetaud, E. *et al.* (2002) 'A new expression vector for *Crithidia fasciculata* and *Leishmania*', *Molecular and Biochemical Parasitology*, 120(2), pp. 195–204. doi: 10.1016/S0166-6851(02)00002-6.

Thiel, M. and Bruchhaus, I. (2001) 'Comparative proteome analysis of *Leishmania donovani* at different stages of transformation from promastigotes to amastigotes', *Medical Microbiology and*

Immunology, 190(1–2), pp. 33–36. doi: 10.1007/s004300100075.

Thomas, S. *et al.* (2006) 'A non-universal transcription factor? The *Leishmania tarentolae* TATA box-binding protein LtTBP associates with a subset of promoters', *International Journal for Parasitology*, 36(10–11), pp. 1217–1226. doi: 10.1016/j.ijpara.2006.04.004.

Toh, J. Y. *et al.* (2021) 'Identification of positive and negative regulators in the stepwise developmental progression towards infectivity in *Trypanosoma brucei*', *Scientific Reports*. Nature Publishing Group UK, 11(1), pp. 1–14. doi: 10.1038/s41598-021-85225-2.

Topisirovic, I. *et al.* (2011) 'Cap and cap-binding proteins in the control of gene expression', *Wiley Interdisciplinary Reviews: RNA*, 2(2), pp. 277–298. doi: 10.1002/wrna.52.

Tsvetanova, N. G. *et al.* (2010) 'Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*', *PLoS ONE*, 5(9), pp. 1–12. doi: 10.1371/journal.pone.0012671.

Ubeda, J. M. *et al.* (2008) 'Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy', *Genome Biology*, 9(7). doi: 10.1186/gb-2008-9-7-r115.

Ubeda, J. M. *et al.* (2014) 'Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite *Leishmania*', *PLoS Biology*, 12(5). doi: 10.1371/journal.pbio.1001868.

Uliana, S. R. B. *et al.* (1996) 'Structural and functional characterization of the *Leishmania amazonensis* ribosomal RNA promoter', *Molecular and Biochemical Parasitology*, 76(1–2), pp. 245–255. doi: 10.1016/0166-6851(95)02562-6.

Ullu, E., Matthews, K. R. and Tschudi, C. (1993) 'Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts.', *Molecular and Cellular Biology*, 13(1), pp. 720–725. doi: 10.1128/mcb.13.1.720.

Vacchina, P. *et al.* (2016) 'Genomic appraisal of the multifactorial basis for in vitro acquisition of Miltefosine resistance in *Leishmania donovani*', *Antimicrobial Agents and Chemotherapy*, 60(7), pp. 4089–4100. doi: 10.1128/AAC.00478-16.

Vakili, B. *et al.* (2018) 'Proteome-scale identification of *Leishmania infantum* for novel vaccine candidates: A hierarchical subtractive approach', *Computational Biology and Chemistry*. Elsevier Ltd, 72, pp. 16–25. doi: 10.1016/j.compbiolchem.2017.12.008.

Varma, N. and Naseem, S. (2010) 'Hematologic changes in visceral Leishmaniasis/Kala Azar', *Indian Journal of Hematology and Blood Transfusion*, 26(3), pp. 78–82. doi: 10.1007/s12288-010-0027-1.

Velez, I. D. *et al.* (2001) 'Gender and cutaneous leishmaniasis in Colombia', *Cadernos de saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública*, 17(1),

pp. 171–180. doi: 10.1590/S0102-311X2001000100018.

Verma, N. K., Singh, G. and Dey, C. S. (2007) ‘Miltefosine induces apoptosis in arsenite-resistant *Leishmania donovani* promastigotes through mitochondrial dysfunction’, *Experimental Parasitology*, 116(1), pp. 1–13. doi: 10.1016/j.exppara.2006.10.007.

Vianna, G. (1912) ‘Tratamento da leishmaniose tegumentar por injeções intravenosas de tártaro emético’, *Annais do 7° Congresso Brasileiro de Medicina e Cirurgia*, 4, pp. 426–428.

Vigneron, A. *et al.* (2020) ‘Single-cell RNA sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens’, *Proceedings of the National Academy of Sciences of the United States of America*, 117(5), pp. 2613–2621. doi: 10.1073/pnas.1914423117.

Volf, P. *et al.* (2004) ‘Blocked stomodeal valve of the insect vector: Similar mechanism of transmission in two trypanosomatid models’, *International Journal for Parasitology*, 34(11), pp. 1221–1227. doi: 10.1016/j.ijpara.2004.07.010.

Volpon, L. *et al.* (2005) ‘NMR structural study of TcUBP1, a single RRM domain protein from *Trypanosoma cruzi* : Contribution of a β hairpin to RNA binding’, *Biochemistry*, 44(10), pp. 3708–3717. doi: 10.1021/bi047450e.

Walker, J. *et al.* (2006) ‘Identification of developmentally-regulated proteins in *Leishmania panamensis* by proteome profiling of promastigotes and axenic amastigotes’, *Molecular and Biochemical Parasitology*, 147(1), pp. 64–73. doi: 10.1016/j.molbiopara.2006.01.008.

Walker, J. *et al.* (2012) ‘Discovery of factors linked to antimony resistance in *Leishmania panamensis* through differential proteome analysis’, *Molecular and Biochemical Parasitology*. Elsevier B.V., 183(2), pp. 166–176. doi: 10.1016/j.molbiopara.2012.03.002.

Walker, R. G. *et al.* (2011) ‘High throughput screens yield small molecule inhibitors of *Leishmania* CRK3:CYC6 cyclin-dependent kinase’, *PLoS Neglected Tropical Diseases*, 5(4). doi: 10.1371/journal.pntd.0001033.

Walrad, P. *et al.* (2009) ‘Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA’, 5(2), pp. 1–10. doi: 10.1371/journal.ppat.1000317.

Walrad, P. B. *et al.* (2012) ‘The post-transcriptional *trans*-acting regulator , enriched mRNAs in Trypanosoma brucei’, 40(7), pp. 2869–2883. doi: 10.1093/nar/gkr1106.

Wang, E. *et al.* (2019) ‘Targeting an RNA-binding protein network in acute myeloid leukemia’, *Cancer Cell*. Elsevier Inc., 35(3), pp. 369-384.e7. doi: 10.1016/j.ccell.2019.01.010.

Wang, E. T. *et al.* (2008) ‘Alternative isoform regulation in human tissue transcriptomes’, *Nature*, 456(7221), pp. 470–476. doi: 10.1038/nature07509.

Wang, L. *et al.* (2010) ‘BindN+ for accurate prediction of DNA and RNA-binding residues from

protein sequence features', *BMC Systems Biology*, 4(SUPPL. 1), pp. 1–9. doi: 10.1186/1752-0509-4-S1-S3.

Wang, L. and Brown, S. J. (2006) 'BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences', *Nucleic Acids Research*, 34(WEB. SERV. ISS.), pp. 243–248. doi: 10.1093/nar/gkl298.

Ware, J. A. M. *et al.* (2021) 'Efficacy and tolerability of Miltefosine in the treatment of cutaneous leishmaniasis', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 73(7), pp. e2457–e2562. doi: 10.1093/cid/ciaa1238.

Weischenfeldt, J. and Porse, B. (2008) 'Bone marrow-derived macrophages (BMM): Isolation and applications', *Cold Spring Harbor Protocols*, 3(12), pp. 1–7. doi: 10.1101/pdb.prot5080.

Wessels, H. H. *et al.* (2016) 'The mRNA-bound proteome of the early fly embryo', *Genome Research*, 26(7), pp. 1000–1009. doi: 10.1101/gr.200386.115.

Wheeler, E. C. *et al.* (2020) 'Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors', *Nature Methods*. Springer US, 17(6), pp. 636–642. doi: 10.1038/s41592-020-0826-8.

Wheeler, R. J. (2021) 'A resource for improved predictions of Trypanosoma and *Leishmania* protein three-dimensional structure', *PLoS ONE*, 16(11 November), pp. 1–12. doi: 10.1371/journal.pone.0259871.

White, T. C. *et al.* (1988) 'The H circles of *Leishmania tarentolae* are a unique amplifiable system of oligomeric DNAs associated with drug resistance', *Journal of Biological Chemistry*, 263(32), pp. 16977–16983. doi: 10.1016/s0021-9258(18)37486-6.

WHO fact sheet: Leishmaniasis (2021). Available at: https://www.who.int/en/news-room/fact-sheets/detail/leishmaniasis (Accessed: 16 December 2021).

Wongsombat, C. *et al.* (2014) 'Molecular characterization of *Plasmodium falciparum* Bruno/CELF RNA binding proteins', *Molecular and Biochemical Parasitology*. Elsevier B.V., 198(1), pp. 1–10. doi: 10.1016/j.molbiopara.2014.10.005.

World Health Organization (2010) 'Control of the leishmaniases.', *World Health Organization technical report series*, (949), pp. 22–26.

World Health Organization (2021) *Leishmaniasis*. Available at: https://www.who.int/news-room/fact-sheets/detail/leishmaniasis (Accessed: 30 December 2021).

Wu, Y. *et al.* (2000) 'A new developmentally regulated gene family in *Leishmania* amastigotes encoding a homolog of amastin surface proteins', *Molecular and Biochemical Parasitology*, 110(2), pp. 345–357. doi: 10.1016/S0166-6851(00)00290-5.

Wurst, M. *et al.* (2009) 'An RNAi screen of the RRM-domain proteins of *Trypanosoma brucei*',

Molecular and Biochemical Parasitology, 163(1), pp. 61–65. doi: 10.1016/j.molbiopara.2008.09.001.

Wyllie, S. *et al.* (2018) 'Cyclin-dependent kinase 12 is a drug target for visceral leishmaniasis', *Nature*. Springer US, 560(7717), pp. 192–197. doi: 10.1038/s41586-018-0356-z.

Wyllie, S., Cunningham, M. L. and Fairlamb, A. H. (2004) 'Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*', *Journal of Biological Chemistry*. Â© 2004 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 279(38), pp. 39925–39932. doi: 10.1074/jbc.M405635200.

Wyllie, S., Vickers, T. J. and Fairlamb, A. H. (2008) 'Roles of trypanothione S-transferase and tryparedoxin peroxidase in resistance to antimonials', *Antimicrobial Agents and Chemotherapy*, 52(4), pp. 1359–1365. doi: 10.1128/AAC.01563-07.

Yagoubat, A. *et al.* (2020) 'Universal highly efficient conditional knockout system in *Leishmania*, with a focus on untranscribed region preservation', *Cellular Microbiology*, 22(December 2019), pp. 1–11. doi: 10.1111/cmi.13159.

Yan, S. *et al.* (1999) 'Characterization of the *Leishmania donovani* ribosomal RNA promoter', *Molecular and Biochemical Parasitology*, 103(2), pp. 197–210. doi: 10.1016/S0166-6851(99)00126-7.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.', *BMC bioinformatics*, 13, p. 134. doi: 10.1186/1471-2105-13-134.

Yilmaz, A. *et al.* (2018) 'Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells', *Nature Cell Biology*. Springer US, 20(5), pp. 610–619. doi: 10.1038/s41556-018-0088-1.

Zhang, H. and Lin, G. (2021) 'Microbial proteasomes as drug targets', *PLOS Pathogens*, 17(12), p. e1010058. doi: 10.1371/journal.ppat.1010058.

Zhang, W.-W. and Matlashewski, G. (2015) 'CRISPR-Cas9-mediated genome editing in *Leishmania donovani*', *mBio*, 6(4), pp. e00861-15. doi: 10.1128/mbio.00861-15.

Zhang, W., Lypaczewski, P. and Matlashewski, G. (2017) 'Optimized CRISPR-Cas9 genome editing for *Leishmania* and its use to target a multigene family, induce chromosomal translocation, and study DNA break repair mechanisms', *mSphere*. Edited by I. J. Blader, 2(1), pp. 1–15. doi: 10.1128/mSphere.00340-16.

Zheng, Q. *et al.* (2017) 'The RNA helicase DDX46 inhibits innate immunity by entrapping m 6 A-demethylated antiviral transcripts in the nucleus', *Nature Immunology*, 18(10), pp. 1094–1103. doi: 10.1038/ni.3830.

Zinoviev, A. and Shapira, M. (2012) ‘Evolutionary conservation and diversification of the translation initiation apparatus in trypanosomatids’, *Comparative and Functional Genomics*, 2012. doi: 10.1155/2012/813718.

Zoltner, M. *et al.* (2018) ‘Comparative proteomics of the two *T. brucei* PABPs suggests that PABP2 controls bulk mRNA’, *PLoS Neglected Tropical Diseases*, 12(7), pp. 1–18. doi: 10.1371/journal.pntd.0006679.

Zulfiqar, B., Shelper, T. B. and Avery, V. M. (2017) ‘Leishmaniasis drug discovery: recent progress and challenges in assay development’, *Drug Discovery Today*. Elsevier Ltd, 22(10), pp. 1516–1531. doi: 10.1016/j.drudis.2017.06.004.