Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
Of the Ruperto-Carola University of Heidelberg, Germany
For the degree of
Doctor of Natural Sciences

Presented by

MSc. Julius Mulindwa
Born in: Kampala, Uganda
Oral-examination: 2.12.2013

An approach to determine the Transcriptome of *T. b. rhodesiense* from Sleeping Sickness patients

Referees: Prof. Dr. Christine Clayton

Dr. Jan Korbel

Acknowledgments

I would like to acknowledge and extend my sincere gratitude to the following persons and organizations that have contributed invaluably to the success of this study:

Professor Christine Clayton for her mentorship, vital encouragement and supervision through the course of the study. Members of the Clayton Lab (2009 – 2013) including; Esteban, Abeer, Clementin, Doro, Connie, Diana, Daniela, Chaitali, Igor, Aditi, Elisha, Bhaskar, Claudia and Ute. You guys made the lab a home away from home. A special thank you to Abeer and Clementine for their hand in the bioinformatics analysis.

Dr. Jan Korbel and Dr. Georg Stoeklin, for their impeccable advice and guidance in ensuring that the study objectives were met. Mr David Ibberson at the Bioquant sequencing core facility for his assistance with library preparations and discussions about the evolving RNASeq technology.

Prof. John Enyaru and Ass. Prof Enock Matovu, for their support and initiating me into the procedures of carrying out field studies. Mr James Ochieng and the entire Laboratory and Medical staff of Lwala hospital for the support during the patient recruitment phase. Members of the Veterinary molecular biology lab and department of Biochemistry, Makerere University, for accommodating and supporting me through the entire field study period.

Michi and Hadi, thank you for the friendship and photography sessions. The Sempijja family, thank you for the love and care over the years.
Dr. Jacqueline Dryer and the whole Adequation team for assisting with the shipment of equipment and donation of Journals to Makerere University.

The DFG grant for infection biology, for the financial support. The HBIGS graduate program and ZMBH institute for facilitating the perfect academic ambience in the University of Heidelberg.

Dedication

To Grace, Joshua and Jerry, you are my inspiration.

Abstract

Human African Trypanosomiasis (HAT) is a neglected tropical disease that mainly affects the poorest people in sub Saharan Africa. HAT is caused by two subspecies of *Trypanosoma brucei*; *T. b. rhodesiense* is found in Eastern Africa and causes the acute form of the disease, while *T. b. gambiense* is found in Western Africa and causes the chronic form of the disease. With no reliable diagnostic screening test, available drugs being rather toxic, and emerging cases of drug resistant strains, research on the molecular aspects of the trypanosome is being carried out with the hope of identifying potential drug targets and diagnostic markers. Since most studies are carried out on cultured blood stream trypanosomes, the extent to which these parasites are representative of a real human infection is not known. Therefore the aim of this study was to analyze the transcriptome of clinical isolates of *T. b. rhodesiense* from patient peripheral blood and cerebral spinal fluid by high throughput sequencing. But given the low parasitaemia during active infection, I developed a splice leader priming based Polymerase chain reaction method to specifically amplify nanogram amounts of trypanosome total RNA in microgram amounts of Human cellular RNA, to an amount sufficient for sequencing.

The amplification method resulted in trypanosome transcripts covering 60% of the *T. brucei* 6772 unique genes, and with an expression threshold of 5 Rpkm. The sequenced amplified libraries (four replicates) were highly reproducible and comparable to unamplified libraries generated in the same way. However a comparison to the conventional RNASeq generated library showed distortions in the transcriptome, which could be corrected for and used to analyze clinical samples. An analysis of methods used to purify trypanosomes from blood showed that; even though DEAE chromatography and reticulocyte lysis resulted in 10 times more parasites than those in the buffy coat isolation method, reticulocyte lysis resulted in distortion of the transcriptome. However the DEAE chromatography trypanosomes transcriptome which was more comparable to the buffy coat, was not highly reproducible. The analysis of genomes from the trypanosomes isolated from the patients showed a level of heterogeneity between the samples with significant gene copy number variations observed mainly in the multi copy genes.

Zusammenfassung

Die humane afrikanische Trypanosomiasis (HAT) ist eine vernachlässigte Krankheit die hauptsächlich die ärmsten Leute in den subsaharischen Ländern Afrikas betrifft. HAT wird von zwei Unterarten von *Trypanosoma brucei* verursacht; *T. b. rhodesiense* kommt in Ostafrika vor und verursacht die akute Form der Krankheit, während *T. b. gambiense* in Westafrika vorkommt und die chronische Form der Krankheit verursacht. Da es keine verläßlichen diagnostischen Auslesetests gibt, die vorhandenen Medikamente ziemlich giftig sind und es neue Fälle von arzneimittelresistenten Stämmen gibt wird die Forschung der molekularen Aspekte der Trypanosome in der Hoffnung ausgeführt, neue Zielmoleküle und diagnostische Marker zu identifizieren. Da die meisten Studien mit kultivierten Blutstromtrypanosomen durchgeführt werden, ist nicht bekannt, in welchem Umfang diese Parasiten eine reale humane Infektion widerspiegeln. Daher ist das Ziel dieser Studie das Transkriptom von klinischen Isolaten aus dem peripheren Blut und der Zerebrospinalflüssigkeit von Patienten durch Hochdurchsatzsequenzierung zu analysieren. Aufgrund der niedrigen Parasitämie während einer aktiven Infektion habe ich eine PCR Methode entwickelt, die einen Primer für den „splice leader“ verwendet, um spezifisch trypanosomiale Gesamt-RNA im Nanogrammbereich aus Microgrammmengen humaner zellulärer RNA zu amplifizieren, um eine zur Sequenzierung ausreichende Menge zu erhalten.

Die Amplifizierungsmethode führte zu Trypanosomentranskripten, die 60% der 6772 einzigartigen *T. brucei* Gene abdecken, bei einer Expressionsschwelle von 5 Rpkm. Die sequenzierten amplifizierten Genbibliotheken (vier Replikate) waren sehr reproduzierbar und vergleichbar mit nicht-amplifizieren Bibliotheken, die auf dieselbe Weise hergestellt wurden. Aber ein Vergleich mit konventionellen RNASeq Ergebnissen zeigte Verzerrungen im Transkriptom, welche korrigiert werden konnten und für die Analyse von klinischen Proben verwendet werden konnten. Eine Analyse der Methoden für die Isolation von Trypanosomen aus Blut zeigte, dass - obwohl durch DEAE Chromatographie und Retikulozytenlyse zehnmal mehr Parasiten isoliert werden können als mit der Leukozytenfilmisolationsmethode - die Retikulozytenlyse zu einer Verzerrung des Transkriptoms führt. Das DEAE Chromatographie Trypanosomentranskriptom jedoch, welches eher mit dem Leukozytenfilm-transkriptom vergleichbar war, war nicht sehr reproduzierbar. Die Analyse der Genome der Trypanosomen, welche von den Patienten isoliert wurden, zeigte ein hohes Ausmaß an Heterogenität zwischen den Proben mit signifikanten Variationen der Genkopien, welche hauptsächlich bei den mehrfach kopierten Genen beobachtet wurden.

Table of Contents

1 Introduction

1.1 Human African Trypanosomiasis (HAT)

African trypanosomes are unicellular flagellated parasites, belonging to the genus *Trypanosoma,* family *Trypanosomatidae* and order *Kinetoplastida* (Stevens and Brisse, 2004). They are called Kinetoplastids due to their peculiar mitochondrial DNA known as kinetoplast DNA (kDNA); which is unique in its structure, function, mode of replication and is visible by Giemsa staining (Liu et al., 2005; Renger and Wolstenholme, 1971; Shapiro and Englund, 1995; Simpson, 1973).

African trypanosomiasis is caused by three subspecies of *Trypanosoma brucei*, *T. b. brucei, T. b. gambiense* and *T. b. rhodesiense*. These parasites are transmitted by tsetse flies of the genus *Glossina* that infest vast areas of sub Saharan Africa (Krafsur, 2009). *T. b. brucei* infects wild and domestic animals causing Nagana in live stock (Connor, 1994). It is not pathogenic to humans and this is mainly due to presence of the trypanolytic factor (TLF) in normal human plasma, which kills the parasite. However *T. b. gambiense* and *T. b. rhodesiense* are resistant to this trypanolytic factor (Vanhamme and Pays, 2004). There are two types of TLFs; TLF-1, which is a high-density lipoprotein particle, (Rifkin, 1978) and TLF-2, which is a serum protein binding complex (Raper et al., 1999). Both of these TLFs contain haptoglobin related protein, HPR, and apolipoprotein LI, APOL1 (Pays et al., 2006; Vanhollebeke and Pays, 2010). *T. b. rhodesiense* encodes the serum resistance associated protein (SRA), which binds TLF-1 making it resistant to lysis (De Greef and Hamers, 1994; Xong et al., 1998). On the other hand, *T. b. gambiense* has no SRA gene and is resistant to TLFs by a mechanism that involves the *T. b. gambiense* specific glycoprotein (TgsGP) (Berberof et al., 2001), which prevents the trypanosome killing effects of APOL1 (Capewell et al., 2011; Kieft et al., 2012; Uzureau et al., 2013). This resistance of *T. b. rhodesiense* and *T. b. gambiense* to lysis by the trypanolytic factor culminates in human pathology resulting in the disease, Sleeping sickness.

1.1.1 Epidemiology of HAT

Human African Trypanosomiasis (HAT), also known as Sleeping sickness, is caused by two sub-species of *T. brucei*; *T. b. gambiense* causes a chronic form of the disease in Western and Central Africa and is transmitted to humans by the *Glossina palpalis or G. fuscipes* tsetse flies; while *T. b. rhodesiense* is responsible for the more acute form of the disease in Eastern and Southern Africa and is transmitted

by the *Glossina moritans* or *G. fuscipes* groups, Fig 1.1 (Brun et al., 2010; Simarro et al., 2010). However, Uganda remains the only country with active foci of both Gambian and Rhodesian sleeping sickness (Welburn and Odiit, 2002). It also represents a region of potential overlap, with fears that the two disease foci could eventually merge (Picozzi et al., 2005). Interestingly, a few sporadic cases of human trypanosomiasis being caused by non human-pathogenic trypanosome species have been reported. These species were *T. b. brucei* (Deborggraeve et al., 2008), *T. congolense* (Truc et al., 1998) and *T. evansi* (Joshi et al., 2005).

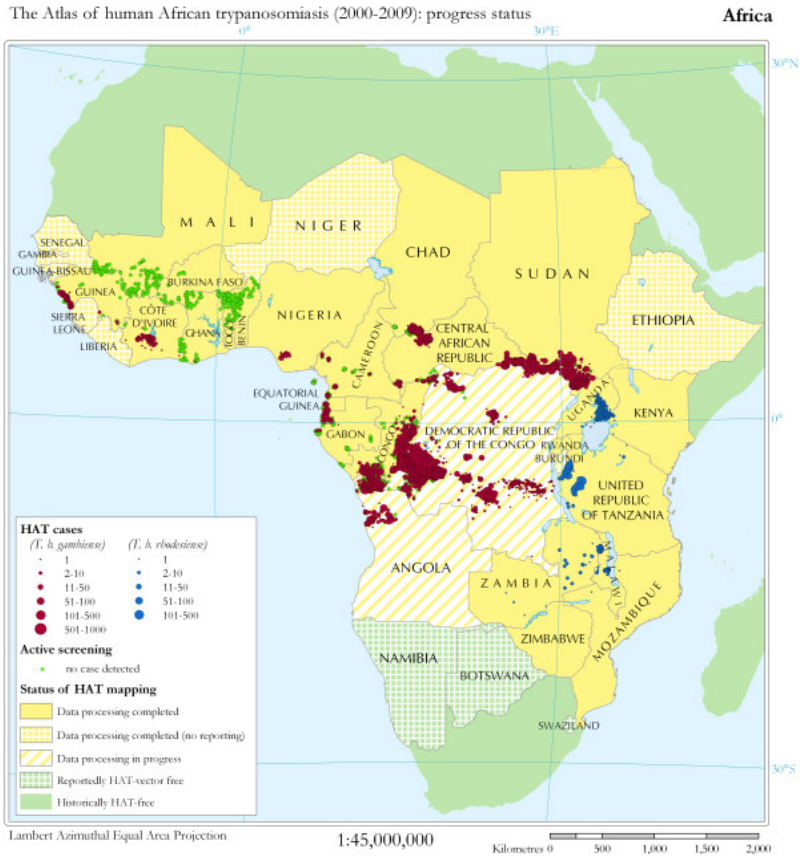


Figure 1.1 Map showing the spatial distribution of Human African Trypanosomiasis on African continent. Uganda represents the only country with active foci for both Gambian and Rhodesian sleeping sickness (Map directly copied from Simarro et al., 2010).

Sleeping sickness mainly affects the poorest people in the rural regions of Africa. Despite being a disabling and life threatening fatal disease, it is still one of the most neglected (Balasegaram et al., 2008; Simarro et al., 2012a). If incidence alone is to be considered, the disease appears to be a minor problem compared to other tropical diseases. However, when disability-adjusted life years are calculated, also considering its severity (untreated HAT results in 100% mortality), HAT ranks third behind malaria and schistosomiasis in sub-Saharan Africa (Cattand, 2001).

In spite of the socio-economic unrest that disrupted disease surveillance (mainly wars), the prevalence of HAT has changed over the years mainly because of control and intervention programs including, early diagnosis, treatment, and vector control (Brun et al., 2010; Kuzoe, 1993). The number of people at risk of *T. b. rhodesiense* is not known precisely however in 2006, 486 cases were reported and this accounts for 3% of the continental HAT cases (Simarro et al., 2008). The *T. b. gambiense* disease has a much more protracted course than the *T. b. rhodesiense* disease (Brun et al., 2010) and is therefore responsible for more than 95% of HAT cases. By 2007 the number of new HAT cases reported in Africa was 10,769 (WHO, 2012).

1.1.2 Clinical Presentation of HAT

T. b. rhodesiense and *T. b. gambiense* cause a clinically similar disease, the major difference between them being their virulence (Dumas et al., 1985). Infection progress to disease is usually acute in *T. b. rhodesiense* with an incubation period of about 2-3 weeks (Odiit et al., 1997), whereas, *T. b. gambiense* infection is usually a slower chronic infection with an incubation period of several weeks to months (Checchi et al., 2008). Following the bite of a trypanosome-infected tsetse fly, the trypanosomes differentiate to the bloodstream stage and spread via the local draining lymph node into the vascular system. In some but not all HAT patients, a local skin reaction or chancre (local inflammatory response) occurs at the site of inoculation (Fairbairn and Godfrey, 1957; Naessens et al., 2003).

There are two recognized stages in the clinical presentation of HAT, the early hemolymphatic stage, and the late encephalitic stage when the parasites cross the blood brain barrier and colonize the central nervous system (Despommier et al., 2005; Kennedy, 2004). However, the transition from the early to the late stage is not always distinct in *T. b. rhodesiense* infections (Atouguia and Kennedy, 2000). The onset of the early (hemolymphatic) stage is variable but usually occurs 1–3 weeks after the bite, with episodes of intermittent fever lasting 1–7 days. During this period the parasites proliferate within the blood and lymphatic system and the leading early signs and symptoms of this first stage are; chronic and intermittent fever, headache, pruritus (itchy rash), arthralgia (joint pains), generalized weakness, and weight loss (Apted, 1970). Multiple organs including the spleen, liver, skin, cardiovascular system, endocrine system, and eyes may then be infected; and this involvement usually underlies the wide spectrum of systemic dysfunction that may occur (Atouguia and Kennedy, 2000; Duggan and Hutchinson, 1966).

The late (encephalitic) stage of the disease begins after weeks in *T. b. rhodesiense* infection and months in *T. b. gambiense* infection. This is when the trypanosomes cross the blood-brain barrier and invade the central nervous system (Greenwood and Whittle, 1980; Masocha et al., 2004; Mulenga et al., 2001). This stage results in a chronic encephalopathy, which is associated with headache and mental changes. These neurologic features can be grouped into general categories such as psychiatric, motor and sensory abnormalities, and sleep disturbances (Atouguia and Kennedy, 2000; Duggan and Hutchinson, 1966). If left untreated at this point, the disease progresses with severe body wasting to coma and eventually leads to death.

1.1.3 Diagnosis of HAT

T. b. rhodesiense and *T. b. gambiense* parasites are morphologically inseparable but they can be distinguished genetically. The human serum resistance-associated (SRA) gene defines *T. b. rhodesiense* (Gibson et al., 2002; Njiru et al., 2004) whereas a receptor-like flagella pocket glycoprotein (TgsGP) is specific to *T. b. gambiense* (Berberof et al., 2001). African trypanosomes are extra cellular parasites that live in blood, lymph and can also cross the blood brain barrier into the cerebral spinal fluid. Therefore for the diagnosis of trypanosomal infections in humans, it is often essential to demonstrate trypanosomes in lymph node aspirate, blood, or cerebrospinal fluid. In this respect, HAT diagnosis follows a three-step procedure which involves screening, diagnostic confirmation and staging of the disease (Chappuis et al., 2005).

In ***T. b. gambiense* endemic areas**, the Card Agglutination Test for Trypanosomiasis (CATT) is a rapid serological test that is used in mass screening (Magnus et al., 1978b). It is based on an agglutination assay consisting of lyophilized bloodstream form *T. b. gambiense* expressing the variable antigen type LiTat 1.3, which detects specific antibodies in blood, plasma or serum of patients. The reported sensitivity of the CATT on undiluted whole blood (CATT-wb) varies from 87 to 98% (Noireau et al., 1987; Robays et al., 2004; Truc et al., 2002). However, false-negative CATT results have been reported (Penchenier et al., 1991), and this is due to patients infected with *T. b. gambiense* strains that lack or do not express the LiTat 1.3 gene (Dukes et al., 1992; Enyaru et al., 1998). For such cases, the LATEX agglutination test, a field alternative to the CATT that combines three purified variable surface antigens (LiTat 1.3, 1.5, and 1.6), was developed (Büscher et al., 1999; Penchenier et al., 2003). Other highly sensitive serological tests, such as immunofluorescence antibody test, IFAT (Magnus et al., 1978a), and enzyme-linked immuno sorbent assays, ELISA (Lejon et al., 1998), were developed but their use is

limited to reference laboratories and not routine field diagnosis.

For all HAT patients, there is a need for parasitological confirmation demonstrating the presence of trypanosomes in any body fluid or tissue. This is often carried out by microscopic examination of chancre aspirate, lymph node aspirate, wet and thick blood films, and/or cerebral spinal fluid (Henry et al., 1981; Simarro et al., 2003; Van Meirvenne, 1999). But given the low sensitivity of these techniques, concentration methods such as the micro-haematocrit centrifugation (Woo, 1970), Quantitative buffy coat (Ancelle et al., 1997), miniature anion-exchange centrifugation (Lumsden et al., 1979) and fluorescence microscopy with acridine orange staining (Biéler et al., 2012), were developed. However these procedures are not routinely used in field diagnosis because they are time consuming and use sophisticated equipment. Due to the absence of sufficiently specific clinical signs and blood tests indicating the progress from first to second-stage HAT, staging of patients still relies on examination of cerebral spinal fluid (CSF) obtained by lumbar puncture (Kennedy, 2008). According to the World Health Organization recommendations (WHO, 1998), second stage HAT is defined by the presence of trypanosomes and/or >5 white blood cells/µL in the CSF or increased protein content (>370mg/L). In addition, a high IgM concentration in CSF is considered a reliable marker of neurological involvement (Greenwood and Whittle, 1980; Lejon et al., 2002b). This then distinguishes the hemolymphatic stage from the encephalitic stage, and is therefore the basis for the choice of drug treatment (Rodgers, 2009); since only few trypanocides are able to cross the blood brain barrier.

In ***T. b. rhodesiense*** **endemic areas**, there is no CATT equivalent test and therefore field screening relies on clinical signs and symptoms. The diagnostic confirmation and staging of *T. b. rhodesiense* HAT is based on the same methods as described for *T. b. gambiense* disease. However parasitological confirmation of *T. b. rhodesiense* infection is relatively easy because of the high parasiteamia, and so most patients are diagnosed by examination of a stained thin or thick blood smear (Chappuis et al., 2005; Odiit et al., 1997). The major drawback to parasite detection is the fact that sensitivity is quite low; for example, direct microscopy of thick blood films has a detection threshold of 5000 trypanosomes per ml of blood. This limit may miss the diagnosis in up to 50% of the infections (Chappuis et al., 2005). Hence, a number of research efforts have been focused on developing new diagnostic tools including; recombinant or native trypanosome antigens that could be used to develop an improved serological test (Hutchinson et al., 2004; Manful et al., 2010), methods to detect parasite antigens in blood or CSF (Papadopoulos et al., 2004), proteomic fingerprinting (Agranoff et al., 2005), loop-mediated isothermal amplification, (Kuboki

et al., 2003; Njiru et al., 2008). But recently, an initiative by World Health Organization (WHO) and the Foundation for Innovative New Diagnostics (FIND) was launched to develop methods for diagnosis and staging of HAT.

1.1.4 Treatment of HAT

The phenomenon of antigenic variation leaves only a slim hope for a vaccine against human African trypanosomiasis (Magez et al., 2010). Disease control also relies on vector control by the use of tsetse fly traps, insecticides and sterile insect technique (SIT), thus reducing the fly populations. However, these initiatives have been difficult to sustain in affected areas and their effectiveness has been compromised in the long-run (Aksoy, 2003). With the limitations from this form of intervention, chemotherapy offers an effective solution. The current chemotherapy for human sleeping sickness is unreliable, with all the main drugs used for early stage (Suramin, Pentamidine) and late-stage disease (Melarsoprol, Eflornithine) being unavailable orally, with a degree of ineffectiveness and sometimes toxic to the patients (Croft et al., 2005; Fairlamb, 2003).

Early stage HAT treatment

The early-stage *T. b. gambiense* HAT is treated with pentamidine, which is usually administered by the intramuscular route (Pepin and Milord, 1994). Its major side effects include, damage to the liver, kidneys and the pancreas, the latter leading to diabetes (Fairlamb, 2003). Early stage *T. b. rhodesiense* HAT is treated with intravenous suramin, which is usually effective especially when given the early stages of infection, but can result in potential life-threatening side-effects which include collapse with nausea, vomiting and shock. Other severe delayed reactions may include kidney damage (particularly in malnourished patients), exfoliative dermatitis, agranulocytosis, haemolytic anaemia, jaundice and severe diarrhoea, all of which can be fatal (Fairlamb, 2003; Fairlamb and Bowman, 1980).

Late stage HAT treatment

The late stage of HAT, characterized by trypanosomes crossing the blood brain barrier into the central nervous system, is more problematic to treat because the drugs used in this case are more toxic (Atouguia and Kennedy, 2000; Kennedy, 2004). The only drug available for treatment of late-stage *T. b. rhodesiense* HAT is melarsoprol. It is an organic arsenical that acts on trypanothione, a trypanosomal molecule that maintains an intracellular reducing environment (Fairlamb et al., 1989). However, melarsoprol also binds to other sulfhydryl-containing agents in the cell, and

its uptake into African trypanosomes has been attributed to the P2 purine nucleoside transporter (Carter et al., 1995; Fairlamb and Cerami, 1992). Melarsoprol is administered intravenously in propylene glycol because it has a very low solubility in water and the injections are extremely painful for the patients (Kennedy, 2012; Kuepfer et al., 2012). But even though melarsoprol is usually effective, treatment failures due to drug resistance have been reported (Legros et al., 1999). Melarsoprol is extremely toxic producing a post-treatment reactive encephalopathy in 10% of patients, half of whom die, leading to an overall mortality from treatment of about 5% (Kennedy, 2008; Pepin and Milord, 1994). The Late-stage *T. b. gambiense* HAT is currently being treated with nifurtimox–eflornithine combination therapy, NECT (Simarro et al., 2012b), although it was previously under monotherapy with eflornithine (Burri and Brun, 2003). Interestingly, the mortality from combination therapy was reduced to 0.7% (one of 143) as compared to 2.1% (three of 144) for eflornithine monotherapy (Priotto et al., 2009). However, eflornithine does have many side effects including bone marrow toxicity, alopecia, seizures, and gastrointestinal symptoms (Burri and Brun, 2003).

Potential new drugs for treatment of HAT

All the drugs highlighted above have been in use for over 50 years and they are deemed undesirable in terms of administration, toxicity and reported resistance (Brun et al., 2010). However the recent development of new compounds that have entered clinical trials, offer hope in new drugs to treat HAT (Barrett, 2010; Brun et al., 2011). For the late stage *T. b. gambiense* HAT, the nitroheterocyclic drug fexinidazole is being tested as an oral treatment for it was found to be both effective and non-toxic in animal models of sleeping sickness. It is yet to undergo phase 2 clinical trials (Kaiser et al., 2011; Torreele et al., 2010). In addition is a new orally active compound benzoxaborole, which was able to cure CNS stage trypanosomal infections in a murine model, and is now under phase 1 trial (Jacobs et al., 2011a; Jacobs et al., 2011b). For the case of *T. b. rhodesiense* HAT, an improvement has been made in the delivery of mode of melarsoprol through melarsoprol-cyclodextrin inclusion complexes, making the drug more soluble in water. When administered orally, these complexes were shown to be effective in curing CNS trypanosomiasis in a mouse model (Kennedy, 2012; Rodgers et al., 2011).

1.1.5 Immunobiology of HAT

The understanding of the immunological events that occur in HAT is essential in identifying the precise role of inflammatory responses in the pathophysiology of the infection. However, various animal models have provided conflicting evidence regarding the immunologic factors that influence the magnitude of resistance to infection (Namangala et al., 2001). But, as a result of African trypanosome adaptation to growth in experimental mouse models, much of the knowledge of the biology of these parasites comes from such studies (Sternberg, 2004). The overall host response to the disease requires the contribution of VSG-specific B and T cell responses (Reinitz and Mansfield, 1990; Schleifer et al., 1993) and the macrophage/monocyte phagocyte system (Dempsey and Mansfield, 1983; Vincendeau et al., 1999) to resolve the infection. In addition, studies suggest that cytokine responses influence the outcome of African trypanosomiasis (Hertz et al., 1998; Namangala et al., 2000; Uzonna et al., 1999). There are also reported cases of asymptomatic carriers and spontaneous cure among *T. b. gambiense* HAT patients (Bucheton et al., 2011; Jamonneau et al., 2012).

Antigenic variation and antibody responses

The primary immune evasion strategy by African trypanosomes is by antigenic variation (Pays et al., 2004; Vanhamme et al., 2001). Trypanosomes are covered with a coat of variant surface glycoproteins, VSG, which are immunodominant antigens that elicit both T-cell dependent and independent B-cell responses, depending on their conformation (Mansfield, 1994). Antibody opsonized trypanosomes are effectively cleared presumably by the host's lymphoreticular system (Macaskill et al., 1981). However, the trypanosomes are able to undergo antigenic variation due to the genomic presence of a repertoire of more than 1000 transcriptionally inactive VSG genes under a single active promoter (Pays et al., 2004), which effects continuous stochastic switching of VSG genes (Taylor and Rudenko, 2006). This then enables the trypanosome to maintain a state of chronic infection in the host. The humoral response to VSG also has an immune-pathological effect whereby, the VSG elicits polyclonal B-cell activation and during human infection, results in the generation of auto-antibodies (Kazyumba et al., 1986) and immune complex disease (Lambert et al., 1981). Furthermore, during the meningoencephalitic stage of infection, trypanosome-specific IgG, IgM and polyclonal IgM responses have been detected in the cerebrospinal fluid (Lejon et al., 1998). These may be derived from modified plasma cells known as Mott or morular

cells in the white matter and also plasma cells which form perivascular infiltrates in the brain (Kennedy, 2004).

Cytokine activation

Following the infection by African trypanosomes, the early innate response is triggered leading to the release of pro-inflammatory mediators including; cytokines, interleukins IL-1, IL-6, IL-12 and tumour necrosis factor alpha (TNF-α) as well as nitric oxide (MacLean et al., 2001a; Magez et al., 1997; Tachado and Schofield, 1994). In a study of *T. b. rhodesiense* patients in Uganda, both early and late stage infections were characterized by elevated levels of IFN-γ, TNF-α and IL-10; although IFN-γ levels did diminish in the late-stage cases (MacLean et al., 1999; MacLean et al., 2001a). In *T. b. gambiense* patients, one study demonstrated increased plasma IL-10, IL-6 and IL-8, but no detectable IFN-γ or TNF-α (Lejon et al., 2002a). However, other studies detected elevated TNF-α levels in *T. b. gambiense* patients, which correlated with the severity of disease (Okomo-Assoumou et al., 1995; Rhind et al., 1997). The infection-associated production and systemic presence of TNF is possibly related to fever induction and to the severity of trypanosomosis induced pathology (Magez et al., 2002). It was indeed observed that the inflammation in the brain and occurrence of the neurological signs during the late-stage meningoencephalitic phase of the infections, were both TNF related (Okomo-Assoumou et al., 1995). In addition, the high incidence of mortality following anti-trypanosomosis treatment is associated with TNF (Lucas et al., 1993). The early release of these pro-inflammatory factors enables the early acquired immune responses to the parasite variant surface glycoprotein (Magez et al., 2002). The early release of IL-12 promotes T helper (Th) cells to release cytokines, particularly interferon-gamma (IFN-γ), which is responsible for host resistance against parasites spreading throughout host tissues (MacLean et al., 2001a). This early resistance is, however, superseded by later immune responses that are associated with the spread of trypanosomes to the CNS. In this subsequent chronic phase of infection, immunosuppression is associated with cytokines, such as IL-10, that down-regulate the inflammatory response in the early phase (Taylor et al., 1998; Uzonna et al., 1998).

1.2 Gene expression in trypanosomes

1.2.1 Genome organization and transcription

The chromatin in the trypanosome nucleus is arranged in nucleosomes for which both conserved and divergent histones have been identified (Hecker et al., 1995). The genomes of *T. b. brucei* (Berriman et al., 2005) and *T. b. gambiense* (Jackson et al., 2010) have been sequenced and annotated. *T. brucei* has a diploid genome with an estimated haploid DNA content of around 3.5 - 4 *x* 10^7 base pairs (Berriman et al., 2005). The nuclear genome consists of three chromosome classes, megabase, intermediate and mini chromosomes. The mega base chromosomes of this diploid genome contain the expressed genes and form at least 11 homologous chromosome pairs, Fig 1.2 (Daniels et al., 2010). In addition, strain dependent chromosomal length polymorphisms exist for this chromosomal set (El-Sayed et al., 2000; Ersfeld et al., 1999). The genome of *T. brucei* contains approximately 9,068 predicted genes including 900 pseudogenes and 1700 genes that are specific to *T. brucei* (Berriman et al., 2005). In trypanosomes, the majority of protein coding genes are organized in polycistronic units that are transcribed by RNA polymerase II (Palenchar and Bellofatto, 2006). These polycistronic RNA precursors are then processed into mature monocistronic mRNAs by trans-splicing of a 39nt leader sequence to the 5' end and polyadenylation of the 3' end (Liang et al., 2003). However, most genes in the polycistronic unit are not functionally related, and gene regulation occurs mainly at the post-transcriptional level operating on mRNA processing and mRNA degradation (Clayton and Shapira, 2007). The only known exceptions are procyclin and *VSG* genes, which are regulated at the transcriptional level. These genes are transcribed by RNA polymerase I (Gunzl et al., 2003), and the transcription rate is thought to be developmentally controlled through epigenetic regulation (Figueiredo et al., 2009; Glover and Horn, 2006; Siegel et al., 2009). Unfortunately, there is little known about RNA polymerase II promoters in trypanosomes, and the only one that is well characterized is the promoter for the spliced leader RNA transcription units (Das et al., 2005; Luo et al., 1999).

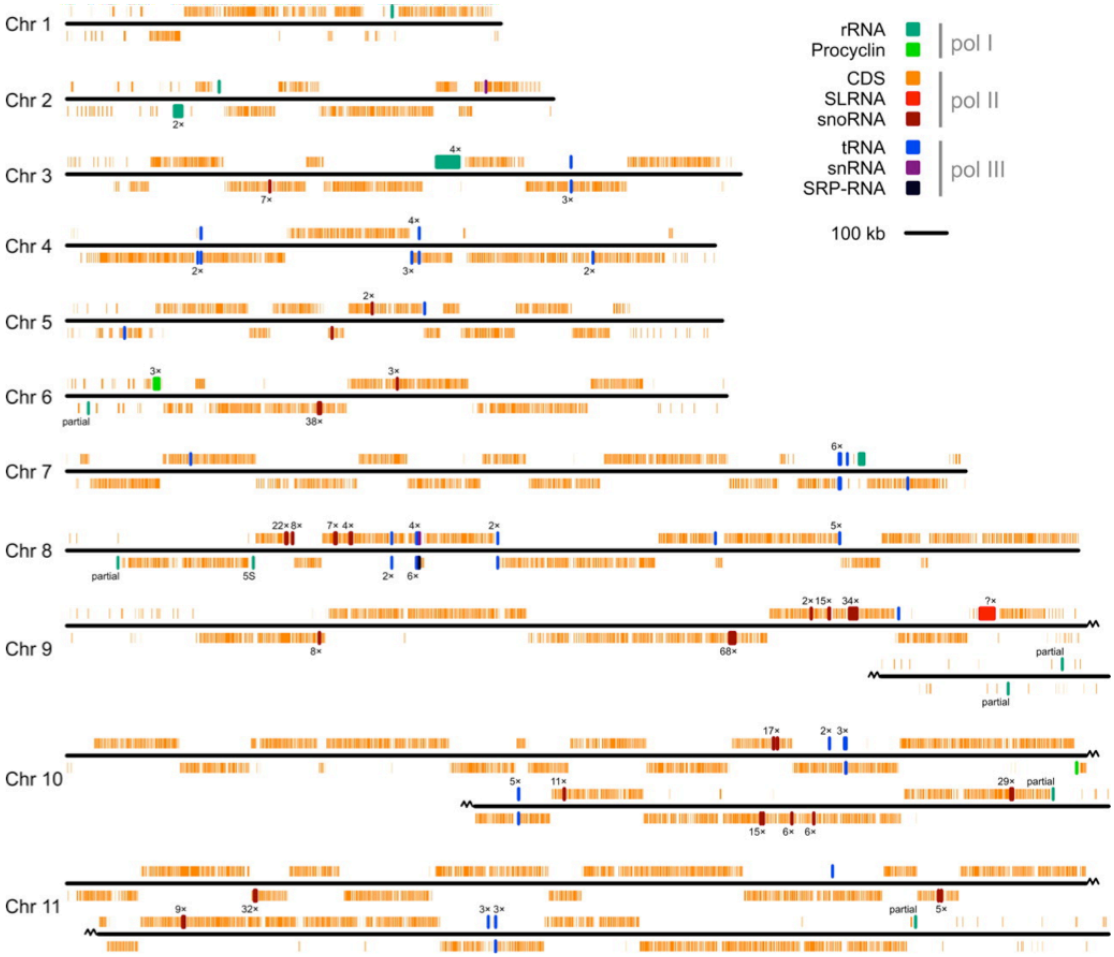


Figure 1.2 Schematic representation of the *T. brucei* genome**.** Chromosomal organization of genes according to their class and transcribing polymerase. The colored bars indicate the positions and lengths of different genetic elements relative to the chromosome backbone (black line). The bars above the lines indicate transcription toward the right; bars below the lines indicate transcription toward the left. When a number of a similar elements are present in close proximity in the genome, this is indicated by nX next to bar (where n is the number of elements). Noncoding RNA genes have been given a minimum bar length to facilitate visualization. Only the largest assembled contig is represented for each chromosome. Chromosomes 9 to 11 have been split across two lines. (Directly copied from Daniels et al., 2010).

1.2.2 Transcriptomic analysis of gene expression

The transcriptome is the complete set of transcripts (genetic code transcribed to RNA molecules) in a cell, for a specific developmental stage or physiological condition. It is also defined as the polyadenylated products of RNA polymerase II activity (Tang et al., 2010; Wang et al., 2009). Even though within an organism or tissue all cells have the same genome, diverse phenotypes exist because of varying types and amounts of mRNA transcripts (Kalisky et al., 2011; Tang et al., 2011). Therefore transcriptomic analyses (of mRNAs, non coding RNAs and small RNAs), enable understanding of the transcriptional structure of genes in terms of their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications. This is in turn relevant in quantifying the changing expression levels of each

transcript during development and under different stimuli or environmental conditions (Wang et al., 2009).

By using high throughput microarrays to study the transcriptome of trypanosomes, it was observed that between 2 to 10% of all mRNA in trypanosomes were developmentally regulated (Jensen et al., 2009; Kabani et al., 2009; Queiroz et al., 2009). However, these hybridisation based microarrays do suffer from background hybridization, limited accuracy of expression for transcripts in low abundance, and cannot be used to detect splice variants or unknown genes (due to reliance upon existing knowledge about genome sequence) (Okoniewski and Miller, 2006; Royce et al., 2007). The drawbacks of microarrays have been circumvented by the recent development of high-throughput DNA sequencing platforms such as Roche/454 (http://www.454.com), Illumina (http://www.illumina.com), Applied Biosystems/Life Technologies (http://www.appliedbiosystems.com), and Helicos (http://www.helicosbio.com) (Fu et al., 2009; Marioni et al., 2008; Metzker, 2010). These next generation sequencing (NGS) methods have provided a new approach for both mapping and quantifying transcriptomes (Mardis, 2008; Ozsolak and Milos, 2011; Wang et al., 2009).

An analysis of the transcriptome of *Trypanosoma brucei* using the next generation sequencing platforms revealed a number of dynamic properties of the gene expression profile of this parasite (Kolev et al., 2010; Nilsson et al., 2010; Siegel et al., 2010; Veitch et al., 2010). For instance, RNA-seq analysis of the trans splicing junctions (coverage of the splice leader and 5' end of the mRNA) enabled the 5' UTR regions for 8,960 transcripts to be mapped. Furthermore, this approach was able to identify polyadenylation sites (PAS) for between 2,081 and 5,948 genes (Kolev et al., 2010; Siegel et al., 2010). In addition, by using the splice leader trapping method (Nilsson et al., 2010), the splice acceptor sites (SAS) for 8,277 genes in the long slender bloodstream forms were identified. These UTRs are important determinants of transcript stability through regulation by *cis*-acting elements (Clayton and Shapira, 2007; Fernández-Moya and Estévez, 2010; Ouellette and Papadopoulou, 2009). RNASeq analysis of the global mRNA degradation rates in bloodstream form trypanosomes showed that most mRNAs had half lives less than 20 min (Manful et al., 2011).

1.2.3 Transcriptome analysis of nanogram amounts of RNA

High through put expression analysis has provided new insights into the biology of trypanosomes (Siegel et al., 2011). However, these studies required relatively large amounts total RNA (typically 10–30 µg of total RNA), which was

obtained from *in vitro* cultured trypanosomes. It is important to extend these analytical methods to much smaller quantities of trypanosomes, such as those found in infected clinical specimens as well as tsetse flies. But most biological specimens (for example blood) directly obtained ex *vivo* in a clinical setting (mainly for diagnostic purposes) don't often yield enough RNA for high throughput gene expression analysis (Marteau et al., 2005; Mitsuhashi et al., 2006). However strategies have been developed to amplify low amounts of total RNA (less than 100ng) to an amount sufficient for gene profiling analysis (Pan et al., 2013; Tang et al., 2011; Tariq et al., 2011; Wang, 2005).

There are two principle methods of amplification of mRNA without skewing relative transcript abundance. First is, linear amplification of cDNA by *in vitro* transcription (IVT) using a DNA dependent RNA polymerase (Stirewalt et al., 2004; Van Gelder et al., 1990) and secondly, exponential amplification of cDNA by DNA polymerase chain reaction, PCR (Brady et al., 1990; Iscove et al., 2002). The combination of PCR amplification to generate sufficient ds-cDNA template followed by *in vitro* transcription has also been used to amplify RNA from minimal stating material (Klur et al., 2004). These methods have been extensively used to study single mammalian cell transcriptomes (Tang et al., 2011). A single mammalian cell contains approximately 20 to 40 pg of total RNA of which 0.5 to 1.0 pg is mRNA (Roozemond, 1976; Uemura, 1980). For standard microarray analysis, this needs to be amplified around ten million–fold (Livesey, 2003). The merit of the PCR strategy is the exponential amplification of cDNAs so that single-cell cDNAs can be amplified millions-fold in several hours. Its disadvantage is the accumulation of primer dimers and other nonspecific byproducts during amplification, especially during later cycles of PCR (Brady et al., 1990; Kurimoto et al., 2006; Livesey, 2003). The advantage of the IVT strategy is its stringent specificity while reducing accumulation of nonspecific byproducts (Eberwine et al., 1992); its disadvantage is that cRNAs typically less than 1Kb are generated. The IVT procedure is also more tedious and time-consuming, and every round of IVT can amplify the cDNAs only up to 1,000-fold (Kawasaki, 2004; Livesey, 2003). By modifying and improving the amplification PCR strategy by (Kurimoto et al., 2007), mRNA-seq on a single cell was carried out and it showed a 64% coverage of expressed genes (Tang et al., 2010; Tang et al., 2009). In this study, the exponential PCR amplification strategy was used to develop a protocol to specifically amplify trypanosome RNA in clinical samples.

1.3 Aim of the study

1.3.1 Study objectives

This study was aimed at analyzing for the differential gene expression of *T. b. rhodesiense* in the peripheral blood and cerebral spinal fluid of Sleeping sickness patients.

The specific objectives involved:

1. Optimizing a method to analyze the transcriptome of trypanosomes from clinical samples.
2. Analysis of the effect of different methods of isolation of trypanosomes from blood on there transcriptomes.
3. Analysis of transcriptomes of trypanosomes isolated from HAT patient peripheral blood and cerebral spinal fluid.
4. Analysis of structural variations within the genomes of trypanosomes isolated from the patients.

1.3.2 Significance of study

Most of the expression-profiling studies on African trypanosomes have been performed on cultured parasites or high-density mouse infections. The extent to which these parasites are representative of a real human infection is still not known. Therefore this study was aimed at high throughput sequencing analysis of the transcriptome of *T. b. rhodesiense* obtained from HAT patients. Analysis of the differential gene expression of *T. b. rhodesiense* in the bloodstream and cerebral spinal fluid would provide an insight into the human-trypanosome interaction, which would thus enable in the identification of possible targets for chemotherapy for the safe treatment of early and late stages of Sleeping sickness.

2 Materials and Methods

Parts of this section are derived from a manuscript that has been submitted for publication and has therefore been written or modified by other people (Fadda, A and Clayton, C.)

2.1 Clinical sample collection

2.1.1 Ethical clearance

The ethical approvals of protocols were obtained from the Ministry of Health, Uganda and Ethics committee of University of Heidelberg, Germany. All patients recruited into this study received written and verbal information explaining the purpose of the study and they gave informed consent. The ethical consent forms were written in English and translated into the local languages. For the children and adolescent participants (below 18 years), the parents or guardians gave the informed consent on their behalf. The patient recruitment was carried out at Lwala hospital in Kaberamaido district, which is situated in the *T. b. rhodesiense* HAT foci of Northeastern Uganda. Diagnosis of HAT was carried out by microscopic detection of motile trypanosomes on a wet blood smear of finger prick blood. For those trypanosome positive cases, staging was carried out by microscopic examination of CSF obtained by lumbar puncture. Diagnosis was carried out on admission and on discharge as part of the normal clinical follow-up procedure. The clinical officers at the out patients department (OPD) carried out the medical examination of all patients recruited into the study. An experienced senior nurse or laboratory technician carried out the phlebotomy and lumbar puncture procedures. I trained a laboratory technologist, James Ochieng, on the sample processing methodology including parasite isolation, counting, and storage. He was based at the hospital and carried out the sample collection.

2.1.2 Peripheral blood collection

Peripheral blood, PB (3-4ml) from confirmed HAT patients, was collected into an EDTA tube (BD Vacutainer) upon phlebotomy. An aliquot of the whole blood (600µl) was cryopreserved, a drop (10µl) was spotted on Whatman FTA and the remaining volume centrifuged at 3000rpm for 10min at room temperature. The plasma was aspirated off for storage and the buffy coat layer (approx. 500 µl) transferred to a 1.5ml eppendoff tube. Drew 10µl onto a nuebaur haemocytometer for a cell count (trypanosomes and WBC). The buffy coat volume was then centrifuged at 6000rpm for 2min at room temperature and the pellet fraction resuspended in 1ml of Trifast

reagent (Peqlab, GmbH), then transferred to cryotube for immediate freezing in liquid nitrogen.

2.1.3 Cerebral spinal fluid collection

Cerebral spinal fluid, CSF (2-4ml) obtained from confirmed HAT patients by lumbar puncture, was centrifuged at 3000rpm for 10min. The supernatant was aspirated off leaving approximately 500µl, which was transferred to a 1.5ml eppendoff tube. Then drew 10µl for the cell count, and centrifuged the remaining volume at 6000rpm for 2min at room temperature. The pellet fraction was then resuspended in 1ml Trifast (Peqlab, GmbH), transferred to a cryotube and immediately frozen in liquid nitrogen.

2.1.4 Haemocytometer count

The total cell count was determined microscopically by drawing 10µl of either buffy coat or CSF sample onto a neubauer counting chamber (Fig 2.1) as described above. The total numbers of cells (WBCs and Trypanosomes), from the 16-square-fields were counted, and the average from the 4 big squares (containing the 16 squares) was multiplied by $1x10^4$; which equated to the cell density in 1ml volume.

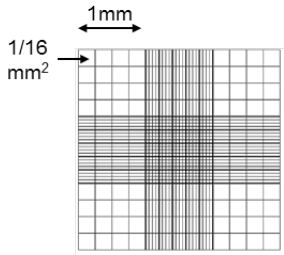


Figure 2.1 Representation of the counting chamber of the neubauer hemocytometer.

2.1.5 Speciation PCR on FTA blood

In order to confirm that all the cases were *T. b. rhodesiense* infections, PCR was carried out on the SRA gene as described (Radwanska et al., 2002). The PCR was carried out on DNA extracted from the patient blood spotted on Whatman FTA. Briefly, the FTA spots were punched with a sterile 7.9mm diameter disc punch (Ted Pella Inc.) into a 1.5ml eppendof tube. The discs were then rinsed in 500µl of distilled water by vortexing 3 times for 3 seconds. The water was drawn off and fresh 100 µl of water was pipetted into the tubes. This was then incubated at 95°C for 30min. 3µl of this was used as a template for PCR amplification of the SRA gene. The DNA templates were amplified in 50µl of PCR mixture containing 1X PCR buffer, 1.5 mM $MgCl_2$, 100µM of each of the four dNTPs, 1µM of each primer, and 2 units Taq DNA polymerase. The nested PCR was carried out using two sets of primers under the same conditions. The first primer set (SRA-1_s, 5'-CCTGATAAAACAAGTATCG GCAGCAA-3', & SRA-1_as, 5'-CGGTGACCAATTCATCTGCTGCTGTT-3') and the

second primer set (SRA-2_s, 5'-ATAGTGACAAGATGCGTACTCAACGC-3' & SRA-2_as, 5'-AATGTGTTCGAGTACTTCGGTCACGCT-3'). PCR conditions were, 1min at 98°C, followed by 25 cycles of 55sec at 98°C, 30sec at 64°C, and 2min at 72°C, and a final extension of 5min at 72°C.

2.2 Trypanosome cultures

2.2.1 *In-vivo* cultures

Field isolates of *T. b. rhodesiense* in cryopreserved blood stabilates were inoculated intra-muscularly in rats. The rats were then tail bled on a daily basis to check for parasitaemia microscopically. At peak parasitaemia (approximately $4x10^6$ cells/ml), the rats were sacrificed and approximately 4ml of blood was drawn by cardiac puncture into an EDTA vacutainer for further experimentation. For the *in vitro* cultured *T. b. rhodesiense* strain Tbr729, approximately 5000 parasites were injected in a rat and peak parasitaemia would be observed within 5 days post infection.

2.2.2 *In-vitro* cultures

Bloodstream-form Trypanosoma brucei strains (Tbr729 and Lister 427) were cultured in loosely capped flat-bottomed T-flasks in an incubator at 37°C, 5% CO_2, in a humified atmosphere. The cell densities were maintained between $2\text{-}10x10^5$ cells/ml in supplemented HMI-9 medium (Table 2.1).

Table 2.1 Trypanosome culture HMI-9 medium

Reagent	Conc.
[a]IMDM (Gibco)	17.66g/L
$NaHCO_3$ (Roth)	36mM
Hypoxanthine (Serva)	1mM
Na-pyruvate (Serva)	1mM
Thymidine (Sigma)	160mM
Bathocupronsulphonate (Serva)	50mM
Supplemented HMI-9 medium, 500ml	
[b]Heat-inactivated FCS	10% (v/v)
Penicillin/Streptomycin (Sigma)	50U/L (5ml)
L-Cysteine-HCl.H_2O	1.5mM
β-mercaptoethanol (Sigma)	0.14%

[a]*Iscove's Modified Dulbecco's Medium, no supplements*
[b]*FCS heat-inactivated by incubation at 56ºC for 30min*

2.3 Isolation of Trypanosomes from blood

2.3.1 Anion Exchange chromatography

Trypanosomes in infected rat blood were harvested by anion exchange on DEAE cellulose as described by (Lanham and Godfrey, 1970). Briefly, 20ml packed slurry volume of DEAE cellulose resin (Sigma) pre-equilibrated in phosphate saline glucose (PSG) buffer, pH 8, was set up in a 30ml syringe. The column was equilibrated with 3 column volumes of PSG and thereafter loaded with anti-coagulated rat blood at room temperature. The trypanosomes were eluted with PSG into a 15ml tube and then centrifuged for 10min at 2500rpm. The parasites were then resuspended in Trifast reagent for RNA analysis or subjected to genomic DNA extraction using the Zymogen kit.

2.3.2 Hemolysis centrifugation

Trypanosomes in infected rat blood were also harvested by hemolysis followed by centrifugation as described by (Njogu and Kiaira, 1982). Briefly, to one volume of infected rat blood was added 3 volumes of Erythrocyte lysis buffer (Qiagen), gently mixed and incubated at room temperature for 7min. This was then centrifuged at 2500rpm for 10min. The pellet was then resuspended in Trifast reagent for subsequent RNA extraction.

2.3.3 Buffy coat

The infected rat blood in an EDTA tube was centrifuged at 3000 rpm for 10min to differentially separate the red blood cells from the white blood cells in the buffy coat. The buffy coat, which contains the trypanosomes, was pipetted off and centrifuged at 6000rpm for 2min. The pellet was then resuspended in Trifast for RNA extraction.

2.4 Trypanosome Transcriptome Analysis

2.4.1 Total RNA extraction

Total RNA from the cellular pellets was extracted using Trifast reagent (Peqlab, GmbH). 200µl of chloroform was added to the 1ml cell suspension, vortexed vigorously and incubated at room temperature for 3min. The mix was then centrifuged at 13000rpm for 10min at 4°C and the aqueous phase pipetted into a 1.5ml eppendoff lobind tube. 500µl of isopropanol was then added, mixed thoroughly and incubated at room temperature for 10min to allow for RNA precipitation. The mix was then centrifuged at 13000rpm, 4°C for 20min and the RNA pellet washed with 1ml of 75% ethanol. The ethanol was removed by centrifugation at 13000rpm for 5min, the RNA pellet air dried and dissolved in 20µl of nuclease free water. The total RNA concentration was then determined using the Qubit 2.0 fluorometer (Invitrogen).

2.4.2 PolyA+ selection of mRNA

RNA molecules with a 3' end polyA tail (mRNA) were selectively purified from the total RNA. For this, the total RNA was resuspended in 500µl of RNA denaturing buffer (20mM HEPES, pH7.4, 10mM EDTA, pH7.4, 1% SDS) and incubated at 65°C for 10min then on ice for 2min. 750µl of RNA dilution buffer (20mM HEPES, pH7.4, 800mM NaCl) and 100µl of pretreated oligo(dT) cellulose (Amersham, Table 2.2) was then added, and the tube placed on a rotatory mixer for 15min at room temperature. The mix was then centrifuged at 12000rpm for 2min and the cellulose pellet washed 2 times with RNA washing buffer (20mM HEPES, pH7.4, 5mM EDTA, pH7.4, 0.5% SDS), centrifuging each time at 12000rpm for 2min. The cellulose was then incubated in 150µl of RNAse free water, mixed for 5min, centrifuged at 12000 for 2min and the supernatant transferred to a 1.5ml eppendorff lobind tube. Repeated the incubation in RNAse free water once and pooled the samples to 300µl total volume. The RNA was then precipitated by adding 30µl of 5M NaAc, 3µl Glycogen (10mg/ml), 900µl of absolute ethanol and placing at -20°C overnight. The precipitated RNA was then recovered by centrifugation at 13000rpm for 20min and resuspended the RNA in 6µl of RNAse free water.

Table 2.2 Oligo-dT cellulose pretreatment protocol

1. Added 5ml of 0.1M NaOH to 100mg oligo dT cellulose in 15ml tube. Inverted for 10min at room temperature, then centrifuged for 5min at 5000rpm.
2. Washed cellulose pellet twice with 15ml RNA binding buffer (20mM HEPES, pH7.4, 5mM EDTA, pH 7.4, 0.4% SDS, 500mM NaCl) at 5000rpm for 5min.
3. Resuspended the oligodT cellulose in 1ml of RNA binding buffer
4. Stored the pretreated oligodT cellulose at 4°C

2.4.3 First strand cDNA synthesis

The cDNA synthesis from the polyA+ RNA was carried out by reverse transcription (superscriptIII kit, Invitrogen), using a T3-promoter (5'GCGCGAAATTAACCCTCACT AAAGGGAGA 3') tagged oligo-dT (T3$(dT)_{24}$) and random nanomer (T3N9) to prime off the mRNA sequences. The reaction conditions are tabulated below.

Reagent	Conc.	Vol.
RNA Mix		6µl
T3dT	5pM	1µl
T3N9	5pM	1µl
RNAse Out (Inviitrogen)	40U	1µl

Incubated the mix at 65°C for 10min in a thermocycler (Biometra Analytic) and placed on ice for 4min. Then added 11µl of the cDNA synthesis mix tabulated below.

RT buffer	10x	2µl
$MgCl_2$	25mM	4µl
DTT	0.1M	2µl
dNTPs	10mM	1µl
T4gene32 protein (Roche)		1µl
Superscript III (Invitrogen)		1µl

Incubated at 50°C for 90min, 85°C for 5min, then added 1µl RNase H (2U) and incubated at 37°C for 20min. The synthesized cDNA was then purified using the RNAeasy mini-elute kit (Qiagen) and eluted in a volume of 12µl with RNAse free water.

2.4.4 Second strand cDNA synthesis

Trypanosome specific sequences in the trypanosome/WBC (HeLa) mix were targeted by specifically priming with the splice leader sequence, SL-4 (5'GATCTACAGTTTCTGTACTAT3') which occurs on the 5' end of every trypanosome mRNA. The reaction mix is shown below and the thermo-cycle conditions were 98°C for 20sec, 95°C for 1min, 50°C for 60min and 72°C for 60min.

Phusion buffer (Finnzymes)	5X	10µl
SL-4 primer	0.05pM	1µl
dNTPs	10mM	1µl
Phusion Pol (Finnzymes)		1µl
cDNA mix		12µl
Nuclease free water		25µl

2.4.5 Amplification of the double stranded cDNA

In order to specifically amplify trypanosome cDNA sequences, a nested PCR using the T3-promoter (5'GCGCGAAATTAACCCTCACTAAAGGGAGA3') and SL20 (5'ACAGTTTCTGTACTATATTG3') primers, was carried out. To the second strand cDNA synthesis mix was added 1µl of each, SL20 (5pM), T3 (5pM), and Phusion polymerase. The PCR conditions were set for 10 cycles of 95°C for 1min, 60°C for 3min and 72°C for 5min with a final extension at 72°C for 10min. The cDNA was purified (QIAquick PCR purification kit, Qiagen) eluted in 30µl and quantified using the Qubit 2.0 fluorometer (Invitrogen).

2.4.6 Radioactive analysis of synthesis

The second strand synthesis of double stranded cDNA was visualized by the incorporation of 25µCi of alpha (^{32}P)-dCTP in the polymerase reaction assay. The synthesized DNA was resolved on an 8% Urea acrylamide gel (Table 2.3), exposed to a Phosphor-imager screen and analyzed by the Fuji FLA7000.

The dscDNA was also analysed by *in-vitro* transcription using the T3 polymerase reaction mix (MAXIscript, Ambion Inc.). The reaction was carried out using 12µL of the purified dscDNA in a mix consisting of 0.5mM of ATP, CTP, GTP, 50µCi alpha (^{32}P)-UTP, 1X transcription buffer, T3 enzyme mix in a final volume of 20µl and incubated at 37°C for 10min. The reaction mix was then treated with DNase1, 1µl, and continued incubation at 37°C for a further 15min. The synthesized RNA was then resolved on an 8% Urea acrylamide gel, exposed to a Phosphor-imager screen and analyzed by the Fuji FLA7000.

Table 2.3 8% Urea-acrylamide gel preparation mix

Reagent	Amount
Urea	39.36g
Acrylamide 40 (Biorad)	20ml
TBE, 5X (Trisborate EDTA, pH 8)	18.8ml
Distilled water	13.4ml
APS, 10X (Ammonium perfsulfate)	562µl
TEMED	50µl

2.4.7 High throughput sequencing

Library preparation of DNA for sequencing was carried out on the cDNA (amplified and unamplified), genomic DNA, and the total RNA samples. The cDNA samples were first fragmented with the Covaris S2 system (Covaris), using the AFA microTube at an Intensity 5, 10% Duty Cycles with 200 cycles per burst for 90 seconds. A quality check on 1µl of the fragmented sample was done on the BioAnalyzer 2100 (Agilent Technologies) using the High Sensitivity chip. The Library was then prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs Inc.). Briefly, end repair, A-tailing and NEBNext singleplex adaptor ligation reactions were sequentially carried out on the fragmented DNA. Size selection of the adaptor ligated DNA (200 – 250bp) was then carried out using the AMPure XP beads followed by product enrichment with a 9 cycle PCR. The amplified DNA was cleaned up and 1µl of the library checked on the Bioanalyzer for quality. The library was then loaded onto a single lane of the flow cell and run on either the Illumina GAII or HiSeq 2000 system for single-end 76bp or 50bp reads respectively. The library preparation from genomic DNA was carried out using the same procedure as above however, 100bp paired end reads were generated for this case. On the other hand, the library preparation from total RNA was carried out using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs Inc.); and similarly run on the HiSeq 2000 system. All the samples were multiplexed.

2.4.8 Analysis of sequencing data

The sequenced DNA reads were aligned to the reference *T. brucei brucei* TREU 927 genome (TriTrypDB V4.0) using the Bowtie aligner and allowing for 2 mismatches (Langmead et al., 2009). All the reads that mapped to the coding sequence (CDS) were extracted. For those that did not align, reads containing the spliced leader (SL tags) or poly(A) and the T3 promoter were extracted, trimmed and assigned to an open reading frame based on their positioning within annotated gene coordinates (Fadda et al., 2013). The subsequent processing and sorting of the aligned reads was carried out using the SAMtools (Li and Durbin, 2009), and the read alignment to the genome visualized by Artemis (Carver et al., 2008; Rutherford et al., 2000). DESeq was used to identify differentially expressed genes, with a cutoff p-adjusted value of 0.05. Functional category enrichment was carried out using Fisher exact test in R. The heat map was generated in R. To plot the average read density along genes, we selected around 1000 short genes of 200-1000b in length, and a similar number of long ones >2.5kb in length, that had a read coverage of at least 1 on average. PERL scripts were written to extract reads aligning to a window of 0.5% of

each gene, which were then normalized to the total read count for the gene, and an average was plotted against the relative location along 1kb of standardized gene length. The gene expression profiles were analyzed as read density per gene and expressed as either RPM (reads per million) or RPKM (reads per kilobase per million mapped reads). The statistical analysis was carried out using Microsoft excel and R (CRAN R-project) tools and the graphs plotted in KaleidaGraph.

2.4.9 Quantitative Real time PCR analysis

The expression levels of four genes in the amplified (TH) samples were analyzed by real-time quantitative PCR. The amplified cDNA from the 5ng trypanosome to 5µg Hela total RNA mix (TH) was used as a template for the qRT-PCR. The qRT-PCR was carried out using the SL20 forward primer and the gene specific reverse primer annealing downstream of the splice acceptor site within the ORF or 5' UTR of the target gene (Table 2.4). Briefly, 2µl of the amplified cDNA was added to 8µl of the master mix (2x SYBR Green, 100pM SL20, 100pM Rev primer) in triplicate in a 96 well plate. The same was done for the HeLa background control and water template negative control. The unamplified trypanosome cDNA template was diluted (1:10, 100, 500, 1000, 1500, 2000) and 2µl of this used as a template in triplicate for the RT-PCR reaction along with the amplified samples. The real time PCR reactions were performed using the LightCycler480 (Roche) and the amplifications were done using the SYBR Green PCR Master Mix (Roche). The thermal cycling conditions were composed of an initial denaturation step at 95ºC for 5 min, 45 cycles at 95°C for 20s, 60°C for 20s and 72°C for 5s. The experiments were repeated three times (3 technical replicates) and the relative quantification in gene expression was determined using the 2- **ΔΔ**Ct method (Schmittgen and Livak 2001, 2008). The fold changes in gene expression were normalized by relative expression of the same genes from the unamplified trypanosome cDNA (T).

Table 2.4 List of reverse primers for qRT-PCR

CZ No.	Description	Sequence
4438	Tb927.1.2330_rev	AGCCTGAACGCAGACGATT
4432	Tb927.4.4220_rev	CCCACAATTGCCACCTTAAA
4446	Tb927.8.5260_rev	TGGTTGGTTCTTGTGTCTGTG
4448	Tb927.10.8710_rev	ACACCCGCACCAGTCAGT

3 Results

3.1 Transcriptome of amplified Trypanosome cDNA

3.1.1 Working hypothesis

Trypanosomes can be isolated from the buffy coat of infected blood (Woo, 1970), which also contains the host white blood cells. The average total RNA yield from adult blood samples collected in PAXgene tubes or Tempus tubes is 5µg/ml of blood (Duale et al., 2012). However, peripheral blood mononuclear cells (PBMC) at a concentration of 10^6 cells/ml yield approximately 2µg of total RNA (Eikmans et al., 2013; Marteau et al., 2005). We estimated the blood parasitaemia in an infected adult to be about 2000 trypanosomes per ml of blood. Since one trypanosome has approximately 0.5pg of total RNA (Haanstra et al., 2008), 2000 parasites would yield approximately 1ng of total RNA. Therefore we estimated that during active infection, the ratio of human cellular RNA to the Trypanosome RNA would be 5ng to 5µg respectively, from 5ml of blood. Based on these estimations, there was a need to amplify the trypanosome RNA to the threshold concentration required for illumina high throughput sequencing (100ng cDNA).

Therefore, by mixing trypanosome total RNA with HeLa cell total RNA (5ng:5µg), I developed an exponential (PCR based) amplification method to specifically amplify the trypanosome cDNA, as illustrated in Fig 3.1 (Kurimoto et al., 2007; Tang et al., 2011). This approach utilized the splice leader sequence that is located at the 5' end of all trypanosomatid mRNA following post transcription trans splicing (Liang et al., 2003). With this approach, we envisaged a specific enrichment of trypanosome cDNA above the human (HeLa) cDNA background to a concentration sufficient for Illumina sequencing.

3.1.2 Single stranded cDNA synthesis

Following the selection of polyadenylated RNA (PolyA+ RNA) by oligo(dT) chromatography, reverse transcription of this mRNA to cDNA was carried out. The mRNA templates were primed with oligo(dT) and random nanomer primers that were modified at the 5' end to contain a promoter sequence of the T3 RNA polymerase, $T3(dT)_{24}$ and T3N9 (Eberwine, 1996). The random primers were to capture degraded mRNA and those mRNAs that are too long to synthesize full length cDNA using oligo(dT) (Xiang et al., 2003). The isothermal reverse transcription reaction by Superscrpit III (Invitrogen) was maintained at 50°C for maximum efficiency (Malboeuf et al., 2001). The efficiency of cDNA synthesis was also enhanced by addition of T4 Gene 32 Protein, which reduces the formation of higher order structures of RNA molecules thus reducing any pause sites during the course of cDNA synthesis (Rapley, 1994; Villalva et al., 2001).

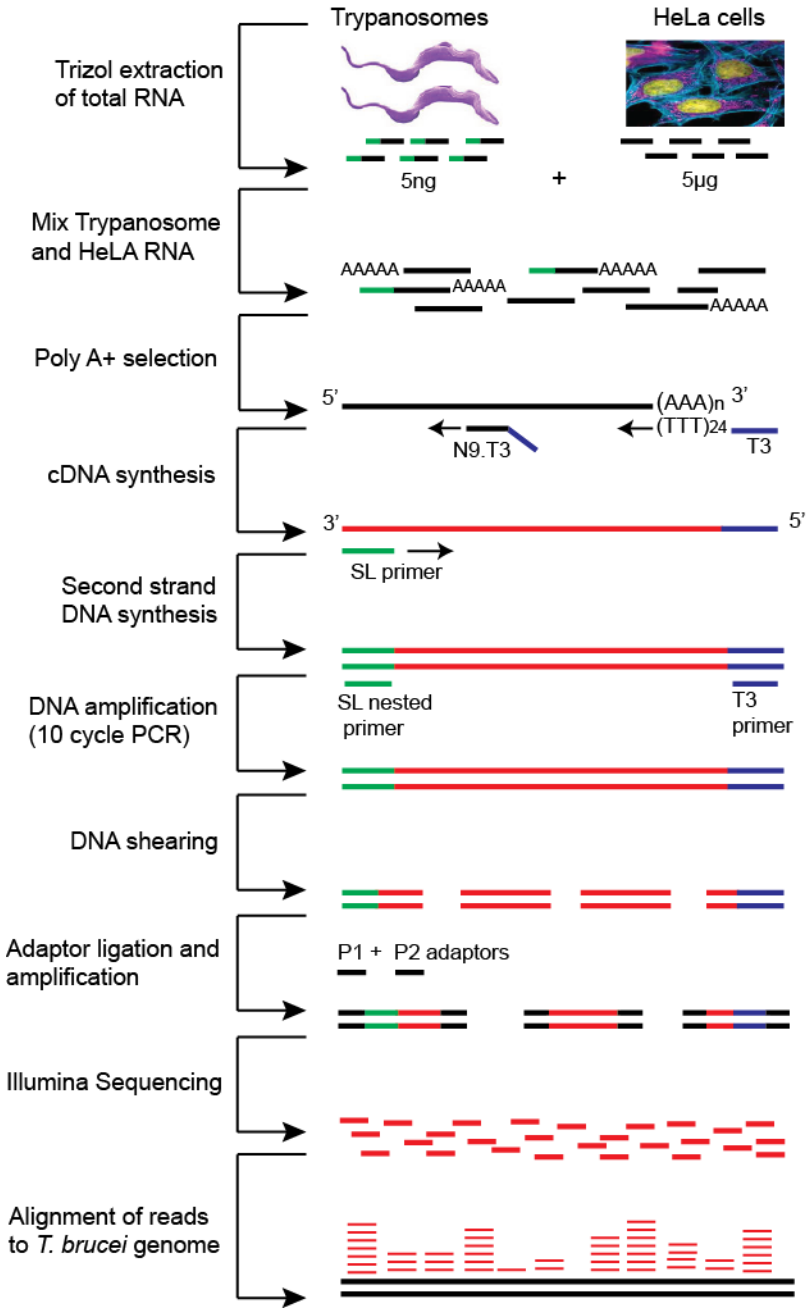


Figure 3.1 Schematic representation of the trypanosome RNA amplification and sequencing workflow. Total RNA extracted from trypanosomes and HeLa cells was mixed at a ratio of 5ng to 5µg respectively. PolyA+ mRNA was reverse transcribed into cDNA using T3 RNA polymerase promoter sequence tagged oligodT and random nanomer primers. The second strand cDNAs were synthesised using the splice leader (SL) primer, which targets only the trypanosome cDNA in the mix. The double stranded cDNA was then amplified by Nested PCR using the splice leader and T3 sequence primers, fragmented, and the P1 and P2 adaptors ligated to the ends. The libraries were enriched by a PCR, size selected and loaded onto the flow cell for Illumina sequencing. The out put reads were then aligned onto the *T. brucei* genome and transcriptomic analysis of the nanogram amplified Trypanosome RNA carried out.

3.1.3 Second strand cDNA synthesis

All trypanosome mRNA transcripts have a 39nt splice leader (SL) sequence at their 5' end (Liang et al., 2003). Furthermore, RNASeq analysis carried out on DNA generated by second strand synthesis using the splice leader sequence primer was able to map the 5' UTRs of the transcriptome of bloodstream trypanosomes (Kolev et al., 2010; Nilsson et al., 2010). Therefore in order to specifically detect trypanosome cDNA in the Tryps-HeLa cDNA mix, a primer annealing to the splice leader sequence was used to initiate the second strand cDNA synthesis by a proofreading Phusion polymerase reaction (Finnzymes). In order to optimize this procedure, various primers, concentrations and temperature conditions were checked and the results presented are a summary of the many experiments that I carried out.

Following cDNA synthesis by reverse transcription, double stranded DNA was synthesized by a splice leader primed DNA polymerase reaction. Different splice leader primers that were derived from the full-length splice leader sequence (Fig 3.2A), were used to optimize the second strand synthesis conditions. In order to test the sensitivity of these primers for trypanosome cDNA in the trypanosome-HeLa mix, 5pM concentration of the primers was initially used and the reaction conditions were; denaturation at 95°C for 2min, annealing at 50°C for 3min and extension at 72°C for 5min (Fig 3.2B). I observed that the SL-4 primer had the highest sensitivity with over 3-fold detection of trypanosome cDNA (Fig 3.2B, lane 10), with minimal HeLa cDNA background as compared to the full-length SL (Fig 3.2B, lane 2) and SL20 (Fig 3.2B, lane 6) primers.

In order to show that these double stranded cDNA transcripts had a T3 promoter sequence at the 3' end, *in-vitro* transcription using the T3 polymerase was carried out (Fig 3.2C). RNA transcripts ranging from less than 25 bases to greater than 2 Kb were observed, implying that the single strand cDNA and subsequent double stranded cDNA synthesis reactions efficiently captured the Trypanosome transcriptome. However it is worth noting that the HeLa background (Fig 3.2C, lane 3) could also show transcripts from the HeLa transcriptome captured by the two reactions.

Using the SL-4 primer, the efficiency of the second strand cDNA synthesis was enhanced by increasing the annealing and extension times to 60min (Fig 3.2D). I observed that by reducing the SL-4 primer concentration to 0.05pM, there was enhanced specificity with a 5.4 fold increase in detection of the trypanosome cDNA (Fig 3.2D, lane 3) and relatively low HeLa cDNA background detection (Fig 3.2D, lane 4). Using these conditions for second strand cDNA synthesis, the choice for either OligodT PolyA+ (Amersham) or RiboMinus (Invitrogen) selection of mRNA transcripts was checked (Fig 3.2E). The PolyA+ method had less HeLa background detection (Fig 3.2E, lane 3) as compared to the RiboMinus method (Fig 3.2E, lane 2).

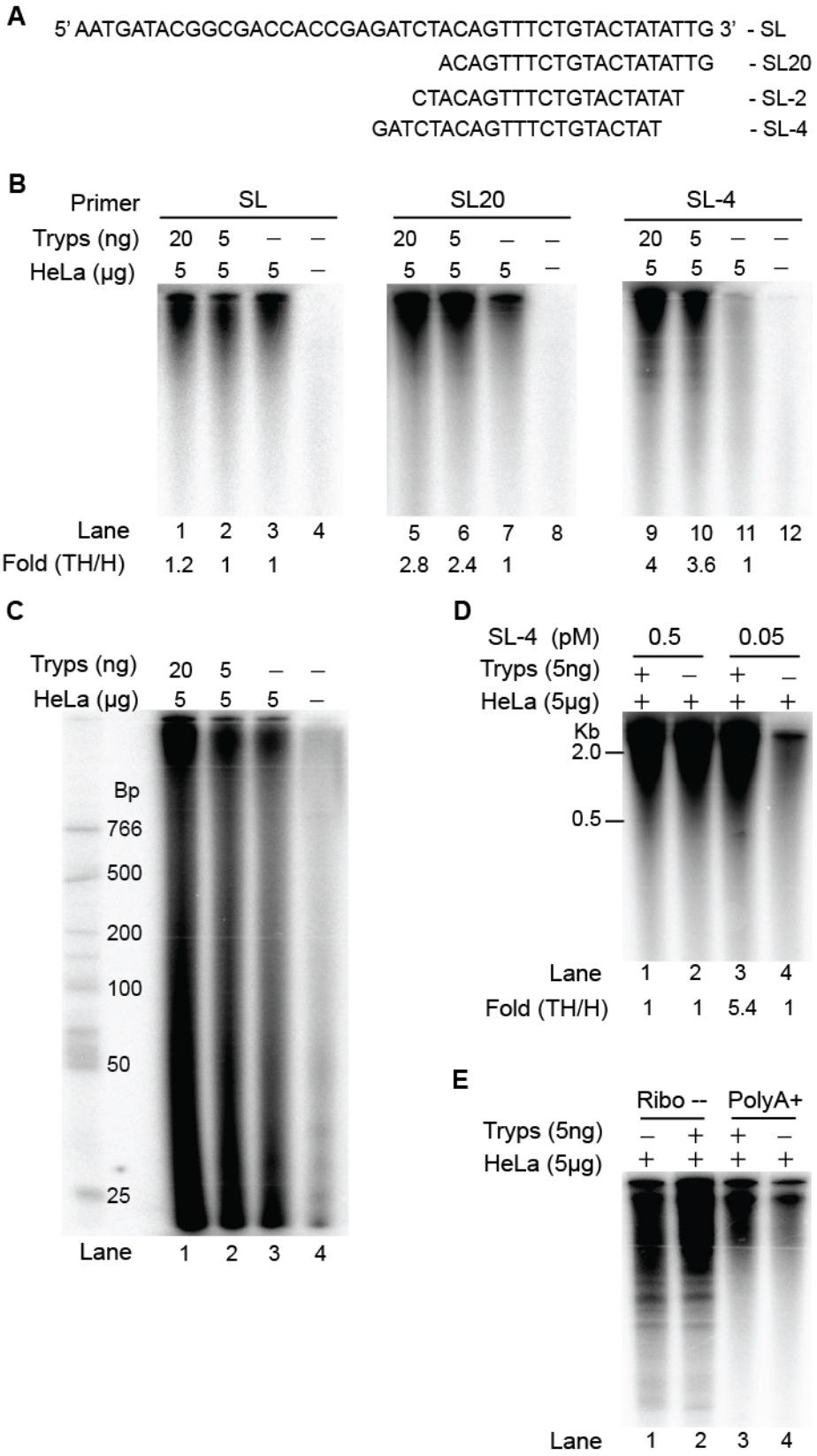


Figure 3.2 Second strand synthesis of double stranded cDNA in the Trypanosome HeLa mix.
A. The splice leader sequence primers; SL, full length splice leader, SL20, primer spanning the first 20bp (from the 3'end) of the full-length, SL-2, primer spanning 20bp after the first 2 bases (TG) of the full-length and SL-4, primer spanning 21bp after the first 4 bases (ATTG).

B. Second strand synthesis reaction using the splice leader primer and Phusion polymerase incorporating the radioisotope α^{32}P-dCTP at 95°C for 2min, 50°C for 3min and 72°C for 5min. The dscDNA was run on 8% Urea-polyacrylamide gel, exposed to a phosphor imager and visualised using the Fuji FLA7000.
C. *In vitro* transcription reaction using T3 polymerase incorporating α^{32}P-dUTP at 37°C for 60min. The RNA transcripts were run on 8% Urea-polyacrylamide gel exposed to a phosphor imager.
D. Second strand synthesis reaction using the SL-4 primer at a concentration of 0.5pM or 0.05pM and Phusion polymerase incorporating the radioisotope α^{32}P-dCTP at 95°C for 2min, 50°C for 60min and 72°C for 60min.
E. Analysis of OligodT PolyA+ and RiboMinus selection methods of RNA transcripts.

3.1.4 Amplification of double stranded cDNA

Following second strand dscDNA synthesis, the next step was to amplify the DNA to a concentration sufficient for high throughput sequencing. Since the dscDNA was a heterogeneous mixture of Trypanosome and HeLa DNA, exponential amplification by PCR rather than linear amplification by *in vitro* transcription (IVT) was opted for (Kurimoto et al., 2007; Tang et al., 2009). There was a high HeLa background observed by *in vitro* transcription (Fig 3.2C). Furthermore, there was a lower yield of RNA following *in vitro* transcription and multiple rounds to obtain the required concentration would have magnified the background effect.

First, I tested for which annealing temperature would have the most favourable primer hybridization efficiency for the amplification. For this, I used the splice leader primer downstream of the SL-4 (second strand synthesis primer) for increased specificity (nested primer). Three annealing temperatures, 55°C, 60°C and 65°C, were tested by second strand synthesis using the 20mer splice leader primer (Fig 3.3A). Different annealing temperatures were analyzed as well. I observed that the annealing temperature of 60°C had a 3 fold detection of trypanosome cDNA using 5pM of SL20 primer, in comparison to 55°C (1 fold) and 65°C (1.6 fold) (Fig 3.3A). It was rather odd that the 60°C was more specific than 65°C, but these observations were reproducible.

The amplification was carried out by the polymerase chain reaction (PCR). For this, the number of PCR cycles was set to 10, which would result in approximately 1,024 copies of the original DNA (2^n, n= No. of cycles). By setting the cycles to a minimum, the effects of saturation and amplification bias would also be avoided. The reaction used the high fidelity proof reading Phusion DNA polymerase (Finnzymes) with enhanced processivity (Wang et al., 2004). The PCR was initiated by adding 5pM of the 20mer splice leader primer (SL-2 or SL20) in combination with 5pM of T3 promoter sequence primer to the second strand synthesis mix. Then, 10 cycles of 95°C for 1min, 60°C for 3min and 72°C for 5min was carried out. I observed that the SL20/T3 primer combination had a 2.4 fold enrichment of trypanosome amplicons in comparison to the SL-2/T3 combination (Fig 3.3B).

Therefore, the overall conditions that were adapted for the amplification protocol were; following PolyA+ selection of mRNA, second strand synthesis was carried out using 0.05pM SL-4 primer under the conditions, 95°C for 2min, 50°C for 60min and 72°C for 60min. The double stranded cDNA (dscDNA) was then amplified in a 10 cycle Nested PCR using 5pM SL20 and T3 primers under the conditions; 95°C for 1min, 60°C for 3min and 72°C for 5min. The amplicons were then used to prepare the library for sequencing.

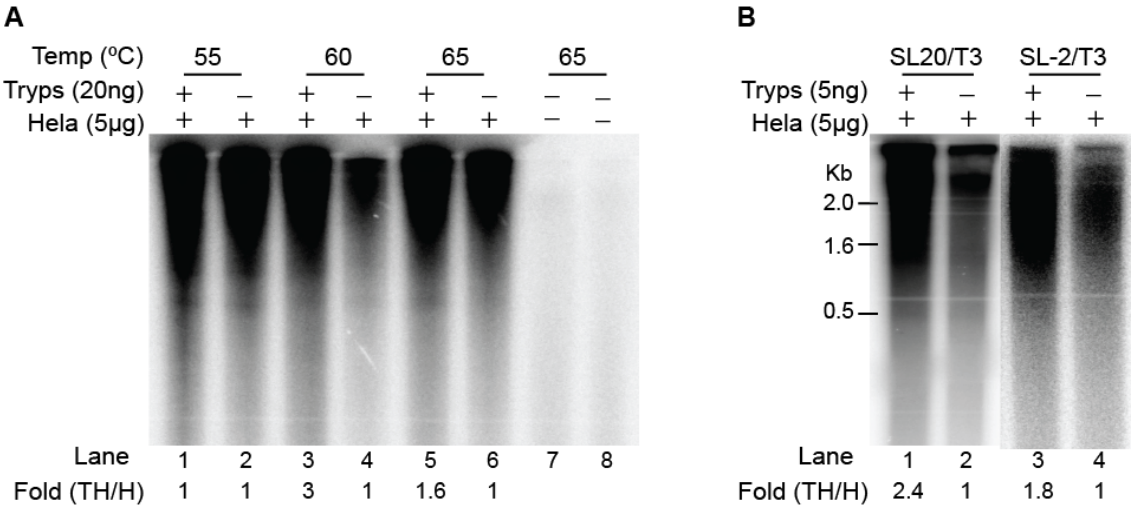


Figure 3.3 Amplification of double stranded cDNA in the Trypanosome HeLa mix.
A. Second strand synthesis reaction using SL20 primer and Phusion polymerase incorporating the radioisotope α^{32}P-dCTP at 95°C for 2min, 55°C or 60°C or 65°C for 3min and 72°C for 5min. The dscDNA was run on 8% Urea-polyacrylamide gel and exposed to a phosphor imager.
B. Amplification of dscDNA by 10 cycles of 95°C for 1min, 60°C for 3min and 72°C for 5min with Phusion polymerase incorporating the radioisotope α^{32}P-dCTP and amplicons run on 8% Urea-polyacrylamide gel. A comparison between the primer pairs SL20/T3 and SL-2/T3 at a concentration of 5pM was done for this nested PCR.

3.1.5 DNA Sequencing analysis

Having optimised the RNA amplification protocol, four technical replicates of the Trypanosome-HeLA RNA mix (TH, 5ng:5µg) were prepared for high throughput sequencing. The amplified DNA amounts of the replicates TH1 and TH2 were 90ng and 120ng respectively, and were sequenced using the Illumina GAII; TH3 and TH4 were sequenced using the Illumina HiSeq2000. In addition, a set up of unamplified trypanosome only cDNA (T), obtained by second strand dscDNA synthesis using 5pM SL20 primer (Fig 3.4A), were prepared for sequencing. In principle, it is the same method as the amplification protocol but without the nested PCR step and the use of microgram amounts of starting Trypanosome total RNA. Four replicates T1, T2, T3 and T4 with DNA amounts of approximately 200ng each were sequenced. T1 and T2 were run on the Illumina GAII where as the T3 and T4 samples on the Illumina HiSeq2000. Analysis of the sheared TH and T DNA samples (Fig 3.4C) showed an even distribution of fragments that were used

for the library preparation. Following the DNA shearing was adaptor ligation and size selection. DNA fragments in the range of 300–400bp were hybridised onto the Illumina flow cell for sequencing; generating reads that were subsequently analysed. Ibberson, D., carried out the library preparation.

3.1.6 Coverage of sequenced reads

The GAII (76mer) and HiSeq2000 (50mer) reads obtained following Illumina sequencing were aligned to the *T. b. brucei* TREU 927 version 4 reference genome using Bowtie-SAM tools (Langmead et al., 2009) and allowing for 2 mismatches (Table 3.1). This was done by Fadda A. The average percentage of reads that mapped to the Tb927 genome in the amplified Trypanosome – HeLa mix was 2.4% from the GAII and 20.4% using the HiSeq2000 system. There was a higher read density with the HiSeq2000 (8.5 fold) and hence more reads that were aligned from this platform. The unamplified trypanosome only transcriptome also showed a significant difference in the sequencing platforms, the GAII having an average of 57.7% and HiSeq with 83.3% reads mapping to the Tb927 genome. Considering that the ratio of trypanosome to HeLa RNA was 1:1000, the bulk of the reads (80% - 98%) aligned to the Human genome. However the number of genes with greater than 2 reads that were detected using a list of 6796 unique single copy ORFs (Siegel et al., 2010) were 4346 for the GAII reads. This implied that the transcriptome coverage of amplified 5ng of Trypanosome RNA in 5µg of HeLa was 64%. On the other hand, the replicate run using the HiSeq2000 system detected all the unique ORFs with 99% coverage. This same coverage was observed for the unamplified trypanosome cDNA (T) for this unique set of ORFs in both the GAII and HiSeq2000 data sets.

A.

Illumina GAIIx	**Unamplified**		**Amplified**	
Sample	**T1**	**T2**	**TH1**	**TH2**
Reads processed	17,790,273	14,796,988	17,909,805	20,350,255
Reads aligned to CDS	10,348,505	8,379,566	241,924	409,262
SL, Poly A, T3 tags	67,355	47,367	100,964	117,369
Total Reads aligned	10,415,860	8,426,933	342,888	586,631
% Reads aligned	58.54%	56.95%	1.91%	2.88%

B.

Illumina HiSeq2000	**Unamplified**		**Amplified**	
Sample	**T3**	**T4**	**TH3**	**TH4**
Reads processed	47,539,876	59,769,712	30,889,345	34,076,525
Reads aligned to CDS	33,441,648	42,639,172	4,585,299	6,508,792
SL, Poly A, T3 tags	6,143,230	7,228,812	869,722	1,403,415
Total Reads aligned	39,584,878	49,867,984	5,455,021	7,912,207
% Reads aligned	83.26%	83.43%	17.66%	23.22%

Table 3.1 Summary of sequenced read output and mapping to the *T. brucei* TREU 927 reference genome. The read output from the Illumina GAII (**A**) and HiSeq2000 (**B**) platforms. The samples T1 and T2, TH1 and TH2 were multiplexed on a single lane of the flow cell, similarly done for T3, T4, TH3 and TH4.

Although the reads obtained from the amplified (TH) transcriptomes were much less than those from the unamplified (T) transcriptomes, they covered the entire 11 chromosome genome of *T. brucei*. I compared the reads from the amplified (TH), unamplified (T) and the fragmented PolyA+ RNA (Fig 3.4B) of the bloodstream form 1313 (Manful et al., 2011; Roberts et al., 2011) mapping to chromosome 2, 4, 8 and 10 by artemis genome viewer (Fig 3.5). I observed that the mapping pattern along the chromosomes was more comparable between the amplified (TH) and unamplified (T) transcriptome than the BF1313 PolyA+ (Fig 3.5A). Furthermore, the mapping pattern for the HiSeq reads was also comparable between the amplified and unamplified transcriptomes, with subtle differences in certain regions within the chromosomes (Fig 3.5B). This was probably due to the considerable technical similarities in the methodologies for obtaining the trypanosome transcriptome from the amplified and unamplified cDNA (Fig 3.1, Fig 3.4A) as compared to the polyA+ fragmented RNA method (Fig 3.4B).

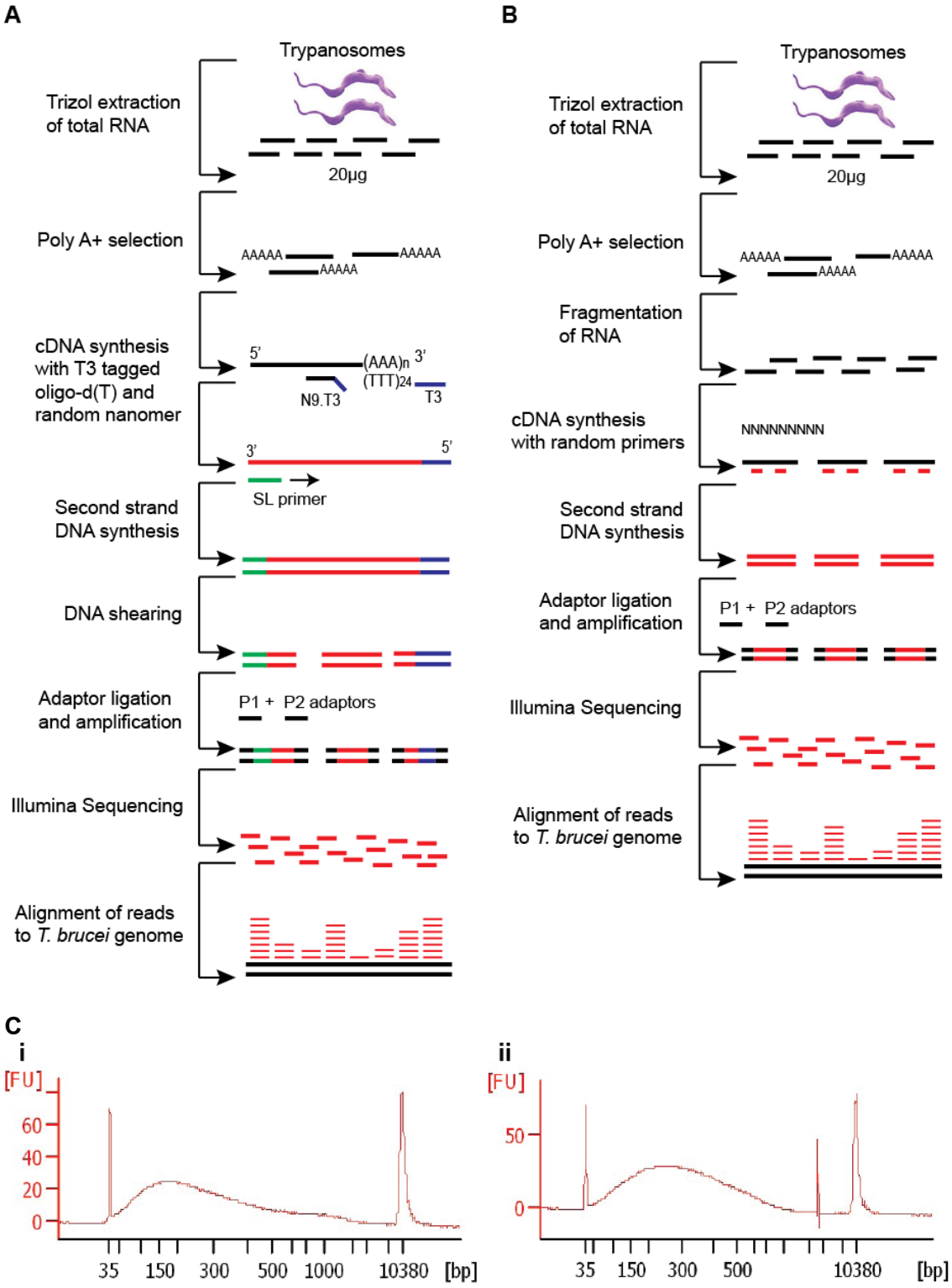


Figure 3.4 Schematic representation of unamplified Trypanosome RNA library preparation for RNASeq. The total RNA from Trypanosomes was subjected to polyA+ selection and either; **A**. Second strand synthesis was carried out on full-length transcripts and then fragmented or **B**. Fragmented the transcripts and carried out the second strand synthesis; this is the reference procedure routinely used for RNASeq analysis. **C.** Aligent bioanalyzer electropherograms of DNA sheared by the Covaris. Electophoretic profiles of sheared amplified (TH) Trypanosome-HeLa mix DNA (**i**) and unamplified (T) Trypanosome only DNA (**ii**).

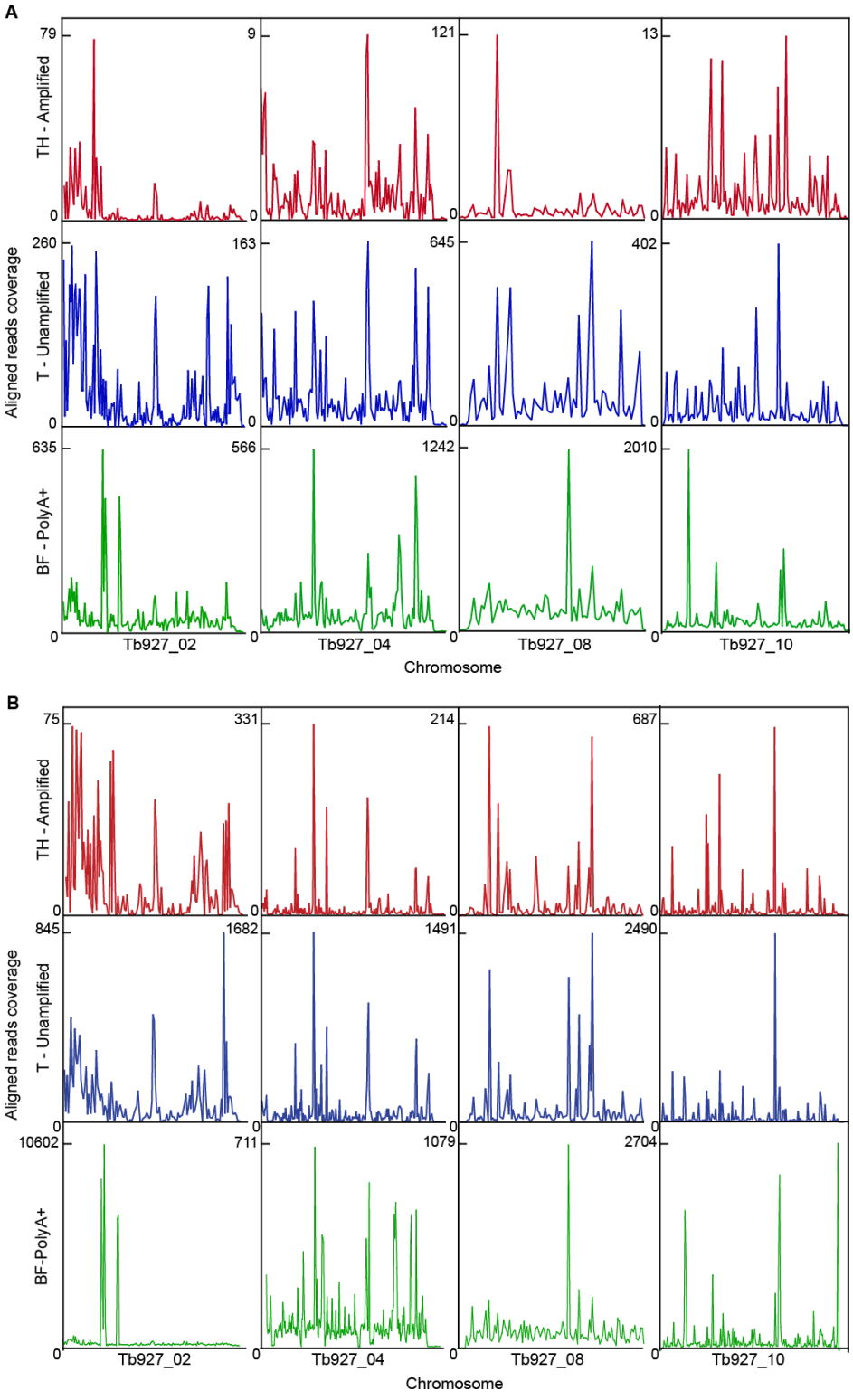


Figure 3.5 Artemis visualization of read coverage over entire chromosome. Reads obtained from the trypanosome amplified (TH, red), unamplified (T, blue) and polyA+ BF1313 (Green) were mapped to the *T. brucei* TREU 927 reference genome v4. The y-axis is a scale representing the read depth of the highest peak of reads. **A.** Comparison of the GAII reads coverage in all three methods on chromosomes 2, 4, 8 and 10, showed a similar pattern between the amplified (TH) and unamplified (T). **B**. Comparison of the HiSeq2000 reads coverage in the amplified (TH) unamplified (T) and BF1313 samples.

3.1.7 Comparison of transcriptome libraries

Comparison of replicates

I compared the sequenced libraries by first, normalising the reads mapped per gene to the total mapped reads, hence obtaining the reads per million (RPM); then determined the Pearson correlation coefficients (Fig 3.6). The transcriptome profiles between the technical replicates that were obtained from either the GAII or the HiSeq2000 were highly similar. The correlation coefficients of the amplified (TH) transcriptome replicates were 0.84 and 0.95 for the GAII and HiSeq2000 libraries respectively (Fig 3.6A, 3.6B). There was low correlation for the low abundance transcripts with fewer read counts. A comparison between the mean of the TH1 and TH2 libraries and mean of the TH3 and TH4 libraries (Fig 3.6C) showed a correlation of 0.79, implying that they had some similarity but with a degree of dispersion. However, this low correlation was probably due to the low read density from the GAII system, thus fewer genes detected in comparison to the HiSeq2000 system. There were over 1700 genes that were not represented in the amplification experiment sequenced on the GAII system. The correlation coefficients of the unamplified (T) transcriptome replicates were 0.92 and 0.93 for the GAII and HiSeq2000 libraries respectively (Fig 3.6D, 3.6E). The mean of the T1 and T2 libraries when compared to the mean of the T3 and T4 libraries showed a correlation of 0.92 (Fig 3.6F), implying they were highly similar. The correlations between the unamplified libraries (T) were better than those from the amplified libraries (TH). This was probably because the unamplified library started with a high amount of trypanosome only cDNA (200ng) for the sequencing library preparation. Since there was no amplification (resulting in bias and artefacts), there was a consistent coverage of the entire transcriptome hence the high correlations. This high correlation observed within replicates and between the platform (GAII and HiSeq2000) libraries suggested that the protocol for these methods was highly reproducible.

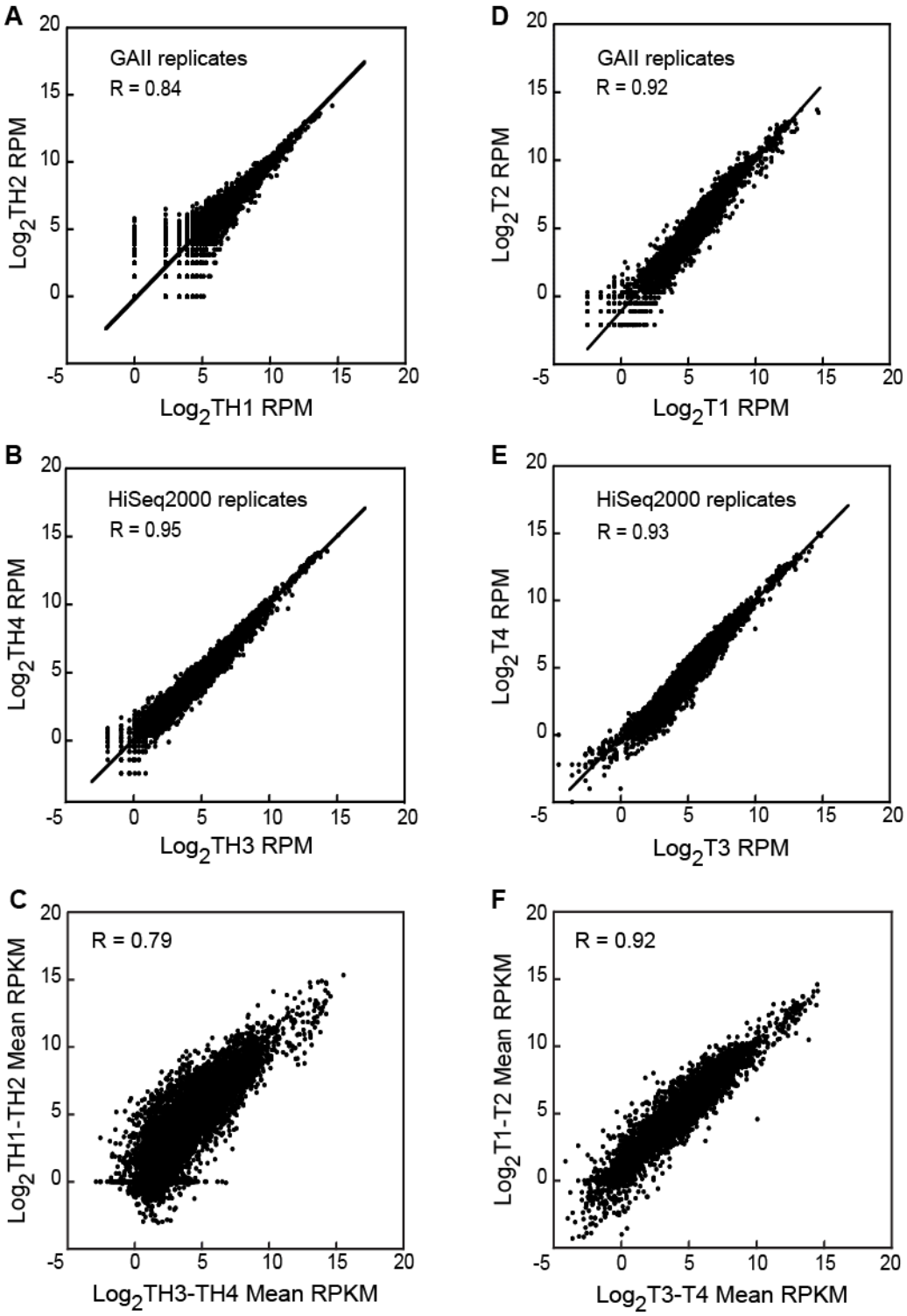


Figure 3.6 Scatter plots comparing the transcriptome libraries of the trypanosome amplified (TH) and unamplified (T) replicates. Each dot represents a single gene. The degree of similarity between libraries was calculated by the Pearson's correlation coefficient, R. Comparison of reads per million (RPM) between technical replicates of the amplified (TH) transcriptomes generated by the GAII (**A)** and HiSeq2000 (**B**) systems. Comparison of the mean read densities (RPKM) between the GAII and HiSeq2000 replicates of the amplified (TH) transcriptomes (**C**). Comparison of RPM between technical replicates of the unamplified (T) transcriptomes generated by the GAII (**D)** and HiSeq2000 (**E**) systems. Comparison of the mean read densities (RPKM) between the GAII and HiSeq2000 replicates of the unamplified (T) transcriptomes (**E**).

Comparison between libraries

The read alignment pipeline involved mapping the sequenced output reads to the *T. brucei* genome. These represented reads that mapped to the coding sequence (CDS). The unaligned reads were then filtered through for reads having the splice leader sequence (SL tag), polyA tag and T3 tag. These reads were then realigned to *T. brucei* genome and the total count of mapped reads was as a summation of CDS, Splice leader tags and polyA-T3 tags.

By using the HiSeq2000 libraries, the mean SL tags from both replicates of the amplified (TH) and unamplified (T) libraries were analyzed. The amplified (TH) had 60% (4048/6796) SL signatures of the unique genes, with reads greater than or equal to 2. The unamplified (T) had 83% (5659/6796) SL signatures with reads greater than or equal to 2. It was also observed that the splice leader tags from the amplified (TH) and unamplified (T) transcriptome libraries were similar with a high correlation of 0.81 (Fig 3.7A). Furthermore, an analysis of the normalised mean per gene read densities (Reads Per Kilobase per Million mapped reads, RPKM) between the TH and T showed a high correlation of 0.85 (Fig 3.7B). This similarity in correlation between the splice leader tags and Rpkms indicated that the transcriptome from only 5ng of amplified trypanosome total RNA (TH) was comparable and similar to the transcriptome of 200ng unamplified trypanosome cDNA (T). However the correlation between the unamplified (T) and the PolyA+ fragmented RNA (BF1313) transcriptomes was low, 0.68 (Fig 3.7C). This was also probably due to the differences in protocols for obtaining the two libraries, which was also exemplified in the read coverage along the genome (Fig 3.5A). Whereas the unamplified trypanosome cDNA (T) method involved selection for polyA+ mRNA, cDNA synthesis without RNA fragmentation, followed by second strand synthesis using a splice leader primer and then fragmentation of the cDNA for library preparation; the BF1313 method involved polyA+ selection, fragmentation of the RNA and cDNA library preparation (Figure 3.4A, 3.4B). These differences in library preparation could have resulted in the low correlation. Furthermore, there was a low correlation between the amplified transcriptome (TH) and the polyA+ BF1313, 0.53 (Fig 3.7D). This was probably due to the low starting RNA concentration, and technically due the amplification effects and process of library preparation. The amplified and unamplified replicates from the HiSeq2000 (Fig 3.7E) and GAII (Fig 3.7F) platforms showed a reproducibly similar correlation with the polyA+ BF1313. However it is worth noting that the amplified (TH) and unamplified (T) trypanosome transcriptomes were more comparable, and this is most probably due to the similar technical approach for the selection of splice leader tagged sequences and subsequent library preparation for sequencing.

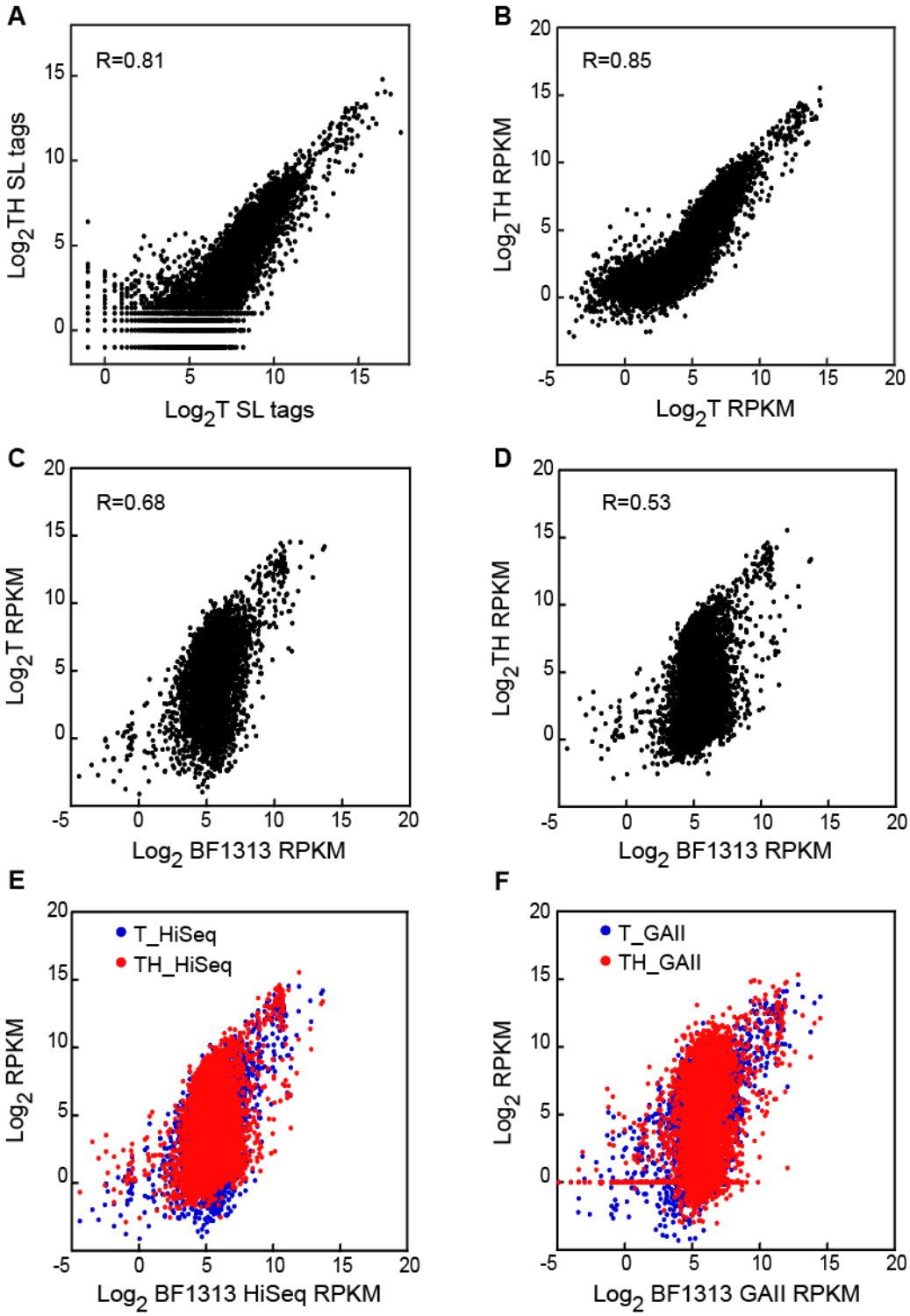


Figure 3.7 Scatter plots comparing the HiSeq2000 transcriptome libraries of the trypanosome amplified (TH), unamplified (T) and polyA+ BF1313. Each dot represents a single gene. **A.** Comparison between the means of the amplified (TH) and unamplified (T) splice leader tags. **B.** Comparison between the amplified (TH) and unamplified (T) mean relative gene expression profiles. **C.** Comparison between the unamplified (T) and polyA+ BF1313 gene expression profiles. **D**. Comparison between the amplified (TH) and polyA+ BF1313 relative gene expression profiles. Comparison between the amplified (TH) and unamplified (T) with the polyA+ 1313 in the HiSeq (**E**) and GAII (**F**) libraries.

3.1.8 Analysis for amplification bias

Analysis of RNAseq data has showed that there are tendencies of both positional and sequence specific bias in sequenced fragments (Hansen et al., 2010; Srivastava and Chen, 2010), which can subsequently affect the gene expression estimates (Li et al., 2010). Therefore in order to empirically investigate the effect of amplification of nanogram concentration of trypanosome RNA transcripts in microgram HeLA RNA background, I analysed for the global effect of transcript length bias against the relative gene expression (RPKM) was carried out on the HiSeq2000 data (Fig 3.8). I observed that there was a bias towards the shorter more highly abundant transcripts in the amplified (TH) transcriptome (Fig 3.8A). A similar trend (slope, -1.8) was observed for the unamplified (T) transcriptome (Fig 3.8B), suggesting that both methods are comparable in terms of transcript bias. This could possibly be due to the biased efficiency of the reverse transcriptase and DNA polymerase towards shorter fragments, during synthesis of full length cDNA (5'UTR-CDS-3'UTR). However, the BF1313 showed no length bias (Fig 3.8C) and this is probably due to the fact that the mRNA was fragmented into shorter strands prior to cDNA synthesis; resulting in even efficiency of the reverse transcriptase and DNA polymerase for cDNA synthesis. But a comparison of all three methods showed that the bulk of transcripts in the amplified (TH) and unamplified (T) transcriptomes had comparable gene expression to the unbiased BF1313 transcriptome (Figure 3.8D).

To further investigate transcript length bias on gene expression, I determined the frequency distribution of the binned average gene length against their average expression (Fig 3.9A). There was high gene expression observed for the short fragments between 0.03 to 1.5kb in both the amplified (TH) and unamplified (T) in comparison to the polyA+ BF1313 transcriptomes. These transcripts account for 38% of genes in the unique list (Siegel et al., 2010). This confirmed the observed length bias and this is probably due to the reasons explained above. However the transcripts greater than 1.5kb (62%) showed an even unbiased distribution of gene expression, a trend that was comparable for all 3 methods; although the BF1313 showed higher gene expression.

The length bias at the gene level was determined by Fadda, A., by analysing the read densities along a standardised gene length of 1Kb in both the short genes (0.2 – 1Kb) and long (greater than 2Kb). There was a uniform distribution of reads for the short genes in all three methods (Fig 3.9B) although the amplified (TH) and unamplified (T) genes had a short peak of the splice leader reads at the 5' end. On the other hand, the long genes showed a lower coverage of reads in the middle of the gene for the amplified and unamplified methods in comparison to the polyA+ sheared BF1313 method (Fig 3.9C). In addition, there was observed 5' end bias with a significantly high peak of reads in this region for the amplified and unamplified methods. This was certainly due to the splice

leader priming used in these methods. A pronounced 3' end peak was also observed in the unamplified method but neither in the amplified nor BF1313 method. This peak corresponds to the T3 priming at the end of the gene, which probably resulted in the observed bias. Overall, the observed effects on the short and long genes showed the efficiency of the DNA polymerase read through of the full-length genes in the amplified and unamplified methods. Whereas the polymerase is highly efficient for the short genes (even coverage of reads), it is not efficient for the long genes resulting in the 5' and 3' end bias, with minimal read coverage in the middle of the gene. The use of the random primers should have improved the read coverage in the middle of the genes.

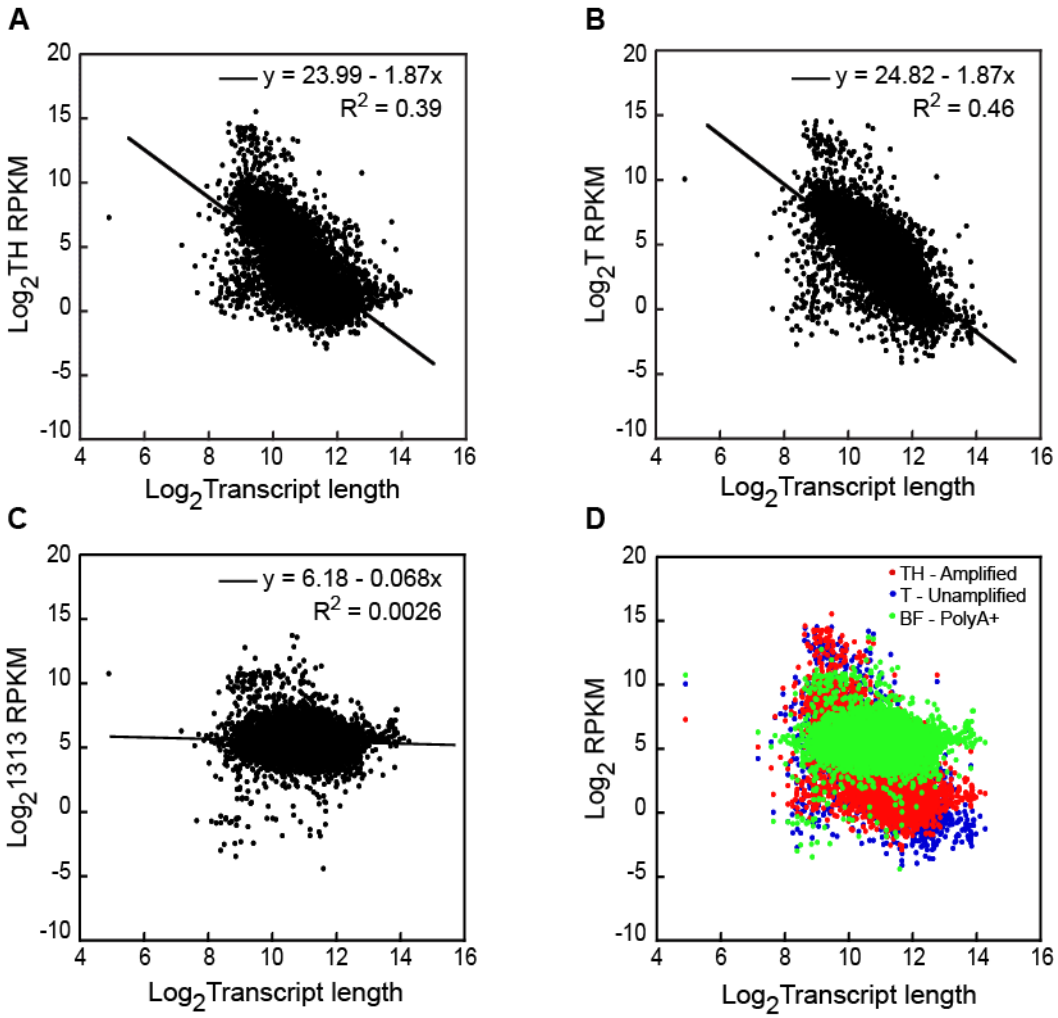


Figure 3.8 Scatter plot analysis of the effect of transcript length on the relative gene expression. Length vs Rpkm plots of the amplified (TH), **A**, unamplified (T), **B**, polyA+ BF1313, **C** and all three methods, **D**. transcriptome libraries. The means of the HiSeq2000 libraries where considered for this analysis.

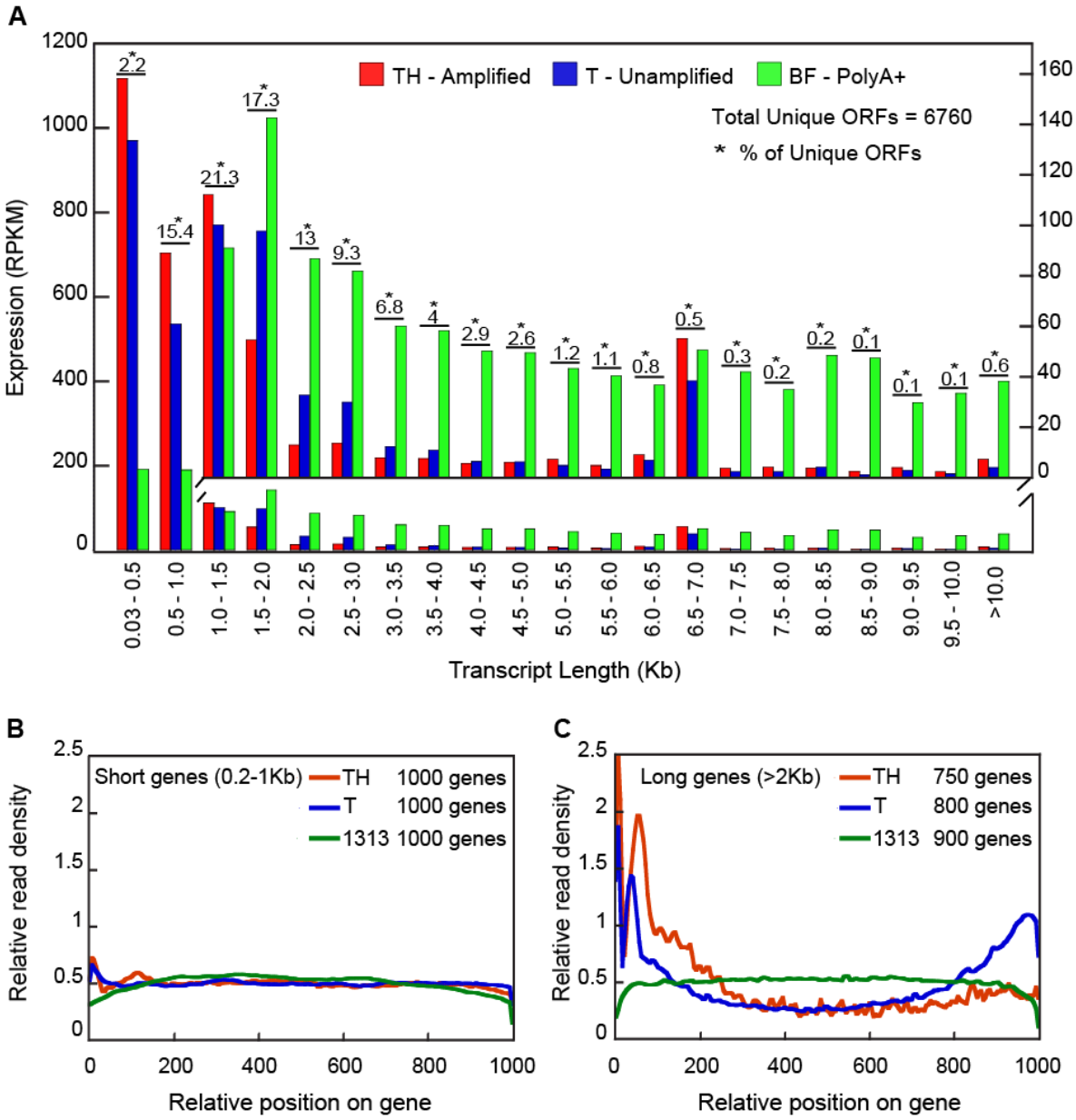


Figure 3.9 Analysis of gene length and gene position bias in the amplified, unamplified and polyA+ BF1313 transcriptome libraries. **A**. Histogram showing the frequency distribution of the average gene expression (RPKMs) according to there binned gene lengths. ***** The percentage of the gene length category in the total unique ORFs (Siegel et al., 2010). A plot of average relative read densities along a standardized gene length of 1kb for a selected group of long genes **B**, and short genes **C.** Perl scripts were used to extract reads aligning to a window of 0.5% of each gene, which were normalized to the total read count of the gene, and an average was plotted against the relative location along the 1Kb of standardized gene length.

3.1.9 Verification of amplification procedure

In order to check and show that the amplification protocol has worked in samples suspected to have trace amounts of trypanosome RNA (such as, Human samples, Rodent samples, Tsetse fly samples), there is a need for markers that confirm amplification. Therefore a quantitative gene expression analysis was carried out on four candidate genes. These genes were selected on basis of length and abundance in the reference polyA+ BF1313, and were located on chromosomes 1, 4, 8 and 10. The genes were; long transcript with a high RPKM (Tb927.1.2330, beta tubulin, ORF-1329bp, 5'UTR-59bp), long transcript with low RPKM (Tb927.4.4220, putative small GTP-binding rab protein, ORF-1692bp, 5'UTR-1707bp), short transcript with high RPKM (Tb927.8.5260, 60S ribosomal protein L39, ORF-156bp, 5'UTR-116bp) and short transcript with low RPKM (Tb927.10.8710, putative centrin, ORF-498bp, 5'UTR-108bp) (Fig 3.10A). The long abundant beta tubulin showed a high read density across the gene in all 3 methods. However for the long less abundant GTP-binding rab transcript, there was a high read density bias in the 5'UTR for the amplified (TH) and unamplified (T) methods, with few reads in the ORF. This was probably due to the inability of the DNA polymerase to efficiently read through the entire 3.4Kb length of this low abundance transcript, hence restricted to the 1.7kb of the 5' UTR following the splice leader priming. On the other hand, the polyA+ BF1313 had reads in both the 5'UTR and ORF, since this library was generated from randomly sheared mRNA rather than full-length cDNA. The short abundant and less abundant transcripts showed an even read density across the gene for all the three methods. In general, it was observed in the amplified and unamplified methods that almost all the genes had the distinct peak of splice leader reads at the 5' end as a gene signature. Therefore in order to quantitatively validate this observation, the expression of these genes in the amplified (TH) samples was determined by quantitative real time PCR (Fig 3.10B). The quantitative RT-PCR was carried out using the splice leader forward primer (SL20) and the gene specific reverse primer annealing downstream of the splice acceptor site within the ORF or 5' UTR of the target gene. Using the unamplified (T) as control cDNA, the relative abundance of the transcripts in the amplified trypanosome (TH) samples and amplified HeLa (H) background was determined. There was no observed amplification of either transcript in the HeLa control implying that the splice leader and gene primers were specific to the trypanosome transcripts and in accordance with the RNASeq abundance profiles.

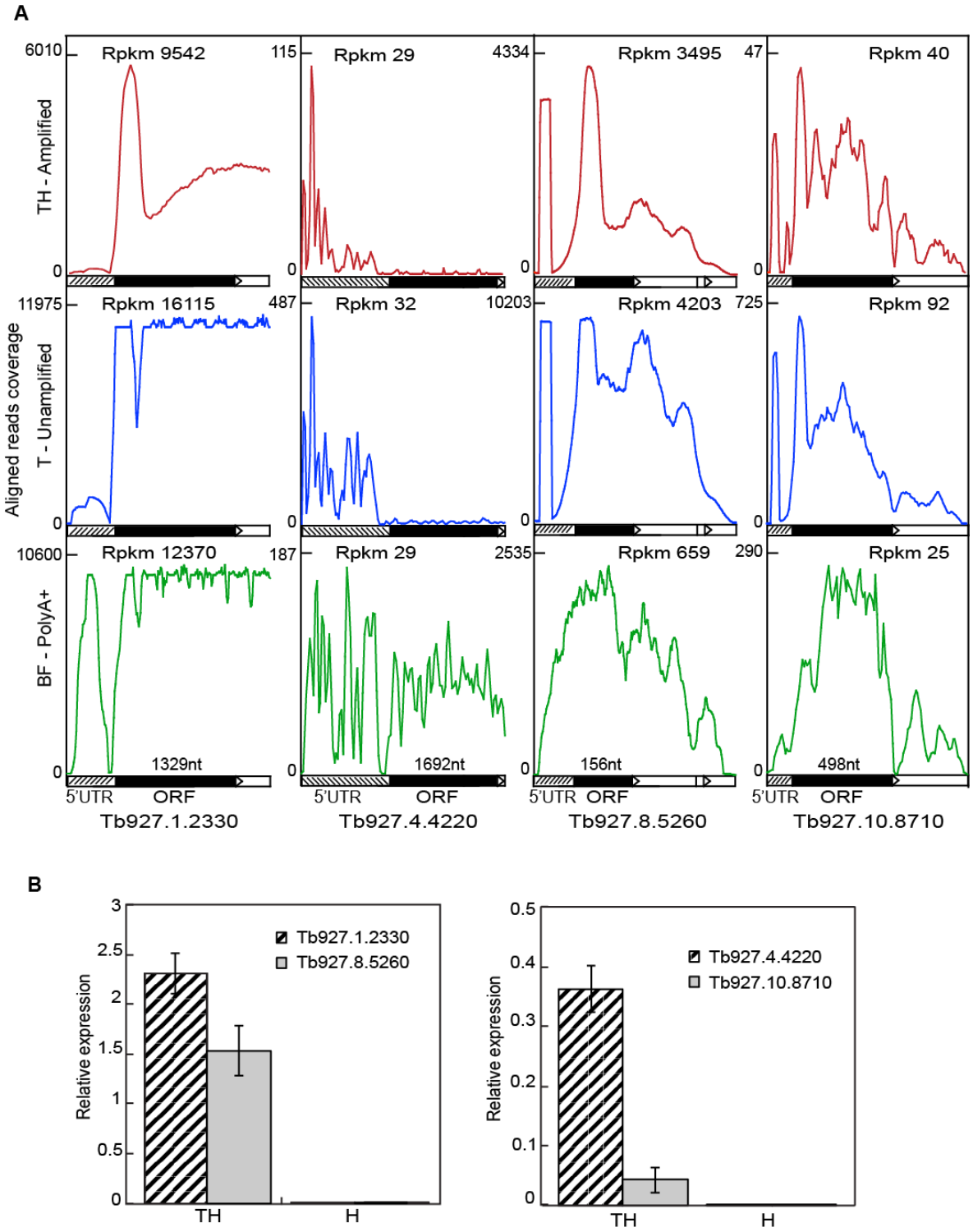


Figure 3.10 Gene expression analysis of four transcripts (Tb927.1.2330, Tb927.4.4220, Tb927.8.5260, Tb927.10.8710) based on length and abundance. **A**. Artemis visualization of the HiSeq2000 reads (TH4, T4 & BF1313) coverage over the open reading frame (ORF) and un-translated region (UTR) of the transcripts. **B**. Quantitative real time PCR on the selected transcripts in three separate technical replicates of amplified trypanosome (TH) and HeLa control. The SL20 forward and gene specific reverse primer were used for each gene. The relative expression level was determined using the unamplified trypanosome cDNA (T), error bars indicate standard deviation SD, (n=3).

3.2 Transcriptome of *T. b. rhodesiense* from HAT patients

3.2.1 Clinical sample collection

The patient recruitment for the study was carried out at Lwala Hospital, Kaberamaido district, which is located in the *T. b. rhodesiense* focus of North Eastern Uganda. The patients were mainly from Kaberamaido and Dokolo districts (Fig 3.11D), which have a combined area of 2740km^2 and population of approximately 356,500 people (UBOS, 2010). The annual number of *T. b. rhodesiense* HAT cases reported from this site in 2010 and 2011 was 94 and 54 respectively. A total of 60 cases were reported in 2012, however for this study, patients were recruited between 1st April and 30th September 2012 and a total of 35 case samples were collected during this period (Fig 3.11A). Of these, 23 showed CNS invasion, 3 were only hemo-lymphatic and 9 were uninfected with *T. b. rhodesiense*. The hospital laboratory staff carried out the screening of suspected HAT cases by microscopic detection of motile trypanosomes on a wet blood smear of finger prick blood. If trypanosomes were present, a lumbar puncture was performed for disease staging. For microscopic detection of trypanosomes in CSF, a pellet obtained after centrifugation of the spinal fluid was used (Chappuis et al., 2005). Furthermore, standard assessment of co-infections including malaria, filariasis and voluntary testing for HIV/AIDS, were carried out. However the patients recruited in the study had neither of these co-infections detected. Following admission of the patients, the treatment regimen for early and late stage *T. b. rhodesiense* infections were Suramin (Antrypol: Bayer: 20 mg/kg) and melarsoprol (MelB) (Arsobal: Rhone-Poulenc: 3.6 mg/kg) respectively for approximately 10 days.

A protocol for the sample collection was designed in line with the routine clinical diagnostic procedures (Fig 3.11B). The buffy coat and CSF cellular fractions (containing White blood cells (WBCs) and Trypanosomes) were counted on a Neubauer haemocytometer and reported as cell count per ml (x10^4/ml). The cell pellets were then resuspended in Trifast (PeQlab) and immediately stored in liquid Nitrogen. The sample processing to storage took approximately 15min and it is worth noting that the cell counts were done after storage of the sample. I trained Ochieng, J., on the above procedure and he carried out the sample collection.

In order to confirm that all case samples collected were indeed *T. b. rhodesiense* infections, I carried out a PCR amplification of the serum resistance associated gene, SRA (Radwanska et al., 2002) on the patient blood that had been spotted on Whatman FTA cards (Fig 3.11C). All the samples were confirmed for *T. b. rhodesiense* infection due to the presence of the SRA gene, which is only found in this species.

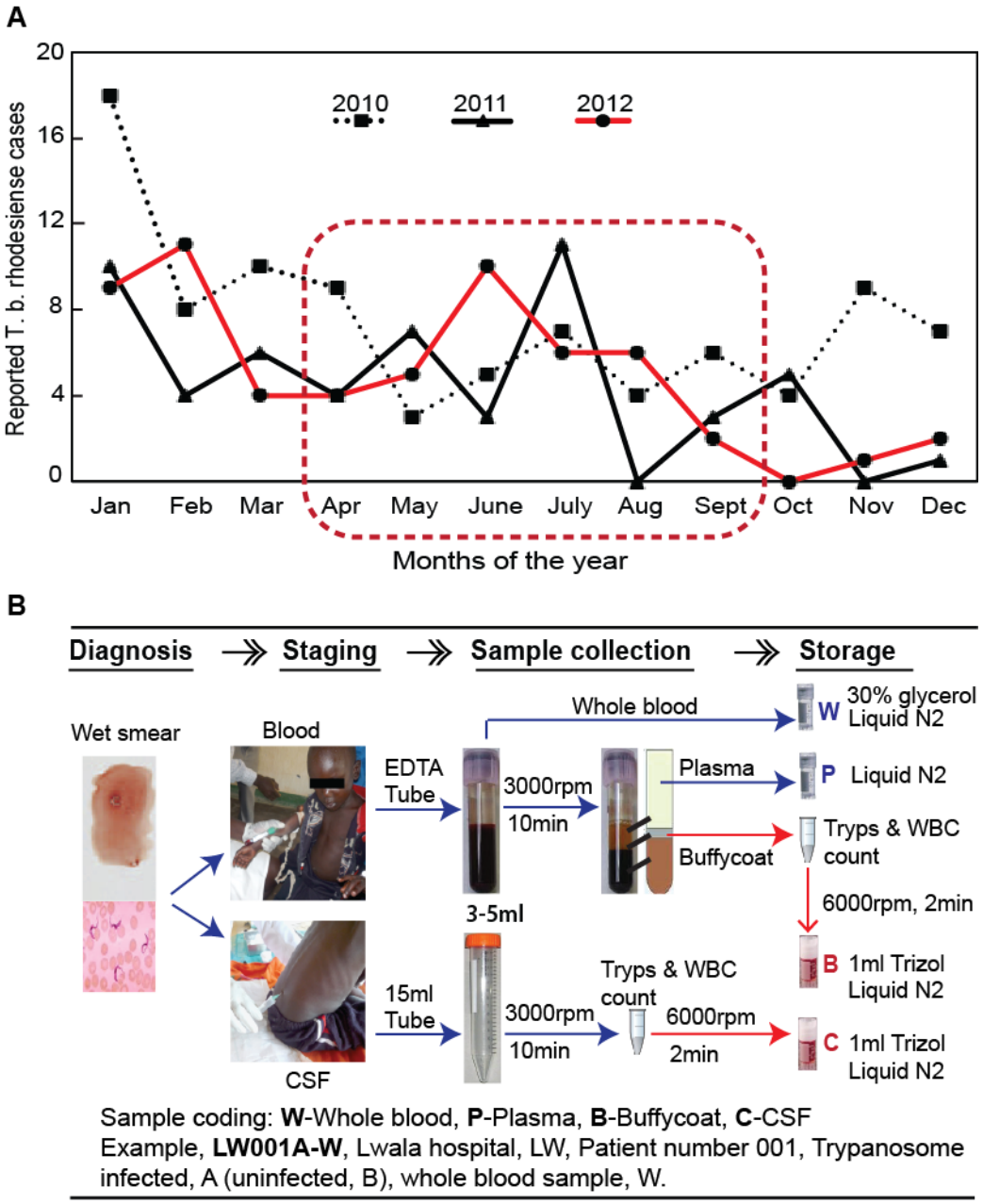


Figure 3.11 Study sample collection from Sleeping sickness patients. **A.** Annual HAT cases reported at Lwala hospital for the years, 2010, 2011 and 2012. A total of 26 case samples were collected for this study in 2012 between the months of April and September. **B.** Schematic representation of the protocol used to process and collect blood and CSF trypanosomes for the transcriptome analysis. **C.** Nested PCR of the SRA gene in DNA extracted from patient blood spotted on Whatman FTA cards. The 284 gene fragments confirms presence of *T. b. rhodesiense* species infection in the samples.

D

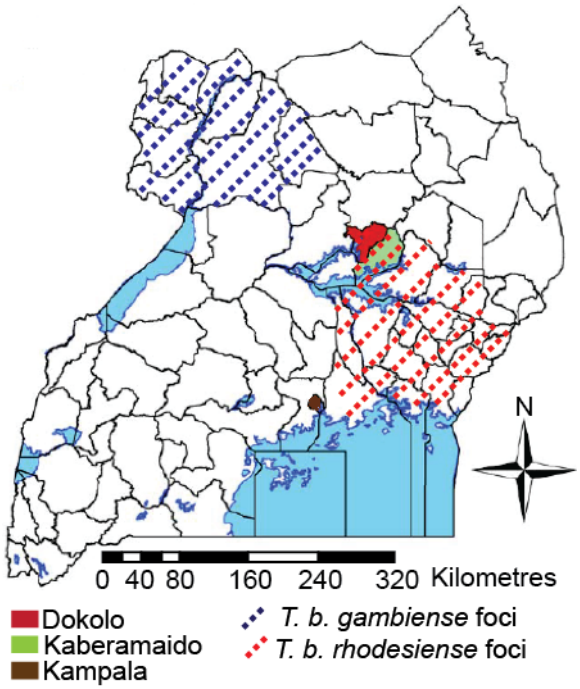
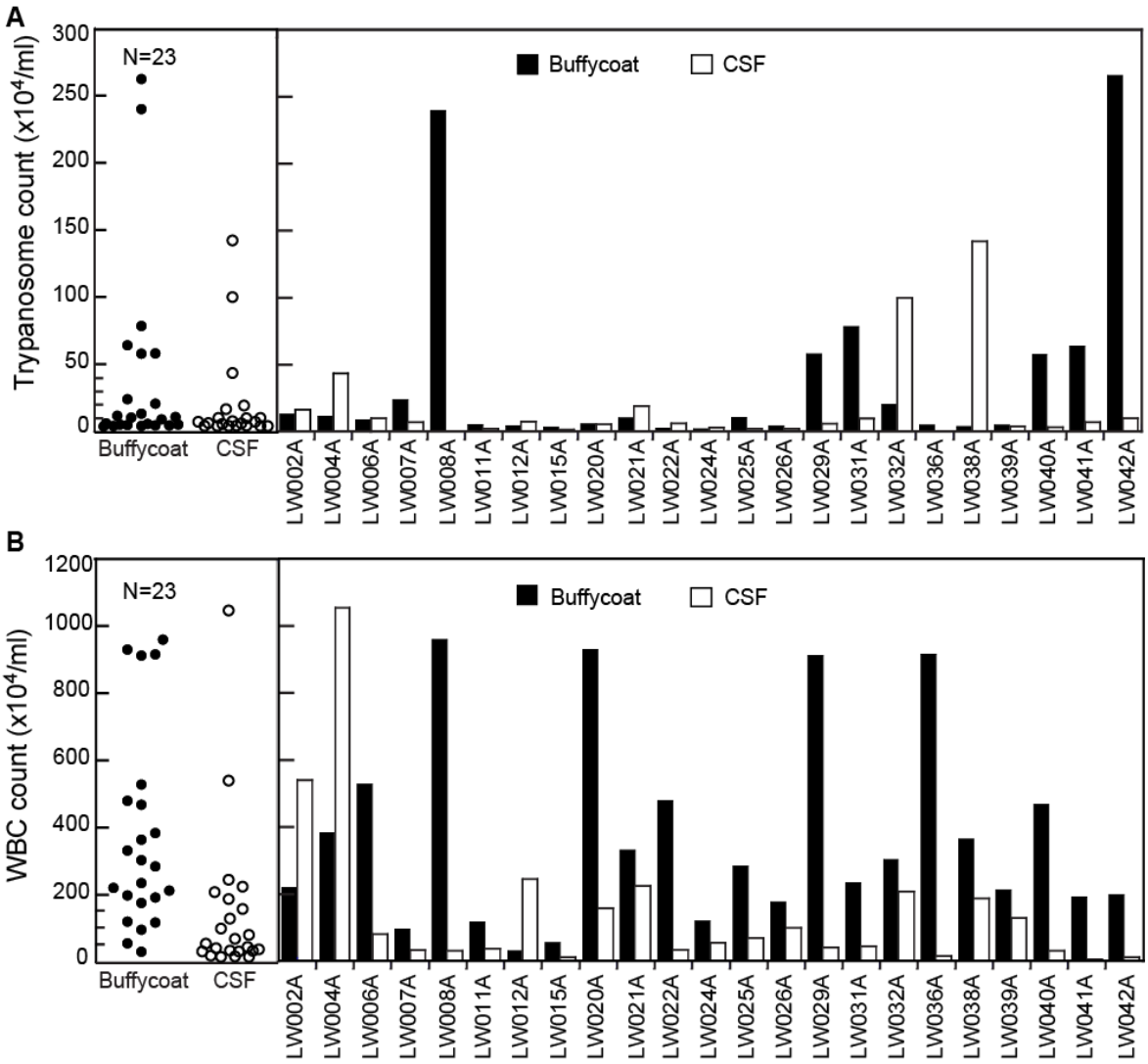


Figure 3.11D Map of Uganda highlighting the disease foci and region from where patients where recruited. Lwala hospital is located in Kaberamaido district. Samples were transported to Makerere University in Kampala for processing and further analysis. The map is modified from the original of (Batchelor et al., 2009).

3.2.2 Cell count analysis

Total cellular counts (WBCs and Trypanosomes) were carried out on the buffy coat and CSF fractions using a hemocytometer prior to pelleting them (Fig 3.12). The range of trypanosomes counted in the 23 patient samples that were analyzed was between $2x10^4$ and $266x10^4$ cells/ml in the buffy coat. The trypanosome count in the CSF fraction was between 0 and $142x10^4$ cells/ml (Fig 3.12A). But given the few cases of outliers who had a high trypanosome count, the median count in the buffy coat and CSF fractions was $10x10^4$ and $6x10^4$ cells/mL respectively (Table 2). On the other hand, the WBC count range was between 29 $x10^4$ and $960x10^4$ cells/ml in the buffy coat and 5 $x10^4$ to 1054 $x10^4$ cells/mL in the CSF (Fig 3.12B). In this case, the median count for the WBCs in the buffy coat and CSF fractions was $284x10^4$ and $54x10^4$ cells/mL respectively (Table 3.2). Except for patient LW08A who showed a high trypanosome count and correspondingly high WBC count in the buffy coat, there was no observed correlation between trypanosome and WBC counts in either buffy coat or CSF. The observed ratio of trypanosomes to WBCs was 1:28 in the buffy coat and 1: 9 in the CSF (considering the median count). During the amplification protocol development phase, we hypothesized a ratio of 1:1000 and this showed that in reality the actual ratio is much less.



	Cell count (x10⁴/ml)			
	Buffycoat		CSF	
N = 23	Tryps	WBC	Tryps	WBC
Maximum	266	960	142	1054
Minimum	2	29	0	5
Median	10	284	6	54
Mean	39	370	17	145

Table 3.2 Summary of cell count data from the 23 patients recruited for the study

Figure 3.12 Dot plot and column graph showing the cell counts in the samples collected from patients. **A.** Shows the cell count of trypanosomes in the buffy coat and CSF samples from the 23 patients studied. **B.** Shows the corresponding cell count of white blood cells (WBCs) in the same 23 patients.

3.2.3 Sequencing patient samples

The samples were transported from Uganda to Germany on dry ice for processing and sequencing. There was an unexpected delay of the samples at the German customs, at which point the ice melted and sampled thawed for a week.

The total RNA was extracted from the Trifast frozen samples using the chloroform method and quantified using the Qubit (Fig 3.13A). The total RNA in the 19 patient samples analyzed ranged from 0 to 656ng in the buffy coat fractions and 0 to 580ng in the CSF fractions. There were 13/19 of the samples that had RNA amounts >40ng in the buffy coat fractions, where as the CSF fractions had 8/19. However, the amount of RNA obtained from the samples was less than expected (Table 3.3). The expected RNA amount for each sample was calculated based on Eikmans et al., 2013, who showed that $1x10^6$ peripheral blood lymphocytes had an average of 2.8µg total RNA extracted by either Trizol or purification columns. This low yield of RNA was probably caused by degradation during either storage in Trifast reagent, thawing at the German customs, RNase activity during extraction procedure or loss of RNA pellet during the extraction procedure.

The feasibility of the amplification protocol for the patient samples was tested by second strand synthesis to detect trypanosome cDNA in samples LW026A (Fig 3.13B). There was detection of trypanosome cDNA in the CSF fraction in comparison to the buffy coat fraction, which had less RNA. Furthermore the RNA quality of this LW026A CSF sample was checked on the Aligent bioanalyzer (Fig 3.13C) for which traces of Human 28s:18s rRNA peaks were detected with a ratio of 1.6, signifying good quality. This suggested that the samples were probably not degraded. The amplification protocol was then checked on the LW042A buffy coat sample. Analysis of the trypanosome amplified cDNA showed a bioanalyser trace of DNA ranging from 300bp to >7kb which covers the entire transcriptome as observed in the protocol development phase (Fig 3.13D). The sheared cDNA showed a peak between 300 to 500bp that would be used for library preparation for sequencing (Fig 3.13E). At this point, library preparations for all the samples were made for the sequencing. A total of 21 amplified cDNA samples were processed for sequencing, that is, 12 buffy coat, 7 CSF, 1 pool of buffy coat (5 samples <5ng), 1 pool of CSF (10 samples <5ng). Unfortunately, except for buffy coat samples LW007 and LW042, the profiles from the rest of the samples were extremely variable (Fig 3.14). The majority of the samples had distinct peaks with high fluorescent units (concentration). The factors that could have affected the yield probably influenced the downstream amplification procedure resulting in this variability. However there was an unprecedented failure in making libraries from the two samples LW007 and LW042, resulting in a low yield, which could not be sequenced.

	Buffy coat				CSF			
Sample	**WBC count (10^6 /ml)**	**Tryps count (10^6 /ml)**	**Exp RNA (µg)**	**Obs RNA (ng)**	**WBC count (10^6/ml)**	**Tryps count (10^6 /ml)**	**Exp RNA (µg)**	**Obs RNA (ng)**
LW002A	2.2	0.13	6.16	ND	5.4	0.16	15.12	ND
LW006A	5.28	0.08	14.78	78.9	0.8	0.1	2.24	ND
LW007A	0.96	0.24	2.68	99.4	0.32	0.07	0.91	134
LW012A	0.29	0.04	0.83	ND	2.45	0.07	6.68	72
LW020A	9.30	0.06	26.05	42	1.57	0.05	4.40	ND
LW021A	3.32	0.10	9.29	438	2.24	0.19	6.27	129
LW022A	4.79	0.02	13.41	656	0.33	0.06	0.94	ND
LW024A	1.20	0.02	3.36	95.8	0.54	0.03	1.52	73.4
LW025A	2.84	0.10	7.95	ND	0.68	0.02	1.91	55.6
LW026A	1.76	0.04	4.92	ND	0.99	0.02	2.77	580
LW029A	9.12	0.58	25.53	75.4	0.41	0.06	1.14	ND
LW031A	2.34	0.78	6.56	258	0.44	0.09	1.25	ND
LW032A	3.02	0.20	8.47	ND	2.08	1	5.82	79.4
LW036A	9.16	0.05	25.64	ND	0.14	0	0.41	ND
LW038A	3.64	0.04	10.19	81	1.87	1.42	5.23	54
LW039A	2.12	0.05	5.93	52	1.29	0.04	3.61	ND
LW040A	4.68	0.57	13.10	157	0.30	0.03	0.86	ND
LW041A	1.92	0.64	5.37	56	0.05	0.07	0.14	ND
LW042A	1.98	2.66	5.54	228	0.11	0.1	0.31	ND

* Expected (Exp.) and Observed (Obs.) RNA amounts in the samples
* Expected RNA amount calculated from 2.8 ± 0.8 µg/ $1x10^6$ peripheral blood lymphocytes (Eikman et al., 2013).
* ND – Samples with RNA amounts less than 5ng/µl and not detectable by the Qubit.

Table 3.3 Comparison of the expected and observed total RNA amounts from the patient buffy coat and CSF samples.

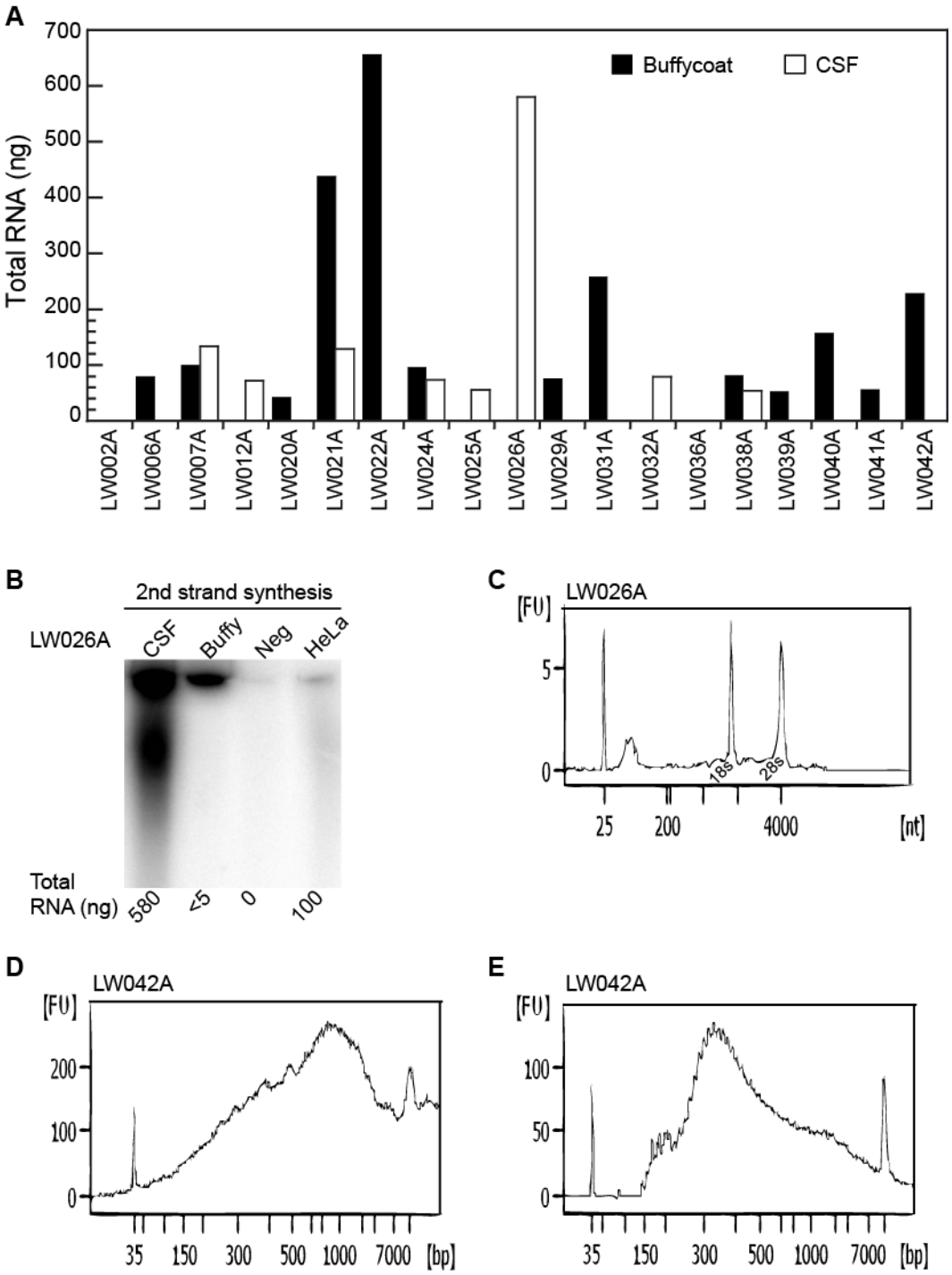


Figure 3.13 Analysis of RNA extracted from the patient samples. **A.** Total amount RNA after Trifast (Peqlab) extraction from the patient samples and quantified using the Qubit (Invitrogen). **B.** Detection of trypanosome cDNA in sample LW026A buffy coat and CSF by second strand synthesis analysis with α^{32}P-dCTP, resolved on 8% Urea-acrylamide gel and visualised by a phosphor-imager. **C.** Agilent Bioanalyzer trace of LW026A CSF-RNA highlighting the 18s and 28s human rRNA, showing that the RNA was of good quality. **D.** Agilent Bioanalyzer trace of amplified LW042A buffy coat cDNA showing the range of DNA transcripts. **E.** Agilent Bioanalyzer trace of the sheared LW042A buffy coat amplified cDNA for library preparation.

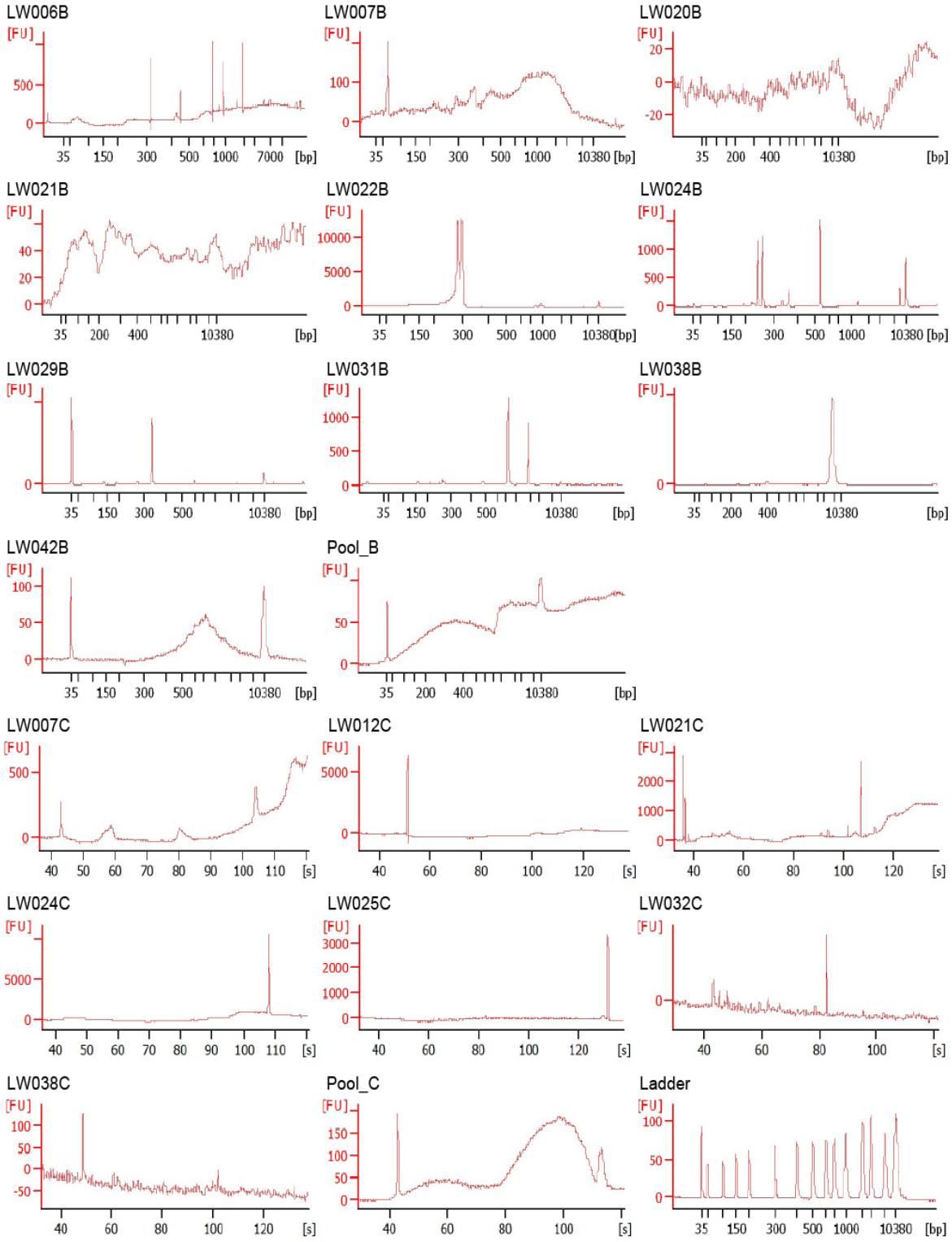


Figure 3.14 Agilent bioanalyzer traces of the amplified trypanosome transcriptomes from the patient samples. The samples are labelled based on the source, that is buffy coat **B** or CSF **C.** The y-axis represents a measure of the fluorescence units (FU) and x-axis the coverage in base pairs (bp) or duration for the coverage in second (s).

3.3 Transcriptome of purified *T. b. rhodesiense* trypanosomes

3.3.1 Comparison of transcriptome libraries

Besides the buffy coat, there are other methods through which trypanosomes can be isolated from blood namely, (a) DEAE cellulose chromatography (Lanham and Godfrey, 1970; Lonsdale-Eccles and Grab, 1987; Lumsden et al., 1979), in which the blood cells by bind to the positively charged DEAE (pH 8) and the trypanosomes are eluted out; (b) Red blood cell lysis (Njogu and Kiaira, 1982) followed by centrifugation, with the pellet fraction containing the trypanosomes and WBCs. These methods result in 10 times more parasite recovery in comparison to the buffy coat method hence could be used in the transcriptome analysis avoiding the need for amplification. However the extent to which these methods affect the transcriptome of the trypanosomes was not known. By comparing to the buffy coat isolated trypanosomes, which don't undergo any form of purification, I investigated the effect of these purification methods on the transcriptome of the bloodstream trypanosomes.

I infected rats with a *T. b. rhodesiense* strain (Tbr729) and the trypanosomes purified from whole blood (Buffy coat, BT, DEAE chromatography, DT, and RBC lysis, RT) for transcriptome analysis. Three biological replicates were performed for each isolation method, RNA extracted from the purified trypanosomes and processed for HiSeq2000 sequencing by selecting for the polyA+ RNA transcripts, randomly shearing this RNA followed by cDNA synthesis (Fig 3.4B). The library preparation was carried out by Ibberson, D., (Fig 3.4B). The number of reads mapping to the open reading frames (ORFs) in the *T. brucei* genome were calculated as reads per kilobase per million mapped reads (Rpkm). The data for individual unique open reading frames (ORFs) was then extracted in order to eliminate multi-copy ORFs and the transcriptome libraries from the different methods were then compared. I observed that the correlation between transcriptome libraries of trypanosomes from buffy coat was high (Fig 3.15A & 3.15B), implying that they were very similar and highly reproducible. The transcriptome libraries from the DEAE purified trypanosomes correlated well between replicates 1 and 2 (Fig 3.15C), however there was more distortion observed between replicates 1 and 3 (Fig 3.15D), especially for the low abundance transcripts. This suggested that the DEAE purified trypanosome transcriptome was not as reproducible as that for the buffy coat trypanosomes. The RBC lysis replicates had a high correlation (Fig 3.15E & 3.15F) suggesting a reproducible effect of the method on the transcriptome of these trypanosomes.

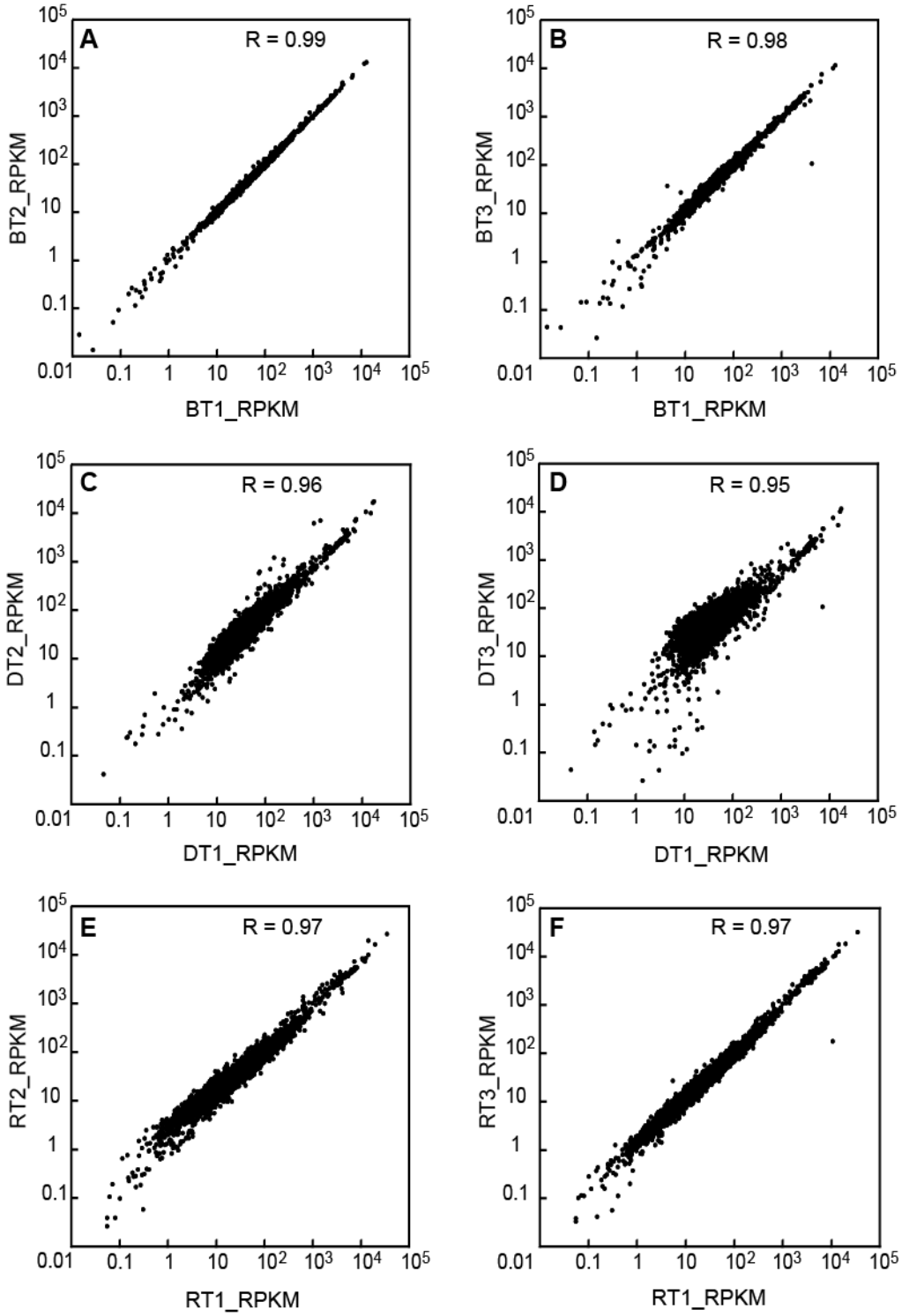


Figure 3.15 Effect of purification on the transcriptome of Trypanosomes. Scatter plot showing the correlation between the transcriptomes (RPKMs) of three biological replicates of trypanosomes obtained from the buffy coat (BT1, BT2, BT3) **A**-**B**, DEAE chromatography (DT1, DT2, DT3), **C**-**D** and RBC lysis (RT1, RT2, RT3), **E-F,** of *T. b. rhodesiense* infected rat blood. Each dot in the scatter plot represents a single gene.

Further analysis of reproducibility was a measure of the relative standard deviation between all three replicates of each isolation method. The mean standard deviations for the individual genes were 0.06 for the buffy coat, 0.3 for DEAE and 0.18 for the RBC lysis. In addition, it was observed that the buffy coat trypanosomes had 99% of their unique genes with relative standard deviation less than 30% (Fig 3.16A); whereas the DEAE (Fig 3.16B) and the RBC lysis (Fig 3.16C) had 55% and 83% respectively. This certainly confirmed that the transcriptomes from the buffy coat trypanosomes were more precise and reproducible. The buffy coat pattern was also comparable to that of the *in vitro* cultured bloodstream forms (Fig 3.16D), which also had 99% of their unique genes with relative standard deviation less than 30%. The RBC lysis trypanosomes also had a fairly reproducible transcriptome however, DEAE purified trypanosome transcriptomes were not precisely reproducible.

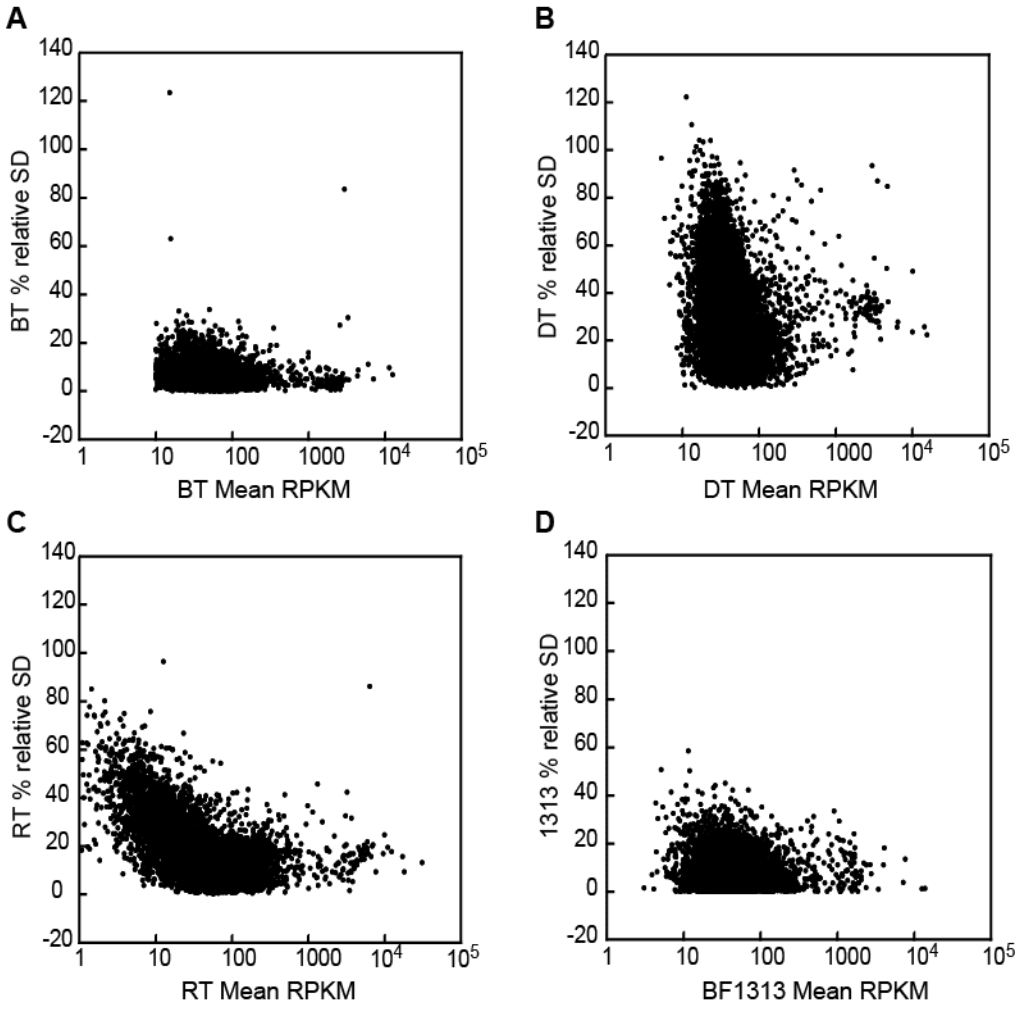


Figure 3.16 Scatter plots relating the percentage relative standard deviation to the mean RPKM of each gene (single dot). The mean of the three replicates of each isolation method was used for the calculation. Trypanosome transcriptomes obtained by **A**, Buffy coat, **B**, DEAE chromatography, **C**, RBC lysis and **D**, *In-vitro* cultured bloodstream form trypanosomes Lister 427 (BF1313).

3.3.2 Effect of purification on the transcriptome

In order to assess the effect of the purification methods on the transcriptome of the trypanosomes, I made a comparison between the purified (DEAE and RBC lysis) and "unpurified" (buffy coat) trypanosome transcriptomes (Fig 3.17). For this, the mean mRNA abundance for each gene in all three replicates was used for the comparison. The transcriptome of the DEAE purified trypanosomes had a good correlation with the buffy coat trypanosomes except for the low abundant transcripts (Fig 3.17A). However, the RBC lysis trypanosomes transcriptome did not correlate well with the buffy coat trypanosomes transcriptome, with even more variation observed with the low abundance transcripts (Fig 3.17A). This seemed to suggest that the DEAE purification procedure did not adversely affect the abundance of the transcripts in comparison to the RBC lysis method.

Furthermore, I analyzed for the expression of transcripts in the DEAE and RBC lysis transcriptomes relative to the buffy coat transcriptomes. This was quantified by calculating the ratio of the mean RPKMs (unique gene list) from the DEAE and RBC lysis, to the buffy coat (DT/BT and RT/BT). By excluding genes with a relative standard deviation greater than 50%, it was observed that 20% (1341/6532) of the unique genes in the RBC lysis treated trypanosomes showed more than 1.5 fold increase and 23% (1546/6532) had less than 0.5 fold decrease in abundance relative to those in the buffy coat. This implied that approximately 43% of the transcripts among the unique genes were affected by this treatment. On the other hand, for the DEAE treatment, only 1% (74/5655) of the unique genes showed 1.5 fold increase whereas none had less than 0.5 fold decrease.

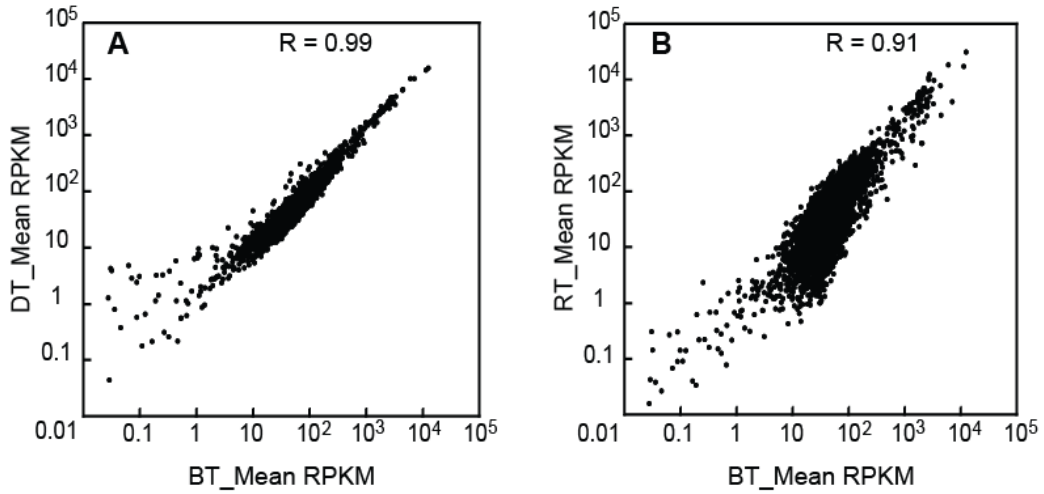


Figure 3.17 Gene expression analysis between purified and un-purified trypanosomes. Scatter plots comparing the average RPKMs of trypanosome transcripts extracted by the buffy coat method to DEAE chromatography **A** and RBC lysis method, **B**.

To further establish the statistical significance of the observed relative expression between the methods, Merce, C., carried out a differential gene expression analysis. For this, the Fisher's exact test was used for which the calculated p-values of the transcripts were manually assigned functional categories. The functional category enrichment between methods was then visualised using a heat map generated in R (Fig 3.18). There were no differentially expressed genes detected between the DEAE and Buffy coat methods probably because their transcript abundances were quite similar. However there was a significant increase in transcripts coding for ribosomal proteins in the RBC lysis treatment. Other mRNAs encoding proteins for mitochondrial pathway, chaperones, redox pathway and RNA processing were also increased by this treatment. There was also an observed decrease in the mRNAs encoding for protein kinases, cytoskeleton, mitochondrial DNA metabolism proteins, RNA helicase and many more. This showed that purification by RBC lysis greatly affected the transcriptome of the trypanosomes. However, purification by the DEAE method did not have such a significant effect on the abundance of mRNAs.

An analysis of transcripts with an abundance of less than 5 Rpkm showed that trypanosomes from the buffy coat had 91, the DEAE, 72 and RBC lysis, 413, out of the 6773 unique genes. This highlighted the low abundance transcripts that could potentially be lost during the course of trypanosome isolation by either method. And indeed the RBC lysis yielded the most, implying that it would not be a method of choice.

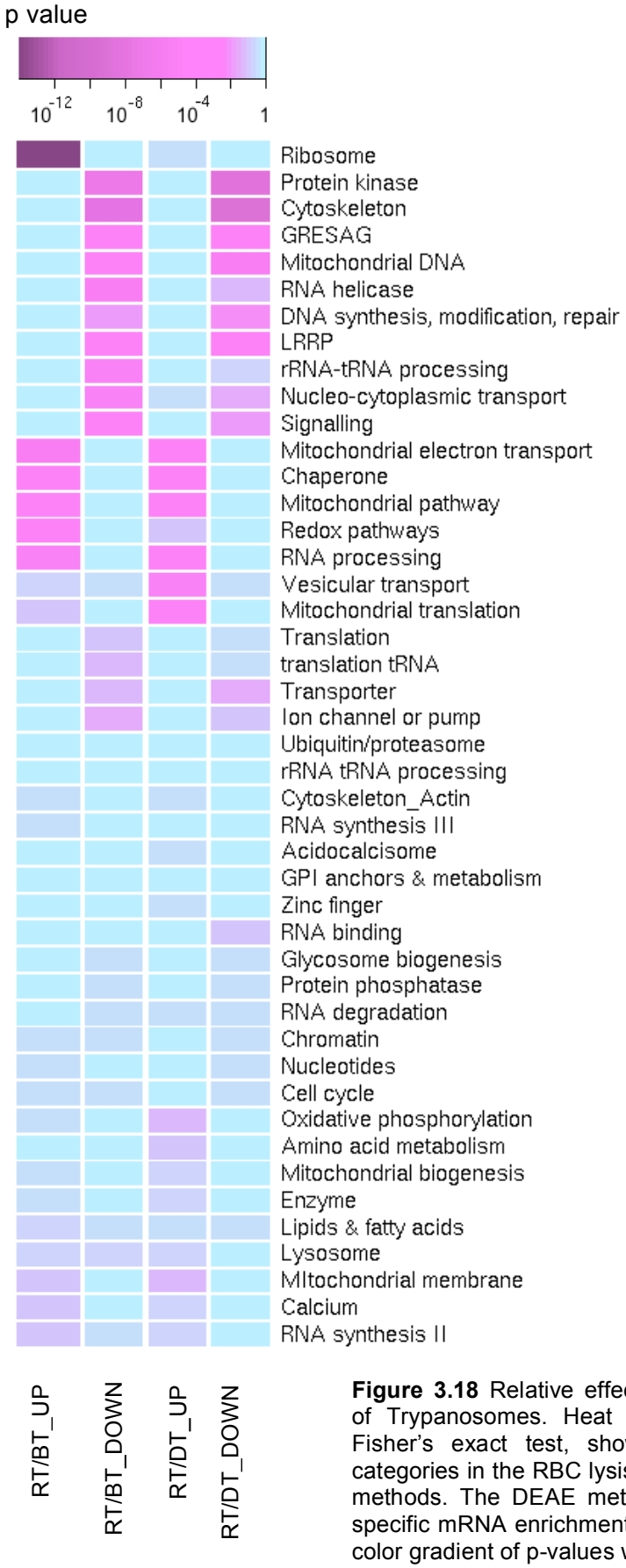


Figure 3.18 Relative effect of purification methods on transcriptome of Trypanosomes. Heat map plot using R of p-values from the Fisher's exact test, showing the enrichment of specific mRNA categories in the RBC lysis method relative to the buffy coat or DEAE methods. The DEAE method relative to the buffy coat showed no specific mRNA enrichment hence not shown. The scale indicates the color gradient of p-values with 1 being no effect.

3.4 Genomic analysis of *T. b. rhodesiense* isolates from patients

3.4.1 Comparison between strain libraries

In order to investigate the structural variation in the *T. b. rhodesiense* strains obtained from the patients, I inoculated rats with cryopreserved infected blood stabilates in order to propagate the parasites for genomic DNA extraction. Four patient isolates, LW024, LW031, LW032 and LW041 were selected for DNA sequencing. They were the isolates that were able to grow to a density sufficient for genomic DNA extraction. The DNA library preparation was carried out by Ibberson, D., multiplexed and sequenced on the Illumina HiSeq2000. For comparison, the *in vitro* cultured *T. brucei* Lister 427 bloodstream (BF1313) and procyclic (Pc2060) forms, and the pleomorphic Antat1.1 were also sequenced. These strains express the tet repressor and have been growing in culture intermittently since 1990 undergoing clonal selection in the process. The Antat1.1 strain has the ability to differentiate into procyclic forms.

Merce, C., aligned the 100bp reads to the *T. brucei* TRUE927 reference genome using Bowtie2 (Li and Durbin, 2009) and the output is summarized in table 3.3. The reads were sorted using the SAM tools (Li et al 2009) and the reads per kilo base per million mapped reads (Rpkm) calculated for each transcript. These Rpkm values were used to measure the gene copy number and not single nucleotide polymorphisms (SNPs). A comparison between the genomes showed a high correlation of 0.99 for the *T. b. rhodesiense* strains obtained from the patients (Fig 3.19 A, B). This implied that the strains were almost identical. However there was a lower correlation between the *in vitro* cultured *T. b. brucei* BF1313 and the field isolated *T. b. rhodesiense* LW031 implying that there were subtle differences between them (Fig 3.19C). A similar phenomenon was observed between the pleomorphic Antat1.1 and the *T. b. rhodesiense* LW031 (Fig 3.19D). The comparison between the two cultured strains of the monomorphic BF1313 and pleomorphic Antat1.1 showed a correlation of 0.95 implying that the two had differences even though they are both *T. brucei* strains (Fig 3.19E). Furthermore, the *in vitro* cultured BF1313 and procyclic form 2060 were more similar (Fig 3.19F) in comparison to the correlation with the Antat1.1 (Fig 3.19E). The Pc2060 is the differentiated procyclic form of BF1313, hence they have the same genome and only difference is at the transcriptome level.

Sample	Total Reads	Reads Aligned to Tb927	% Aligned
LW031	35,361,888	29,045,683	82%
LW041	51,994,734	41,540,615	80%
LW032	54,616,570	43,669,502	80%
LW024	50,374,813	40,515,524	80%
Bf1313	35,863,223	28,577,153	79%
Antat 1.1	32,876,011	24,927,041	76%
Pc2060	36,864,835	27,162,418	74%

Table 3.3 Summary of sequenced 100bp paired end reads from genomes from Trypanosome strains. The patient *T. b. rhodesiense* isolates (LW031, LW041, LW032, LW024), *in vitro* cultured *T. brucei* 427 strains, bloodstream form (Bf1313) and procyclic form (Pc2060), and pleomorphic *T. brucei* strain (Antat 1.1). The reads were mapped on to the *T. brucei* TREU 927 reference genome.

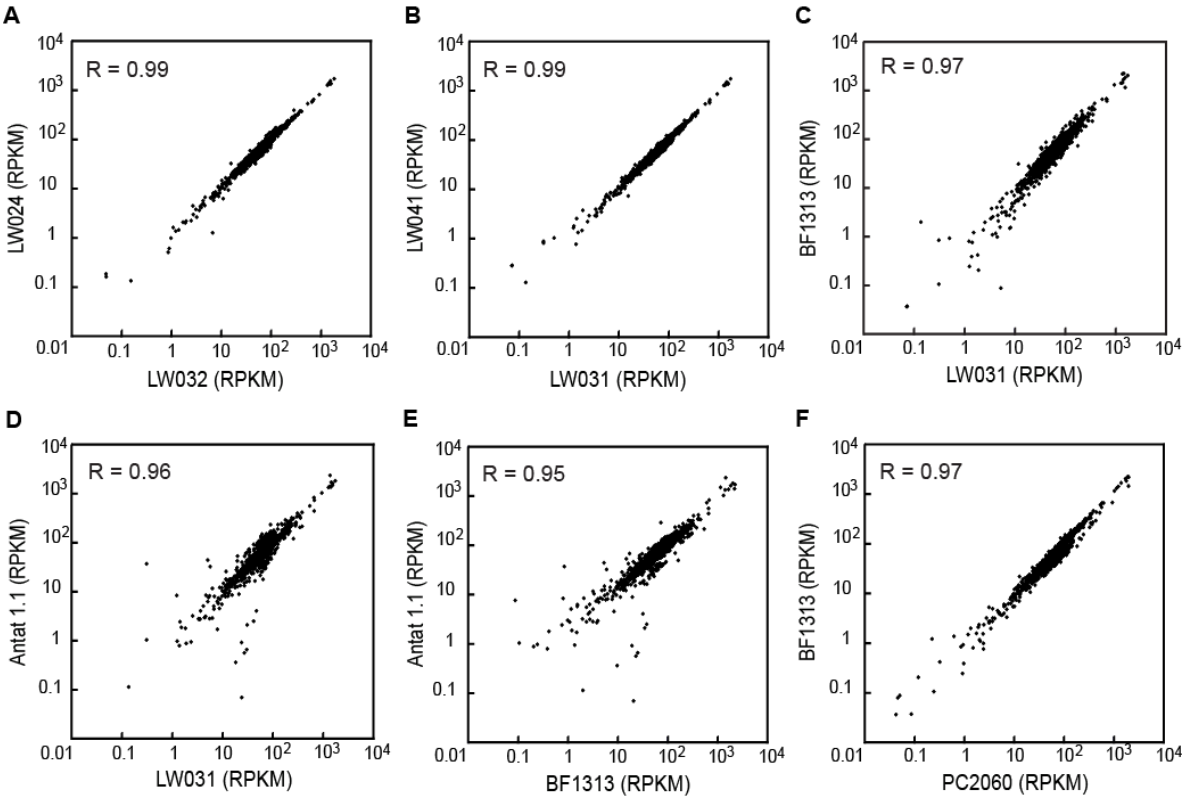


Figure 3.19 Alignment of genomic reads to the *T. brucei* 927 reference genome. Correlation of Rpkm values between the *T. b. rhodesiense* isolates from patients LW024 and LW032, **A**, then LW031 and LW041, **B**. Comparison between the field isolate LW031 and the *in vitro* cultured *T. b. brucei* bloodstream form BF1313, **C**, and Antat1.1, **D.** Comparison between the *in vitro* cultured *T. brucei* strains BF1313 with pleomorphic Antat1.1, **E**, and with procyclic form PC2060, **F**.

3.4.2 Gene copy number variation

In order to analyze for heterogeneity in the field isolates as well as the laboratory cultured strains, the number of copies per gene in the whole genome was determined by Merce, C. First was to determine the gene copy number of the reference *T. brucei* TREU927 genome. This was carried out by aligning the whole genome sequence reads of *T. brucei* 927 (downloaded from the sanger institute data base) to itself (*T. brucei* TRUE927 genome) using Bowtie2, allowing 1 and 300 mappings per read (2 alignments, k1 and k300). Next was to determine the modal Rpkm value (most frequently occurring Rpkm value, Fig 3.20A) that we used to calculate the number of copies per gene, by dividing the individual gene Rpkm by this mode value (Rpkm/mode). The reads from the strains were then aligned to the TRUE927 reference genome allowing 1 mapping per read and also determined the modal value for each strain data set. The gene copy number for each strain was then determined relative to the copy number obtained from the reference reads alignment.

Most of the genes in the clinical isolates had the same copy number however there were a number of subtle differences observed in the *in vitro* cultured strains. The major differences in all the isolates were observed in a few selected multi copy genes (Fig 3.20B). In comparison to all isolates, the cultured strain Bf1313 showed the highest copy numbers for the genes encoding the histone proteins (H4, H3, H2A), elongation factor 1-alpha (TEF1), the chaperone HSP70, paraflagellar rod component (PFR1) and PTP1 interacting protein (PIP39). The S-adenosylmethionine synthetase, putative (METK1) gene showed the same copy number in three of the four field isolates (LW031, LW041, LW032) whereas there was one and two extra copies in the Bf1313 and Antat1.1 strains respectively. Furthermore, the field isolates also showed differences within these multi-copy genes implying that they were heterogeneous, in spite of having a high correlation between them. The field isolate LW032 had a higher copy number for a number of genes where as LW024 had the least copy number amongst all the four isolates.

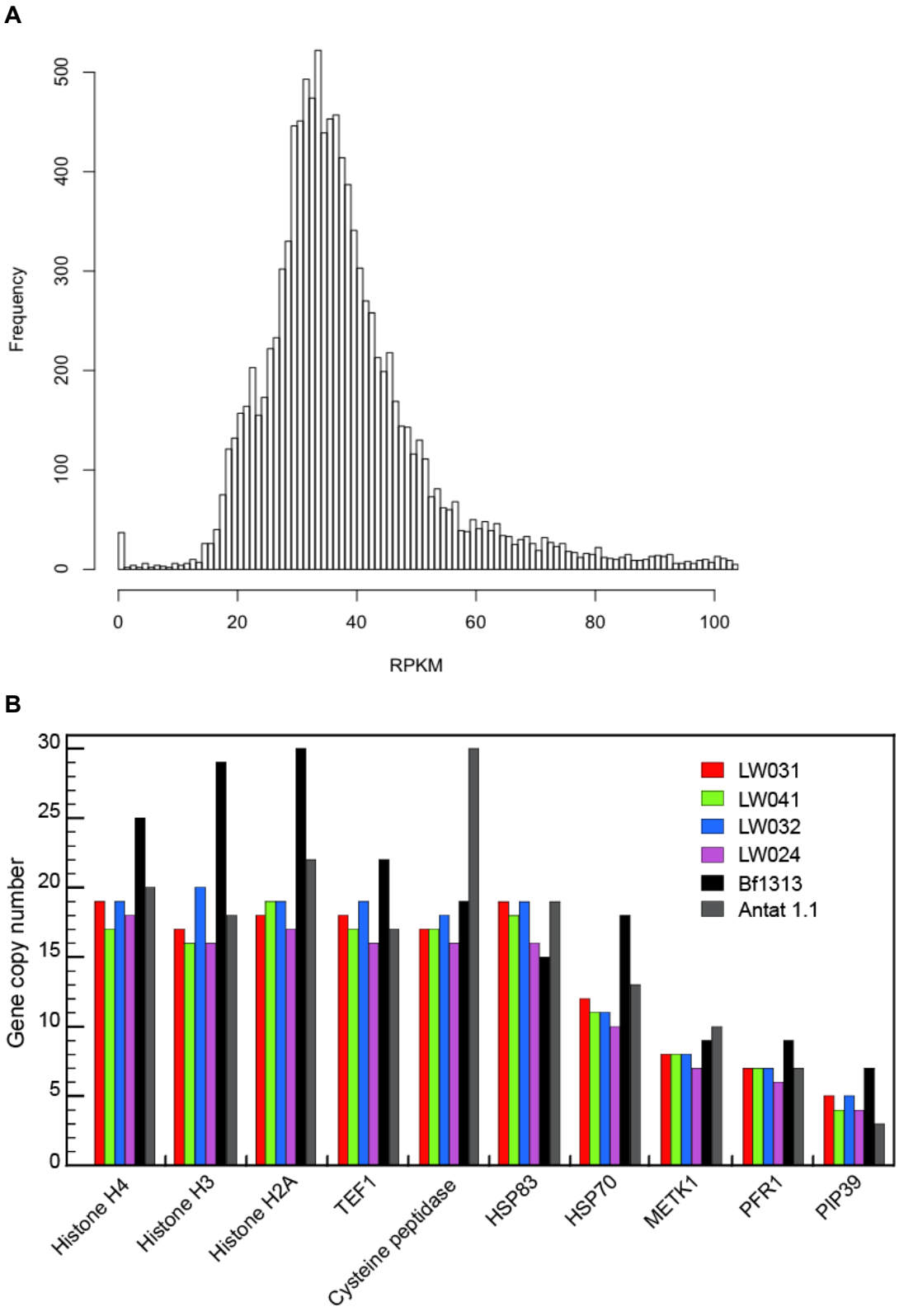


Figure 3.20 Gene copy number variation in *T. brucei* strains. **A.** Histogram plot of the frequency of gene occurrence against the Rpkm for the reference genome reads, TRUE927. **B.** Analysis of multi copy number genes in the *T. b. rhodesiense* field isolates (LW031, LW041, LW032, LW024) and *T. brucei* strain cultured *in vitro*, Lister 427 (Bf1313, Antat1.1).

4. Discussion

Gene expression studies on African trypanosomes have been carried out by high throughput microarrays ((Jensen et al., 2009; Kabani et al. 2009; Queiroz et al. 2009) or by next generation high throughput sequencing platforms (Nilsson et al., 2010; Kolev et al., 2010; Siegel et al., 2010 Veitch et al., 2010). These studies were done on cultured trypanosomes. This advancement in expression analysis provided new insights into the biology of trypanosomes however, the extent to which this phenomenon is representative of a real human infection is still not known. Therefore this study was aimed at analyzing the transcriptomes of clinical isolates of *T. b. rhodesiense* from sleeping sickness patients using high throughput RNA sequencing technology. This transcriptomics approach could be used to study pathophysiological mechanisms and identification of disease biomarkers (Le-Niculescu et al., 2009).

4.1 The Transcriptome of *T. brucei*

4.1.1 Amplification of nanogram *T. brucei* RNA in microgram Human RNA

Clinical samples obtained *ex vivo* from patients don't often yield enough RNA for high throughput gene expression analysis (Marteau et al., 2005, Mitsuhashi et al., 2006). During the course of human infection, there are often few circulating trypanosomes in the blood and CSF (Mumba Ngoyi et al., 2013). Given this low parasitaemia, we hypothesized that the ratio of trypanosome to human cellular total RNA in clinical samples of blood or CSF would be 1 to 1000. Therefore in order to mimic the amounts in the *ex vivo* sample, we mixed 5ng of trypanosome total RNA with 5µg of HeLa total RNA. But this amount of total RNA was 4000 fold less than what was used for the recently published trypanosome high throughput expression analysis ((Nilsson et al., 2010; Kolev et al., 2010; Siegel et al., 2010). Hence, there was a need to specifically amplify the nanogram amounts of trypanosome RNA in the microgram human RNA background to an amount sufficient for the Illumina sequencing platform. For this, I rationally designed an exponential amplification based method (Iscove et al., 2002, Livesey, 2003) utilizing the splice leader sequence located at the 5' end of all trypanosome mRNA transcripts (Liang et al., 2003). This amplification-based method has been successfully used to analyze the transcriptome of single cells (Kurimoto et al., 2007, Tang et al., 2009, Tang et al., 2011), so I adapted the same principles to specifically amplify the trypanosome RNA from the human RNA background. An alternative method would have been linear

amplification by *in vitro* transcription, IVT (Van Gelder et al., 1990, Stirewalt et al., 2003). However, there was 1000 fold more HeLa RNA than Trypanosome RNA in the heterogeneous mixture and multiple rounds of *in vitro* transcription would have increased this nonspecific background; since all the polyA selected mRNA fragments (Trypanosome and HeLa) had the T3 promoter sequence tag at the 3' end. Another reason why this was not the method of choice was that cRNAs less than 1 Kb are generated and that every round of IVT can amplify the cDNAs only up to 1000-fold (Livesey, 2003, Kawasaki, 2004).

Having optimized this exponential amplification-based approach, a size distribution ranging from 35 base pairs to 10Kb of the amplified cDNAs was traced on the bioanalyzer. The smallest gene in the trypanosome is approximately 30bp (Tb972.9.15030) coding for the DNA polymerase epsilon subunit b, whereas the largest is 19.8kb (Tb927.4.310) coding for a conserved hypothetical protein (tritryp db). This suggested that the majority of the expressed genes were captured by the developed method, giving a global representation of the transcriptome. This observed coverage of genes was probably due to the enhanced full length cDNA synthesis procedure which included; use of the tagged random primers (in addition to the oligo(dT) primer) and T4gene32 protein, which reduces the formation of higher order structures of RNA molecules thus eliminating any pause sites during the course of cDNA synthesis (Rapley, 1994, Villalva et al., 2001, Xiang et al., 2003). These factors therefore ensured an efficient read through by the reverse transcriptase in synthesis of the cDNA. Furthermore, by using a 21mer splice leader oligonucleotide, which is 4 bases upstream of the full length splice leader sequence, the specific synthesis of the double stranded cDNA from trypanosome cDNA transcripts in the HeLa background was enhanced. The sensitivity of this reaction was further improved by lowering the concentration of this splice leader primer and prolonging the annealing and extension times to 60 minutes. This adjustment in conditions probably enhanced the hybridization and selectivity properties of the splice leader primer for trypanosome sequences in the tryphanosome-HeLa mix. By using a 20mer splice leader (nested) primer and T3 promoter primer, a 10 cycle PCR ensued in specific amplification of trypanosome DNA with a minimal human DNA background. Limiting the amplification cycles to 10 was meant to minimize the effects of saturation, which would lead to bias towards the more abundant transcripts, resulting in loss of proportionality of the amplification process (Wang et al., 2004). Overall, this optimized amplification procedure gave an average of 120ng of DNA which was sufficient for Illumina sequencing.

4.1.2 Gene expression analysis of amplified nanogram *T. brucei* RNA

Two Illumina platforms that is, Illumina Genome analyzer, GAIIx and Illumina HiSeq2000 (http//www.illumina.com) were used to generate reads from the four technical replicates of the amplified and unamplified transcriptomes. Overall, the replicates from the HiSeq2000 yielded 5 times more reads than the GAIIx. According to Illumina, the HiSEq2000 system is shown to have a maximum output of 600Gb composed of 3 billion single-end reads whereas the GAIIx has a maximum output of 95Gb of 320 million single-end reads. This could probably explain the fold difference in the replicates between the two platforms. The technical replicates from either platform were highly reproducible for both the amplified and unamplified transcriptome libraries.

By considering the HiSeq2000 data, it was observed that over 20% reads from the amplified Trypanosoma-HeLa (5ng:5µg) transcriptome library mapped to the *T. brucei* TRUE97 genome, representing a 200-fold enrichment over the input. There were as 83% reads of the unamplified trypanosoma transcriptome library that mapped to the genome. Even though there was a much less density of reads that aligned to the genome from the amplified nanogram RNA trypanosome-HeLa library, these reads were able to cover the entire genome with profiles which were comparable to the unamplified trypanosome cDNA. Furthermore, the reads obtained were not directly proportional to the mRNA abundance, but were instead strongly influenced by the splice leader priming and also to a lesser extent, by the PCR amplification. The transcriptome libraries from the amplification method were reproducible and comparable to the unamplified method but not to the fragmented mRNA method. This was probably due to the technical similarity (splice leader priming) in the methodologies for generating the trypanosome transcriptome from the amplified and unamplified cDNA, as compared to the fragmented RNA method. This was suggestive of a methodological bias, which was skewed towards the shorter transcripts that are between 0.1 – 1Kb in length. These transcripts had, on average, high RPKMs in both amplified and unamplified transcripts, signifying more expression relative to the fragmented RNA transcripts. This bias was probably due to the enhanced efficiency of the splice leader primed cDNA synthesis by the DNA polymerase towards the shorter than the longer fragments. By analysing a unique list of 6772 single copy genes (Siegel et al., 2010), it was observed that 60% of the transcripts amplified from the 5ng:5µg trypanosome-HeLa mix had an expression threshold of 5 RPKM. The unamplified and fragmented mRNA transcriptomes had 80% and 98% respectively.

The standard RNASeq approach (fragmented RNA method) gives a direct measure of the mRNA abundance (Roberts et al., 2011). In order to obtain "real" mRNA abundances from the amplification method, an algorithm would be required to simulate and relate the amplified and unamplified transcriptome libraries to the fragmented RNA library. Never the less, the amplification method showed ample coverage of the transcriptome starting with only 5ng of trypanosome RNA, implying that it can be used for samples that contain about 10^4 parasites. Such is the case with *T. b. rhodesiense* sleeping sickness patients where we observed parasitaemias (median) of $1x10^5$ cells/ml and $6x10^4$ cells/ml in the buffy coat and CSF respectively. Even much less numbers are observed in the chronic *T. b. gambiense* sleeping sickness patients (Mumba Ngoyi et al., 2013), implying that amplification would be inevitable for their transcriptome analysis. In addition, this amplification method could be used in the study of other Kinetoplastid natural infections in which the parasites have mRNA with a splice leader such as, *T. congolense* and *T. vivax* infected cattle, and also samples of infected vectors (Tsetse fly, Sand fly).

4.1.3 Transcriptome of clinical isolates of *T. brucei rhodesiense*

Considering the population at risk of *T. b. rhodesiense* infection in the study region (Population of Dokolo and Kaberamaido districts), the prevalence rate was reported at 0.017% (60/365,500) in the study period (2012). The average prevalence rate reported in the same risk population reported in 2010, 2011 and 2012 was 0.019%. In this study, 23 patients were recruited between April and September 2012. These patient samples had a median trypanosome count of $10x10^4$ and $6x10^4$ cells/ml in the buffy coat and CSF samples respectively. The routine screening of *T. b rhodesiense* sleeping sickness patients involves microscopic observation of wet smears (Chappius et al., 2005). This method resulted in under reporting of the actual parasitaemia (5000 parasites/ml of blood), from which we derived our working hypothesis at the start of this study. However, we observed a higher parasite count than expected because of the concentration by centrifugation for both the buffy coat and CSF, before counting on the heamocytometer. The ratio of trypanosomes to WBCs was higher in the buffy coat (1:28) than in the CSF (1:9). In addition, there was no correlation between the parasitaemia and the WBC count in either buffy coat or CSF from the individual patient samples. One would expect a high parasitaemia to coincide with a high WBC count following the cellular immune response mounted by the patient's immune system (Sternberg, 2004). However the parasite's state of antigenic variation (Pays et al., 2004) and the poor nutritional status of 90% of the patients recruited (Kuepfer et al., 2011), could have resulted in the observed

immunodeficiency. However this phenomenon would need to be validated by in depth clinical analysis of the individual patients and also cytokine analysis (MacLean et al., 2001b; MacLean et al., 2010).

Buffy coat and CSF cellular pellets from infected patients were collected and stored. The time taken for sample collection to storage was 15min, which was crucial to avoid RNA degradation *ex vivo* (Fleige and Pfaffl, 2006). Since we were dealing with few cells in the clinical samples, a single extraction reagent (phenol-based Trizol Reagent) was used for storage and purification of RNA for subsequent analyses (Chomczynski and Sacchi, 2006; Hummon et al., 2007). However, the total RNA extracted from the Trifast frozen samples was much less than expected. It was assumed that storage of the fresh samples in Trifast reagent (Trizol) could have degraded it. However studies have shown that storage of fresh blood cells in trizol preserves RNA yield and quality (Eikmans et al., 2013; Kang et al., 2011; Ma et al., 2010). The prolonged thawing of samples at the German customs could have resulted in degradation of the samples. Inadequate shipping and handling of samples could easily cause degradation of RNA (Perez-Novo et al., 2005). However an RNA check carried out on four samples in Uganda (before shipment) also showed low RNA yield. Therefore this meant that the observed low RNA yield was probably due to loss of the RNA pellet during the precipitation step of the extraction, or due to renatured RNases.

The other alternatives to the Trizol procedure would be the use of *ex vivo* RNA stabilizing procedures such as the whole blood PAXgene tubes (PreAnalytix, Qiagen) for sample collection, transport and storage (Feezor et al., 2004; Rainen et al., 2002; Thach et al., 2003). In addition, this PAXgene system enables samples to be frozen for up to 2 years without affecting the expression profile (Ovstebo et al., 2007). Considering the field hospital conditions, the PAXgene system offers an advantage since the RNA in whole blood has been shown to be stable at room temperature for 5 days, following storage for up to 12 months at -20°C and -80°C, and also after repeated freeze-thaw cycles (Rainen et al., 2002). But one disadvantage with this method is that, cell counting is not possible because cells are lysed directly after collection. In addition, the globin transcripts from the reticulocytes may limit the sensitivity of gene expression profiling experiments (Debey et al., 2004), since these constitute up to 70% of the total whole blood mRNA population (Field et al., 2007; Mastrokolias et al., 2012). This would necessitate a globin mRNA removal step prior to polyA+ mRNA selection, which might reduce on the overall yield of the target RNA.

An advantage of total RNA extraction using the Paxgene is that it would yield both the trypanosome and lymphocyte transcriptomes. For patient blood samples with a parasite: lymphocyte ratio of 1:10 ($4x10^5$: $4x10^6$ cells/ml), total RNA extracted from this whole blood and sequenced on a single lane would result in approximately 1 million reads mapping to the trypanosome. A rather expensive alternative would be applying the sequencing library to several lanes to get more reads. But in addition to trypanosome transcripts, lymphocyte transcripts from the same sample would also be mapped to the human genome hence determining the immunological outcome from parasite infection, at the transcriptome level.

Another alternative to Trizol would be the use of procedures used in isolation of RNA from very few/ single cells such as those from laser capture micro dissection (LCM), for instance the Picopure RNA isolation kit (Mikulowska-Mennis et al., 2002), could be used to recover the required yield of RNA from these clinical buffy coat and CFS cell pellets.

4.2 Effect of purification method on transcriptome of *T. b. rhodesiense*

The purification of trypanosomes from blood is employed in parasitological confirmation during the diagnosis of trypanosome infections. Increased sensitivity for detection is obtained through concentration techniques such as Quantitative buffy coat (Ancelle et al., 1997), miniature anion-exchange centrifugation, mAECT (Lumsden et al., 1979), and reticulocyte lysis (Njogu and Kiaira, 1982). In addition, methods such as DEAE chromatography (similar to mAECT) are used in the purification of trypanosomes from host blood in experimental studies of trypanosome biology. For the analysis of transcriptomes of trypanosomes from patient blood, purification of the trypanosomes would result in increased recovery of parasites and certainly reduce on the host's cells in the sample. The DEAE purification would be the method of choice for transcriptome studies since reticulocyte lysis results in major distortions in the mRNA abundance. The DEAE purification method can therefore be used as an alternative to amplification especially for samples with a relatively high parasite count. This method also has the advantage of reducing the lymphocyte cellular background from the purified parasites, which would be good for those samples with a high WBC to parasite count. Since standard Illumina library preparations work best with at least 50ng of mRNA sample, this could be obtained from at least $2x10^6$ trypanosomes. However the results from this DEAE purification would be less reproducible in comparison to the buffy coat (unpurified) trypanosomes from the different patients. Under field conditions in the clinical set up, it would take

no less than 40min to process a sample by DEAE purification, whereas the buffy coat would conveniently take about 15min.

4.3 *T. b. rhodesiense* isolates heterogeneity

The genomes of four *T. b. rhodesiense* strains isolated from patients showed that even though the sequenced libraries were highly similar, there were variations in the gene copy numbers between the isolates. The *in vitro* cultured *T. brucei* 427 strains had a significant difference to the field isolates with respect to the multi copy genes. Most notably were the genes coding for the histone proteins for which there were 5 to 10 more copies in the cultured *T. brucei* 427 strains. These proteins play a crucial role in transcription initiation and termination (Siegel et al., 2009), and continuous culturing of these parasites probably evolved their transcription landscape with respect to the *in vitro* environment resulting in selection of certain growth phenotypes (such as, high growth rate, non stumpy formation). Further analysis of structural variations in the field isolates was to be carried out. Identifying of single nucleotide polymorphisms would enable genotyping these recent isolates and compare them to the past isolates from the same disease foci (Goodhead et al., 2013). Furthermore, the virulence phenotype of these isolates should have been carried out in a mouse infection. This would enable the association of the genotypes with the virulence pattern and possibly the clinical picture in the patients.

4.4 Conclusion

The splice leader primed amplification method is an approach that can be used to determine the transcriptome of Kinetoplastids in the course of natural infection in the mammalian or invertebrate host. Sample collection from these hosts should be carried out using RNA stabilizing agents since Trizol storage seems to adversely affect the transcripts in low cell numbers. Upon generating the mRNA abundance profiles, a correction algorithm is then required to relate the amplified libraries to the real (unamplified, mRNA fragmented) libraries. This would thus enable the relative comparison between different samples.

Manuscript

Compilation of the methods data into research article for publication;

Mulindwa, J., Fadda, A., Merce, C., Matovu, E., Enyaru, J., Clayton, C. 2013. Methods to determine the transcriptomes of trypanosomes in clinical samples and other natural infections: the effects of parasite purification and selective cDNA amplification.

5. References

Agranoff, D., A. Stich, P. Abel, and S. Krishna. 2005. Proteomic fingerprinting for the diagnosis of human African trypanosomiasis. *Trends Parasitol*. 21:154-157.

Aksoy, S. 2003. Control of tsetse flies and trypanosomes using molecular genetics. *Vet Parasitol*. 115:125-145.

Ancelle, T., A. Paugam, F. Bourlioux, A. Merad, and J.P. Vigier. 1997. [Detection of trypanosomes in blood by the Quantitative Buffy Coat (QBC) technique: experimental evaluation]. *Med Trop (Mars)*. 57:245-248.

Apted, F.I.C. 1970. Clinical manifestations and diagnosis of sleeping sickness. *In* The African trypanosomiasis. H.W. Mulligan, editor. George Allen & Unwin, London. 661-683.

Atouguia, J.L.M., and P.G.E. Kennedy. 2000. Neurological aspects of human African trypanosomiasis. *In* Infectious diseases of the nervous system. L.E. Davis and P.G.E. Kennedy, editors. Butterworth-Heinemann, Oxford. 321-372.

Balasegaram, M., S. Balasegaram, D. Malvy, and P. Millet. 2008. Neglected diseases in the news: a content analysis of recent international media coverage focussing on leishmaniasis and trypanosomiasis. *PLoS neglected tropical diseases*. 2:e234.

Barrett, M. 2010. Potential new drugs for human African trypanosomiasis: some progress at last. *Curr Opin Infect Dis*. 23:603-608.

Batchelor, N.A., P.M. Atkinson, P.W. Gething, K. Picozzi, E.M. Fèvre, A.S.L. Kakembo, and S.C. Welburn. 2009. Spatial predictions of Rhodesian Human African Trypanosomiasis (sleeping sickness) prevalence in Kaberamaido and Dokolo, two newly affected districts of Uganda. *PLoS neglected tropical diseases*. 3:e563.

Berberof, M., D. Perez-Morga, and E. Pays. 2001. A receptor-like flagellar pocket glycoprotein specific to Trypanosoma brucei gambiense. *Molecular and biochemical parasitology*. 113:127-138.

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D.C. Bartholomeu, N.J. Lennard, E. Caler, N.E. Hamlin, B. Haas, U. Bohme, L. Hannick, M.A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U.C. Alsmark, C. Arrowsmith, R.J. Atkin, A.J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.J. Chillingworth, C. Churcher, L.N. Clark, C.H. Corton, A. Cronin, R.M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M.C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B.R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A.X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P.J. Mooney, S. Moule, D.M. Martin, G.W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M.A. Quail, E. Rabbinowitsch, M.A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A.J. Simpson, L. Tallon, C.M. Turner, A. Tait, A.R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M.D. Adams, T.M. Embley, K. Gull, E. Ullu, J.D. Barry, A.H. Fairlamb, F. Opperdoes, B.G. Barrell, J.E. Donelson, N. Hall, C.M. Fraser, et al. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science*. 309:416-422.

Biéler, S., E. Matovu, P. Mitashi, E. Ssewannyana, S.K.B. Shamamba, P.R. Bessell, and J.M. Ndung'u. 2012. Improved detection of Trypanosoma brucei by lysis of red blood cells, concentration and LED fluorescence microscopy. *Acta Tropica*. 121:135-140.

Brady, G., M. Barbara, and N.N. Iscove. 1990. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol. Cell. Biol.* 2:17-25.

Brun, R., J. Blum, F. Chappuis, and C. Burri. 2010. Human African trypanosomiasis. *Lancet*. 375:148-159.

Brun, R., R. Don, R.T. Jacobs, M.Z. Wang, and M.P. Barrett. 2011. Development of novel drugs for human African trypanosomiasis. *Future Microbiol*. 6:677-691.

Bucheton, B., A. Macleod, and V. Jamonneau. 2011. Human host determinants influencing the outcome of Trypanosoma brucei gambiense infections. *Parasite Immunol*. 33:438-447.

Burri, C., and R. Brun. 2003. Eflornithine for the treatment of human African trypanosomiasis. *Parasitology research*. 90 Supp 1:S49-52.

Büscher, P., V. Lejon, E. Magnus, and N. Van Meirvenne. 1999. Improved latex agglutination test for detection of antibodies in serum and cerebrospinal fluid of Trypanosoma brucei gambiense infected patients. *Acta Tropica*. 73:11-20.

Capewell, P., N.J. Veitch, C.M.R. Turner, J. Raper, M. Berriman, S.L. Hajduk, and A. Macleod. 2011. Differences between Trypanosoma brucei gambiense Groups 1 and 2 in Their Resistance to Killing by Trypanolytic Factor 1. *PLoS neglected tropical diseases*. 5:e1287.

Carter, N.S., B.J. Berger, and A.H. Fairlamb. 1995. Uptake of diamidine drugs by the P2 nucleoside transporter in melarsen-sensitive and -resistant Trypanosoma brucei brucei. *J Biol Chem*. 270:28153-28157.

Carver, T., M. Berriman, A. Tivey, C. Patel, U. Böhme, B.G. Barrell, J. Parkhill, and M.-A. Rajandream. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. ….

Cattand, P. 2001. The epidemiology of human African trypanosomiasis: a complex multifactorial history. *Med Trop (Mars)*. 61:313-322.

Chappuis, F., L. Loutan, P. Simarro, V. Lejon, and P. Buscher. 2005. Options for field diagnosis of human african trypanosomiasis. *Clin Microbiol Rev*. 18:133-146.

Checchi, F., J.A. Filipe, D.T. Haydon, D. Chandramohan, and F. Chappuis. 2008. Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC infectious diseases*. 8:16.

Chomczynski, P., and N. Sacchi. 2006. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc*. 1:581-585.

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

Connor, R.J. 1994. The impact of nagana. *The Onderstepoort journal of veterinary research*. 61:379-383.

Croft, S.L., M.P. Barrett, and J.A. Urbina. 2005. Chemotherapy of trypanosomiases and leishmaniasis. *Trends in parasitology*. 21:508-512.

Daniels, J.P., K. Gull, and B. Wickstead. 2010. Cell biology of the trypanosome genome. *Microbiology and molecular biology reviews : MMBR*. 74:552-569.

Das, A., Q. Zhang, J.B. Palenchar, B. Chatterjee, G.A. Cross, and V. Bellofatto. 2005. Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Molecular and cellular biology*. 25:7314-7322.

De Greef, C., and R. Hamers. 1994. The serum resistance-associated (SRA) gene of Trypanosoma brucei rhodesiense encodes a variant surface glycoprotein-like protein. *Molecular and biochemical parasitology*. 68:277-284.

Debey, S., U. Schoenbeck, M. Hellmich, B.S. Gathof, R. Pillai, T. Zander, and J.L. Schultze. 2004. Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. *The pharmacogenomics journal*. 4:193-207.

Deborggraeve, S., M. Koffi, V. Jamonneau, F.A. Bonsu, R. Queyson, P.P. Simarro, P. Herdewijn, and P. Buscher. 2008. Molecular analysis of archived blood slides reveals an atypical human Trypanosoma infection. *Diagnostic microbiology and infectious disease*. 61:428-433.

Dempsey, W.L., and J.M. Mansfield. 1983. Lymphocyte function in experimental African trypanosomiasis. V. Role of antibody and the mononuclear phagocyte system in variant-specific immunity. *Journal of immunology*. 130:405-411.

Despommier, D., R. Gwardz, P. Hotez, and C. Knirsch. 2005. Parasitic diseases. Apple trees productions, L.L.C, New York.

Duale, N., G. Brunborg, K.S. Rønningen, T. Briese, J. Aarem, K.K. Aas, P. Magnus, C. Stoltenberg, E. Susser, and W.I. Lipkin. 2012. Human blood RNA stabilization in samples collected and transported for a large biobank. *BMC research notes*. 5:510.

Duggan, A.J., and M.P. Hutchinson. 1966. Sleeping sickness in Europeans: a review of 109 cases. *J Trop Med Hyg*. 69:124-131.

Dukes, P., W.C. Gibson, J.K. Gashumba, K.M. Hudson, T.J. Bromidge, A. Kaukus, T. Asonganyi, and E. Magnus. 1992. Absence of the LiTat 1.3 (CATT antigen) gene in Trypanosoma brucei gambiense stocks from Cameroon. *Acta Tropica*. 51:123-134.

Dumas, M., J.C. Breton, M. Pestre Alexandre, P.L. Girard, and C. Giordano. 1985. Current status of the therapy of human African trypanosomiasis. *Presse Med*. 14:253-256.

Eberwine, J. 1996. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *BioTechniques*:584-594.

Eberwine, J., H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman. 1992. Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences of the United States of America*. 89:3010-3014.

Eikmans, M., N.V. Rekers, J.D.H. Anholts, S. Heidt, and F.H.J. Claas. 2013. Blood cell mRNAs and microRNAs: optimized protocols for extraction and preservation. *Blood*. 121:e81-89.

El-Sayed, N.M., P. Hegde, J. Quackenbush, S.E. Melville, and J.E. Donelson. 2000. The African trypanosome genome. *Int. J. Parasitol.* 30:329-345.

Enyaru, J.C., E. Matovu, M. Akol, C. Sebikali, J. Kyambadde, C. Schmidt, R. Brun, R. Kaminsky, L.M. Ogwal, and F. Kansiime. 1998. Parasitological detection of Trypanosoma brucei gambiense in serologically negative sleeping-sickness suspects from north-western Uganda. *Annals of tropical medicine and parasitology*. 92:845-850.

Ersfeld, K., S.E. Melville, and K. Gull. 1999. Nuclear and genome organization of *Trypanosoma brucei*. *Parasitol. Today*. 15:58-63.

Fadda, A., V. Farber, D. Droll, and C. Clayton. 2013. The roles of 3'-exoribonucleases and the exosome in trypanosome mRNA degradation. *RNA*. 19:937-947.

Fairbairn, H., and D.G. Godfrey. 1957. The local reaction in man at the site of infection with Trypanosoma rhodesiense. *Ann Trop Med Parasitol*. 51:464-470.

Fairlamb, A.H. 2003. Chemotherapy of human African trypanosomiasis: current and future prospects. *Trends in parasitology*. 19:488-494.

Fairlamb, A.H., and I.B. Bowman. 1980. Uptake of the trypanocidal drug suramin by bloodstream forms of Trypanosoma brucei and its effect on respiration and growth rate in vivo. *Mol Biochem Parasitol*. 1:315-333.

Fairlamb, A.H., and A. Cerami. 1992. Metabolism and functions of trypanothione in the Kinetoplastida. *Annu. Rev. Microbiol.* 46:695-729.

Fairlamb, A.H., G.B. Henderson, and A. Cerami. 1989. Trypanothione is the primary target for arsenical drugs against African trypanosomiasis. *Proc. Natl. Acad. Sci. USA*. 86:2607-2611.

Feezor, R.J., H.V. Baker, M. Mindrinos, D. Hayden, C.L. Tannahill, B.H. Brownstein, A. Fay, S. MacMillan, J. Laramie, W. Xiao, L.L. Moldawer, J.P. Cobb, K. Laudanski, C.L. Miller-Graziano, R.V. Maier, D. Schoenfeld, R.W. Davis, R.G. Tompkins, Inflammation, and L.-S.C.R.P. Host Response to Injury. 2004. Whole blood and leukocyte RNA isolation for gene expression analyses. *Physiol Genomics*. 19:247-254.

Fernández-Moya, S.M., and A.M. Estévez. 2010. Posttranscriptional control and the role of RNA-binding proteins in gene regulation in trypanosomatid protozoan parasites. *Wiley interdisciplinary reviews. RNA*. 1:34-46.

Field, L.A., R.M. Jordan, J.A. Hadix, M.A. Dunn, C.D. Shriver, R.E. Ellsworth, and D.L. Ellsworth. 2007. Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. *Clin Biochem*. 40:499-502.

Figueiredo, L.M., G.A.M. Cross, and C.J. Janzen. 2009. Epigenetic regulation in African trypanosomes: a new kid on the block. *Nature Publishing Group*. 7:504-513.

Fleige, S., and M.W. Pfaffl. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine*. 27:126-139.

Fu, X., N. Fu, S. Guo, Z. Yan, Y. Xu, H. Hu, C. Menzel, W. Chen, Y. Li, R. Zeng, and P. Khaitovich. 2009. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC genomics*. 10:161.

Gibson, W., T. Backhouse, and A. Griffiths. 2002. The human serum resistance associated gene is ubiquitous and conserved in Trypanosoma brucei rhodesiense throughout East Africa. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 1:207-214.

Glover, L., and D. Horn. 2006. Repression of polymerase I-mediated gene expression at Trypanosoma brucei telomeres. *EMBO Rep.* 7:93-99.

Goodhead, I., P. Capewell, J.W. Bailey, T. Beament, M. Chance, S. Kay, S. Forrester, A. Macleod, M. Taylor, H. Noyes, and N. Hall. 2013. Whole-Genome Sequencing of Trypanosoma brucei Reveals Introgression between Subspecies That Is Associated with Virulence. *mBio*. 4.

Greenwood, B.M., and H.C. Whittle. 1980. The pathogenesis of sleeping sickness. *Trans R Soc Trop Med Hyg*. 74:716-725.

Gunzl, A., T. Bruderer, G. Laufer, B. Schimanski, L.C. Tu, H.M. Chung, P.T. Lee, and M.G. Lee. 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryot Cell*. 2:542-551.

Haanstra, J.R., M. Stewart, V.D. Luu, A. van Tuijl, H.V. Westerhoff, C. Clayton, and B.M. Bakker. 2008. Control and regulation of gene expression: quantitative analysis of the expression of phosphoglycerate kinase in bloodstream form Trypanosoma brucei. *The Journal of biological chemistry*. 283:2495-2507.

Hansen, K.D., S.E. Brenner, and S. Dudoit. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*. 38:e131.

Hecker, H., B. Betschart, M. Burri, and W. Schlimme. 1995. Functional morphology of trypanosome chromatin. *Parasitology Today*. 11:79.

Henry, M.C., P. Kageruka, J.F. Ruppol, H. Bruneel, and Y. Claes. 1981. Evaluation of field diagnosis of trypanosomiasis caused by Trypanosoma brucei gambiense. *Ann Soc Belg Med Trop*. 61:79-92.

Hertz, C.J., H. Filutowicz, and J.M. Mansfield. 1998. Resistance to the African trypanosomes is IFN-gamma dependent. *J Immunol*. 161:6775-6783.

Hummon, A.B., S.R. Lim, M.J. Difilippantonio, and T. Ried. 2007. Isolation and solubilization of proteins after TRIzol extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques*. 42:467-470, 472.

Hutchinson, O.C., H. Webb, K. Picozzi, S. Welburn, and M. Carrington. 2004. Candidate protein selection for diagnostic markers of African trypanosomiasis. *Trends Parasitol.* 20:519-523.

Iscove, N.N., M. Barbara, M. Gu, M. Gibson, C. Modi, and N. Winegarden. 2002. Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat Biotechnol.* 20:940-943.

Jackson, A.P., M. Sanders, A. Berry, J. McQuillan, M.A. Aslett, M.A. Quail, B. Chukualim, P. Capewell, A. Macleod, S.E. Melville, W. Gibson, J.D. Barry, M. Berriman, and C. Hertz-Fowler. 2010. The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis. *PLoS neglected tropical diseases*. 4:e658.

Jacobs, R.T., B. Nare, S.A. Wring, M.D. Orr, D. Chen, J.M. Sligar, M.X. Jenks, R.A. Noe, T.S. Bowling, L.T. Mercer, C. Rewerts, E. Gaukel, J. Owens, R. Parham, R. Randolph, B. Beaudet, C.J. Bacchi, N. Yarlett, J.J. Plattner, Y. Freund, C. Ding, T. Akama, Y.K. Zhang, R. Brun, M. Kaiser, I. Scandale, and R. Don. 2011a. SCYX-7158, an Orally-Active Benzoxaborole for the Treatment of Stage 2 Human African Trypanosomiasis. *PLoS Negl Trop Dis*. 5:e1151.

Jacobs, R.T., J.J. Plattner, B. Nare, S.A. Wring, D. Chen, Y. Freund, E.G. Gaukel, M.D. Orr, J.B. Perales, M. Jenks, R.A. Noe, J.M. Sligar, Y.K. Zhang, C.J. Bacchi, N. Yarlett, and R. Don. 2011b. Benzoxaboroles: a new class of potential drugs for human African trypanosomiasis. *Future Med Chem*. 3:1259-1278.

Jamonneau, V., H. Ilboudo, J. Kabore, D. Kaba, M. Koffi, P. Solano, A. Garcia, D. Courtin, C. Laveissiere, K. Lingue, P. Buscher, and B. Bucheton. 2012. Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal. *PLoS Negl Trop Dis*. 6:e1691.

Jensen, B.C., D. Sivam, C.T. Kifer, P.J. Myler, and M. Parsons. 2009. Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. *BMC Genomics*. 10:482.

Joshi, P.P., V.R. Shegokar, R.M. Powar, S. Herder, R. Katti, H.R. Salkar, V.S. Dani, A. Bhargava, J. Jannin, and P. Truc. 2005. Human trypanosomiasis caused by Trypanosoma evansi in India: the first case report. *The American journal of tropical medicine and hygiene*. 73:491-495.

Kabani, S., K. Fenn, A. Ross, A. Ivens, T.K. Smith, P. Ghazal, and K. Matthews. 2009. Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and

dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. *BMC genomics*. 10:427.

Kaiser, M., M.A. Bray, M. Cal, B. Bourdin Trunz, E. Torreele, and R. Brun. 2011. Antitrypanosomal Activity of Fexinidazole, a New Oral Nitroimidazole Drug Candidate for Treatment of Sleeping Sickness. *Antimicrobial Agents and Chemotherapy*. 55:5602-5608.

Kalisky, T., P. Blainey, and S.R. Quake. 2011. Genomic analysis at the single-cell level. *Annu Rev Genet.* 45:431-445.

Kang, J.-E., S.-H. Hwang, J.H. Lee, D.Y. Park, and H.-H. Kim. 2011. Effects of RBC removal and TRIzol of peripheral blood samples on RNA stability. *Clinica chimica acta; international journal of clinical chemistry*. 412:1883-1885.

Kawasaki, E.S. 2004. Microarrays and the gene expression profile of a single cell. *Ann N Y Acad Sci*. 1020:92-100.

Kazyumba, G., M. Berney, G. Brighouse, A. Cruchaud, and P.H. Lambert. 1986. Expression of the B cell repertoire and autoantibodies in human African trypanosomiasis. *Clinical and experimental immunology*. 65:10-18.

Kennedy, P.G. 2008. The continuing problem of human African trypanosomiasis (sleeping sickness). *Ann Neurol*. 64:116-126.

Kennedy, P.G. 2012. An alternative form of melarsoprol in sleeping sickness. *Trends in parasitology*. 28:307-310.

Kennedy, P.G.E. 2004. Human African trypanosomiasis of the CNS: current issues and challenges. *The Journal of clinical investigation*. 113:496-504.

Kieft, R., N.A. Stephens, P. Capewell, A. Macleod, and S.L. Hajduk. 2012. Role of expression site switching in the development of resistance to human Trypanosome Lytic Factor-1 in Trypanosoma brucei brucei. *Molecular and biochemical parasitology*. 183:8-14.

Klur, S., K. Toy, M.P. Williams, and U. Certa. 2004. Evaluation of procedures for amplification of small-size samples for hybridization on microarrays. *Genomics*. 83:508-517.

Kolev, N.G., J.B. Franklin, S. Carmi, H. Shi, S. Michaeli, and C. Tschudi. 2010. The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS pathogens*. 6:e1001090.

Krafsur, E.S. 2009. Tsetse flies: genetics, evolution, and role as vectors. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 9:124-141.

Kuboki, N., N. Inoue, T. Sakurai, F. Di Cello, D.J. Grab, H. Suzuki, C. Sugimoto, and I. Igarashi. 2003. Loop-mediated isothermal amplification for detection of African trypanosomes. *Journal of clinical microbiology*. 41:5517-5524.

Kuepfer, I., E.P. Hhary, M. Allan, A. Edielu, C. Burri, and J.A. Blum. 2011. Clinical presentation of T.b. rhodesiense sleeping sickness in second stage patients from Tanzania and Uganda. *PLoS neglected tropical diseases*. 5:e968.

Kuepfer, I., C. Schmid, M. Allan, A. Edielu, E.P. Haary, A. Kakembo, S. Kibona, J. Blum, and C. Burri. 2012. Safety and efficacy of the 10-day melarsoprol schedule for the treatment of second stage Rhodesiense sleeping sickness. *PLoS neglected tropical diseases*. 6:e1695.

Kurimoto, K., Y. Yabuta, Y. Ohinata, Y. Ono, K.D. Uno, R.G. Yamada, H.R. Ueda, and M. Saitou. 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic acids research*. 34:e42.

Kurimoto, K., Y. Yabuta, Y. Ohinata, and M. Saitou. 2007. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nature protocols*. 2:739-752.

Kuzoe, F.A. 1993. Current situation of African trypanosomiasis. *Acta Tropica*. 54:153-162.

Lambert, P.H., M. Berney, and G. Kazyumba. 1981. Immune complexes in serum and in cerebrospinal fluid in African trypanosomiasis. Correlation with polyclonal B cell activation and with intracerebral immunoglobulin synthesis. *The Journal of clinical investigation*. 67:77-85.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 10:R25.

Lanham, S.M., and D.G. Godfrey. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. *Exp Parasitol*. 28:195-202.

Le-Niculescu, H., S.M. Kurian, N. Yehyawi, C. Dike, S.D. Patel, H.J. Edenberg, M.T. Tsuang, D.R. Salomon, J.I. Nurnberger, Jr., and A.B. Niculescu. 2009. Identifying blood

biomarkers for mood disorders using convergent functional genomics. *Molecular psychiatry*. 14:156-174.

Legros, D., S. Evans, F. Maiso, J. Enyaru, and D. Mbulamberi. 1999. Risk factors for treatment failure after melarsoprol for *Trypanosoma brucei gambiense* trypanosomiasis in Uganda. *Trans R Soc Trop Med Hyg*. 93:439-442.

Lejon, V., P. Buscher, E. Magnus, A. Moons, I. Wouters, and N. Van Meirvenne. 1998. A semi-quantitative ELISA for detection of Trypanosoma brucei gambiense specific antibodies in serum and cerebrospinal fluid of sleeping sickness patients. *Acta Tropica*. 69:151-164.

Lejon, V., J. Lardon, G. Kenis, L. Pinoges, D. Legros, S. Bisser, X. N'Siesi, E. Bosmans, and P. Büscher. 2002a. Interleukin (IL)-6, IL-8 and IL-10 in serum and CSF of *Trypanosoma brucei gambiense* sleeping sickness patients before and after treatment. *Trans R Soc Trop Med Hyg*. 96:329-333.

Lejon, V., D. Legros, M. Richer, J.A. Ruiz, V. Jamonneau, P. Truc, F. Doua, N. Dje, F.X. N'Siesi, S. Bisser, E. Magnus, I. Wouters, J. Konings, T. Vervoort, F. Sultan, and P. Buscher. 2002b. IgM quantification in the cerebrospinal fluid of sleeping sickness patients by a latex card agglutination test. *Tropical medicine & international health : TM & IH*. 7:685-692.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760.

Li, J., H. Jiang, and W.H. Wong. 2010. Modeling non-uniformity in short-read rates in RNA-Seq data. *Genome biology*. 11:R50.

Liang, X., A. Haritan, S. Uliel, and S. Michaeli. 2003. *Trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Euk. Cell*. 2:830-840.

Liu, Y., S.A. Motyka, and P.T. Englund. 2005. Effects of RNA interference of Trypanosoma brucei structure-specific endonuclease-I on kinetoplast DNA replication. *The Journal of biological chemistry*. 280:35513-35520.

Livesey, F.J. 2003. Strategies for microarray analysis of limiting amounts of RNA. *Brief Funct Genomic Proteomic*. 2:31-36.

Lonsdale-Eccles, J.D., and D.J. Grab. 1987. Purification of African trypanosomes can cause biochemical changes in the parasites. *The Journal of protozoology*. 34:405-408.

Lucas, R., S. Magez, E. Bajyana Songa, A. Darji, R. Hamers, and P. De Baetselier. 1993. A role for TNF-α during African trypanosomiasis: involvement in parasite control, immunosuppression and pathology. *Res Immunol*. 144:370-376.

Lumsden, W., C. Kimber, D. Evans, and S. Doig. 1979. *Trypanosoma brucei*: miniature anion-exchange centrifugation technique for detection of low parasitaemias: adaptation to field use. *Trans R Soc Trop Med Hyg*. 73:312-317.

Luo, H., G. Gilinger, D. Mukherjee, and V. Bellofatto. 1999. Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. *The Journal of biological chemistry*. 274:31947-31954.

Ma, W., M. Wang, Z.-Q. Wang, L. Sun, D. Graber, J. Matthews, R. Champlin, Q. Yi, R.Z. Orlowski, L.W. Kwak, D.M. Weber, S.K. Thomas, J. Shah, S. Kornblau, and R.E. Davis. 2010. Effect of long-term storage in TRIzol on microarray-based gene expression profiling. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 19:2445-2452.

Macaskill, J.A., P.H. Holmes, F.W. Jennings, and G.M. Urquhart. 1981. Immunological clearance of 75Se-labelled Trypanosoma brucei in mice. III. Studies in animals with acute infections. *Immunology*. 43:691-698.

MacLean, L., M. Odiit, D. Okitoi, and J.M. Sternberg. 1999. Plasma nitrate and interferon-gamma in Trypanosoma brucei rhodesiense infections: evidence that nitric oxide production is induced during both early blood-stage and late meningoencephalitic-stage infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 93:169-170.

MacLean, L., M. Odiit, and J. Sternberg. 2001a. Nitric oxide and cytokine synthesis in human African trypanosomiasis. *J Infect Dis.* 184:1086-1090.

MacLean, L., M. Odiit, and J.M. Sternberg. 2001b. Nitric oxide and cytokine synthesis in human African trypanosomiasis. *The Journal of infectious diseases*. 184:1086-1090.

MacLean, L.M., M. Odiit, J.E. Chisi, P.G.E. Kennedy, and J.M. Sternberg. 2010. Focus-specific clinical profiles in human African Trypanosomiasis caused by Trypanosoma brucei rhodesiense. *PLoS neglected tropical diseases*. 4:e906.

Magez, S., G. Caljon, T. TRAN, B. Stijlemans, and M. Radwanska. 2010. Current status of vaccination against African trypanosomiasis. *Parasitology*. 137:2017-2027.

Magez, S., M. Geuskens, A. Beschin, H. del Favero, H. Verschueren, R. Lucas, E. Pays, and P. de Baetselier. 1997. Specific uptake of tumor necrosis factor-α is involved in growth control of *Trypanosoma brucei*. *J. Cell. Biol.* 137:715-727.

Magez, S., B. Stijlemans, T. Baral, and P. De Baetselier. 2002. VSG-GPI anchors of African trypanosomes: their role in macrophage activation and induction of infection-associated immunopathology. *Microbes and Infection*. 4:999-1006.

Magnus, E., N. Van Meirvenne, T. Vervoort, D. Le Ray, and M. Wery. 1978a. Use of freeze-dried trypanosomes in the indirect fluorescent antibody test for the serodiagnosis of sleeping sickness. *Ann Soc Belg Med Trop*. 58:103-109.

Magnus, E., T. Vervoort, and N. Van Meirvenne. 1978b. A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of T. B. gambiense trypanosomiasis. *Ann Soc Belg Med Trop*. 58:169-176.

Malboeuf, C.M., S.J. Isaacs, N.H. Tran, and B. Kim. 2001. Thermal effects on reverse transcription: improvement of accuracy and processivity in cDNA synthesis. *BioTechniques*. 30:1074-1078, 1080, 1082, passim.

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

Manful, T., J. Mulindwa, F.M. Frank, C.E. Clayton, and E. Matovu. 2010. A search for Trypanosoma brucei rhodesiense diagnostic antigens by proteomic screening and targeted cloning. *PloS one*. 5:e9630.

Mansfield, J.M. 1994. T-cell responses to the trypanosome variant surface glycoprotein: a new paradigm? *Parasitology Today*. 10:267-270.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 9:387-402.

Marioni, J.C., C.E. Mason, S.M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*. 18:1509-1517.

Marteau, J.-B., S. Mohr, M. Pfister, and S. Visvikis-Siest. 2005. Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. *Clinical chemistry*. 51:1250-1252.

Masocha, W., B. Robertson, M.E. Rottenberg, J. Mhlanga, L. Sorokin, and K. Kristensson. 2004. Cerebral vessel laminins and IFN-gamma define Trypanosoma brucei brucei penetration of the blood-brain barrier. *J Clin Invest*. 114:689-694.

Mastrokolias, A., J.T. den Dunnen, G.B. van Ommen, P.A.C. 't Hoen, and W.M.C. van Roon-Mom. 2012. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC genomics*. 13:28.

Metzker, M.L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet.* 11:31-46.

Mikulowska-Mennis, A., T.B. Taylor, P. Vishnu, S.A. Michie, R. Raja, N. Horner, and S.T. Kunitake. 2002. High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques*. 33:176-179.

Mitsuhashi, M., S. Tomozawa, K. Endo, and A. Shinagawa. 2006. Quantification of mRNA in whole blood by assessing recovery of RNA and efficiency of cDNA synthesis. *Clinical chemistry*. 52:634-642.

Mulenga, C., J.D. Mhlanga, K. Kristensson, and B. Robertson. 2001. Trypanosoma brucei brucei crosses the blood-brain barrier while tight junction proteins are preserved in a rat chronic disease model. *Neuropathology and applied neurobiology*. 27:77-85.

Mumba Ngoyi, D., J. Menten, P.P. Pyana, P. Buscher, and V. Lejon. 2013. Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. *Tropical medicine & international health : TM & IH*. 18:778-782.

Naessens, J., D.M. Mwangi, J. Buza, and S.K. Moloo. 2003. Local skin reaction (chancre) induced following inoculation of metacyclic trypanosomes in cattle by tsetse flies is dependent on CD4 T lymphocytes. *Parasite Immunol*. 25:413-419.

Namangala, B., P. de Baetselier, L. Brijs, B. Stijlemans, W. Noel, E. Pays, M. Carrington, and A. Beschin. 2000. Attenuation of *Trypanosoma brucei* is associated with reduced immunosuppression and concomitant production of Th2 lymphokines. *J Infect Dis.* 181:1110-1120.

Namangala, B., W. Noël, P. De Baetselier, L. Brys, and A. Beschin. 2001. Relative contribution of interferon- γ and interleukin-10 to resistance to murine African trypanosomosis. *The Journal of infectious diseases*. 183:1794-1800.

Nilsson, D., K. Gunasekera, J. Mani, M. Osteras, L. Farinelli, L. Baerlocher, I. Roditi, and T. Ochsenreiter. 2010. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS pathogens*. 6:e1001037.

Njiru, Z., K. Ndung'u, G. Matete, J. Ndungu, and W. Gibson. 2004. Detection of *Trypanosoma brucei rhodesiense* in animals from sleeping sickness foci in East Africa using the serum resistance associated (SRA) gene. *Acta Trop.* 90:249-254.

Njiru, Z.K., A.S.J. Mikosza, T. Armstrong, J.C. Enyaru, J.M. Ndung'u, and A.R.C. Thompson. 2008. Loop-mediated isothermal amplification (LAMP) method for rapid detection of Trypanosoma brucei rhodesiense. *PLoS neglected tropical diseases*. 2:e147.

Njogu, R.M., and J.K. Kiaira. 1982. Trypanosoma brucei: a quick method for separating blood-stream trypomastigotes from infected blood by differential osmotic lysis. *Comp Biochem Physiol A Comp Physiol*. 71:265-269.

Noireau, F., J.P. Gouteux, and J.P. Duteurtre. 1987. Diagnostic value of a card agglutination test (Testryp CATT) in the mass screening of human trypanosomiasis in the Congo. *Bull Soc Pathol Exot Filiales*. 80:797-803.

Odiit, M., F. Kansiime, and J.C. Enyaru. 1997. Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. *East Afr Med J*. 74:792-795.

Okomo-Assoumou, M.C., S. Daulouede, J.L. Lemesre, A. N'Zila-Mouanda, and P. Vincendeau. 1995. Correlation of high serum levels of tumor necrosis factor-alpha with disease severity in human African trypanosomiasis. *The American journal of tropical medicine and hygiene*. 53:539-543.

Okoniewski, M.J., and C.J. Miller. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*. 2:276.

Ouellette, M., and B. Papadopoulou. 2009. Coordinated gene expression by post-transcriptional regulons in African trypanosomes. *Journal of biology*. 8:100.

Ovstebo, R., K. Lande, P. Kierulf, and K.B. Haug. 2007. Quantification of relative changes in specific mRNAs from frozen whole blood - methodological considerations and clinical implications. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 45:171-176.

Ozsolak, F., and P.M. Milos. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Publishing Group*. 12:87-98.

Palenchar, J.B., and V. Bellofatto. 2006. Gene transcription in trypanosomes. *Molecular and biochemical parasitology*. 146:135-141.

Pan, X., R.E. Durrett, H. Zhu, Y. Tanaka, Y. Li, X. Zi, S.L. Marjani, G. Euskirchen, C. Ma, R.H. Lamotte, I.-H. Park, M.P. Snyder, C.E. Mason, and S.M. Weissman. 2013. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proceedings of the National Academy of Sciences of the United States of America*. 110:594-599.

Papadopoulos, M.C., P.M. Abel, D. Agranoff, A. Stich, E. Tarelli, B.A. Bell, T. Planche, A. Loosemore, S. Saadoun, P. Wilkins, and S. Krishna. 2004. A novel and accurate diagnostic test for human African trypanosomiasis. *Lancet*. 363:1358-1363.

Pays, E., L. Vanhamme, and D. Perez-Morga. 2004. Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. *Current opinion in microbiology*. 7:369-374.

Pays, E., B. Vanhollebeke, L. Vanhamme, F. Paturiaux-Hanocq, D. Nolan, and D. Perez-Morga. 2006. The trypanolytic factor of human serum. *Nat Rev Microbiol*. 4:477-486.

Penchenier, L., P. Grebaut, F. Njokou, V. Eboo Eyenga, and P. Buscher. 2003. Evaluation of LATEX/T.b.gambiense for mass screening of Trypanosoma brucei gambiense sleeping sickness in Central Africa. *Acta Tropica*. 85:31-37.

Penchenier, L., J. Jannin, J.P. Moulia-Pelat, F. Elfassi de la Baume, G. Fadat, B. Chanfreau, and P. Eozenou. 1991. Interpretation of the CATT (Card Agglutination

Trypanosomiasis Test) in the screening for human trypanosomiasis due to Trypanosoma brucei gambiense. *Ann Soc Belg Med Trop*. 71:221-228.

Pepin, J., and F. Milord. 1994. The treatment of human African trypanosomiasis. *Advances in Parasitology*. 33:1-47.

Perez-Novo, C.A., C. Claeys, F. Speleman, P. Van Cauwenberge, C. Bachert, and J. Vandesompele. 2005. Impact of RNA quality on reference gene expression stability. *Biotechniques*. 39:52, 54, 56.

Picozzi, K., E. Fevre, M. Odiit, M. Carrington, M. Eisler, I. Maudlin, and S. Welburn. 2005. Sleeping sickness in Uganda: a thin line between two fatal diseases. *BMJ*. 331:1238-1241.

Priotto, G., S. Kasparian, W. Mutombo, D. Ngouama, S. Ghorashian, U. Arnold, S. Ghabri, E. Baudin, V. Buard, S. Kazadi-Kyanza, M. Ilunga, W. Mutangala, G. Pohlig, C. Schmid, U. Karunakara, E. Torreele, and V. Kande. 2009. Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *Lancet*. 374:56-64.

Queiroz, R., C. Benz, K. Fellenberg, J.D. Hoheisel, and C. Clayton. 2009. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC genomics*. 10:495.

Radwanska, M., M. Chamekh, L. Vanhamme, F. Claes, S. Magez, E. Magnus, P. de Baetselier, P. Buscher, and E. Pays. 2002. The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. *Am J Trop Med Hyg.* 67:684-690.

Rainen, L., U. Oelmueller, S. Jurgensen, R. Wyrich, C. Ballas, J. Schram, C. Herdman, D. Bankaitis-Davis, N. Nicholls, D. Trollinger, and V. Tryon. 2002. Stabilization of mRNA expression in whole blood samples. *Clinical chemistry*. 48:1883-1890.

Raper, J., R. Fung, J. Ghiso, V. Nussenzweig, and S. Tomlinson. 1999. Characterization of a novel trypanosome lytic factor from human serum. *Infection and immunity*. 67:1910-1916.

Rapley, R. 1994. Enhancing PCR amplification and sequencing using DNA-binding proteins. *Mol Biotechnol*. 2:295-298.

Reinitz, D.M., and J.M. Mansfield. 1990. T-cell-independent and T-cell-dependent B-cell responses to exposed variant surface glycoprotein epitopes in trypanosome-infected mice. *Infection and immunity*. 58:2337-2342.

Renger, H.C., and D.R. Wolstenholme. 1971. Kinetoplast and other satellite DNAs of kinetoplastic and dyskinetoplastic strains of Trypanosoma. *The Journal of cell biology*. 50:533-540.

Rhind, S.G., B.H. Sabiston, P.N. Shek, A. Buguet, G. Muanga, A. Stanghellini, M. Dumas, and M.W. Radomski. 1997. Effect of melarsoprol treatment on circulating IL-10 and TNF-alpha levels in human African trypanosomiasis. *Clinical immunology and immunopathology*. 83:185-189.

Rifkin, M.R. 1978. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proceedings of the National Academy of Sciences of the United States of America*. 75:3450-3454.

Robays, J., M.M. Bilengue, P. Van der Stuyft, and M. Boelaert. 2004. The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo. *Tropical medicine & international health : TM & IH*. 9:542-550.

Roberts, A., C. Trapnell, J. Donaghey, J.L. Rinn, and L. Pachter. 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology*. 12:R22.

Rodgers, J. 2009. Human African trypanosomiasis, chemotherapy and CNS disease. *J Neuroimmunol*. 211:16-22.

Rodgers, J., A. Jones, S. Gibaud, B. Bradley, C. McCabe, M.P. Barrett, G. Gettinby, and P.G. Kennedy. 2011. Melarsoprol cyclodextrin inclusion complexes as promising oral candidates for the treatment of human African trypanosomiasis. *PLoS Negl Trop Dis*. 5:e1308.

Roozemond, R.C. 1976. Ultramicrochemical determination of nucleic acids in individual cells using the Zeiss UMSP-I microspectrophotometer. Application to isolated rat hepatocytes of different ploidy classes. *Histochem J.* 8:625-638.

Royce, T.E., J.S. Rozowsky, and M.B. Gerstein. 2007. Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Res.* 35:99.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics*. 16:944-945.

Schleifer, K.W., H. Filutowicz, L.R. Schopf, and J.M. Mansfield. 1993. Characterization of T helper cell responses to the Trypanosoma variant surface glycoprotein. *J Immunol*. 150:2910-2919.

Shapiro, T.A., and P.T. Englund. 1995. The structure and replication of kinetoplast DNA. *Annual review of microbiology*. 49:117-143.

Siegel, T.N., K. Gunasekera, G.A.M. Cross, and T. Ochsenreiter. 2011. Gene expression in Trypanosoma brucei: lessons from high-throughput RNA sequencing. *Trends in parasitology*. 27:434-441.

Siegel, T.N., D.R. Hekstra, L.E. Kemp, L.M. Figueiredo, J.E. Lowell, D. Fenyo, X. Wang, S. Dewell, and G.A. Cross. 2009. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev*. 23:1063-1076.

Siegel, T.N., D.R. Hekstra, X. Wang, S. Dewell, and G.A.M. Cross. 2010. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic acids research*. 38:4946-4957.

Simarro, P., F.J. Louis, and J. Jannin. 2003. Sleeping sickness, forgotten illness: what are the consequences in the field? *Med Trop (Mars)*. 63:231-235.

Simarro, P.P., G. Cecchi, J.R. Franco, M. Paone, A. Diarra, J.A. Ruiz-Postigo, E.M. Fèvre, R.C. Mattioli, and J.G. Jannin. 2012a. Estimating and mapping the population at risk of sleeping sickness. *PLoS neglected tropical diseases*. 6:e1859.

Simarro, P.P., G. Cecchi, M. Paone, and J.R. Franco. 2010. The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *Int J Health …*.

Simarro, P.P., J. Franco, A. Diarra, J.A.R. Postigo, and J. Jannin. 2012b. Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis. *Parasitology*. 139:842-846.

Simarro, P.P., J. Jannin, and P. Cattand. 2008. Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS medicine*. 5:e55.

Simpson, L. 1973. Structure and function of kinetoplast DNA. *The Journal of protozoology*. 20:2-8.

Srivastava, S., and L. Chen. 2010. A two-parameter generalized Poisson model to improve the analysis of RNA-seq data. *Nucleic acids research*. 38:e170.

Sternberg, J.M. 2004. Human African trypanosomiasis: clinical presentation and immune response. *Parasite immunology*. 26:469-476.

Stevens, J.R., and S. Brisse. 2004. Systematics of Trypanosomes of medical and Veterinary Importance. *In* The trypanosomiasis. I. Maudlin, P.H. Holmes, and M.A. Miles, editors. CABI Wallingford. 1-23.

Stirewalt, D.L., E.L. Pogosova-Agadjanyan, N. Khalid, D.R. Hare, P.A. Ladne, O. Sala-Torra, L.P. Zhao, and J.P. Radich. 2004. Single-stranded linear amplification protocol results in reproducible and reliable microarray data from nanogram amounts of starting RNA. *Genomics*. 83:321-331.

Tachado, D.S., and L. Schofield. 1994. Glycosylphosphatidylinositol toxin of Trypanosoma brucei regulates IL-1alpha and TNF-alpha expression in macrophages by protein tyrosine kinase mediated signal transduction. *Biochem Biophys Res Commun*. 205:984-991.

Tang, F., C. Barbacioru, E. Nordman, B. Li, N. Xu, V.I. Bashkirov, K. Lao, and M.A. Surani. 2010. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc*. 5:516-535.

Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B.B. Tuch, A. Siddiqui, K. Lao, and M.A. Surani. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods*. 6:377-382.

Tang, F., K. Lao, and M.A. Surani. 2011. Development and applications of single-cell transcriptome analysis. *Nature methods*. 8:S6-11.

Tariq, M.A., H.J. Kim, O. Jejelowo, and N. Pourmand. 2011. Whole-transcriptome RNAseq analysis from minute amount of total RNA. *Nucleic acids research*. 39:e120-e120.

Taylor, J.E., and G. Rudenko. 2006. Switching trypanosome coats: what's in the wardrobe? *Trends Genet*. 22:614-620.

Taylor, K., B. Mertens, V. Lutje, and R. Saya. 1998. Trypanosoma congolense infection of trypanotolerant N'Dama (Bos taurus) cattle is associated with decreased secretion of nitric oxide by interferon-gamma-activated monocytes and increased transcription of interleukin-10. *Parasite Immunol.* 20:421-429.

Thach, D.C., B. Lin, E. Walter, R. Kruzelock, R.K. Rowley, C. Tibbetts, and D.A. Stenger. 2003. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. *J Immunol Methods*. 283:269-279.

Torreele, E., B. Bourdin Trunz, D. Tweats, M. Kaiser, R. Brun, G. Mazue, M.A. Bray, and B. Pecoul. 2010. Fexinidazole--a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness. *PLoS Negl Trop Dis*. 4:e923.

Truc, P., V. Jamonneau, P. N'Guessan, L. N'Dri, P.B. Diallo, and G. Cuny. 1998. Trypanosoma brucei ssp. and T congolense: mixed human infection in Cote d'Ivoire. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 92:537-538.

Truc, P., V. Lejon, E. Magnus, V. Jamonneau, A. Nangouma, D. Verloo, L. Penchenier, and P. Büscher. 2002. Evaluation of the micro-CATT, CATT/Trypanosoma brucei gambiense, and LATEX/T. b. gambiense methods for serodiagnosis and surveillance of human African trypanosomiasis in West and Central Africa. *Bulletin of the World Health Organization*. 80:882-886.

UBOS. 2010. Statistical abstract. *Uganda bureau of Statistics*.

Uemura, E. 1980. Age-related changes in neuronal RNA content in rhesus monkeys (Macaca mulatta). *Brain Res Bull*. 5:117-119.

Uzonna, J.E., R.S. Kaushik, J.R. Gordon, and H. Tabel. 1998. Immunoregulation in experimental murine Trypanosoma congolense infection: anti-IL-10 antibodies reverse trypanosome-mediated suppression of lymphocyte proliferation in vitro and moderately prolong the lifespan of genetically susceptible BALB/c mice. *Parasite Immunol*. 20:293-302.

Uzonna, J.E., R.S. Kaushik, J.R. Gordon, and H. Tabel. 1999. Cytokines and antibody responses during Trypanosoma congolense infections in two inbred mouse strains that differ in resistance. *Parasite Immunol*. 21:57-71.

Uzureau, P., S. Uzureau, L. Lecordier, F. Fontaine, P. Tebabi, F. Homble, A. Grelard, V. Zhendre, D.P. Nolan, L. Lins, J.M. Crowet, A. Pays, C. Felu, P. Poelvoorde, B. Vanhollebeke, S.K. Moestrup, J. Lyngso, J.S. Pedersen, J.C. Mottram, E.J. Dufourc, D. Perez-Morga, and E. Pays. 2013. Mechanism of Trypanosoma brucei gambiense resistance to human serum. *Nature*:430-434.

Van Gelder, R.N., M.E. von Zastrow, A. Yool, W.C. Dement, J.D. Barchas, and J.H. Eberwine. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A*. 87:1663-1667.

Van Meirvenne, N. 1999. Biological diagnosis of human African trypanosomiasis. *In* Progress in Human African Trypanosomiasis, Sleeping Sickness. M. Dumas, B. Bouteille, and A. Buguet, editors. Springer–Verlag, Berlin. 273-292.

Vanhamme, L., and E. Pays. 2004. The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. *Int J Parasitol.* 34:887-898.

Vanhamme, L., E. Pays, R. McCulloch, and J.D. Barry. 2001. An update on antigenic variation in African trypanosomes. *Trends Parasitol.* 17:338-343.

Vanhollebeke, B., and E. Pays. 2010. The trypanolytic factor of human serum: many ways to enter
the parasite, a single way to kill. *Mol Microbiol*. 76:806–814.

Veitch, N.J., P.C. Johnson, U. Trivedi, S. Terry, D. Wildridge, and A. Macleod. 2010. Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. *BMC genomics*. 11:124.

Villalva, C., C. Touriol, P. Seurat, P. Trempat, G. Delsol, and P. Brousset. 2001. Increased yield of PCR products by addition of T4 gene 32 protein to the SMART PCR cDNA synthesis system. *BioTechniques*. 31:81-83, 86.

Vincendeau, P., M.O. Jauberteau-Marchan, S. Dauloue`de, and Z. Ayed. 1999. Immunology of African trypanosomiasis. *In* Progress in human African trypanosomiasis. M. Dumas, B. Bouteille, and A. Buguet, editors. Springer-Verlag, Berlin. 137-156.

Wang, E. 2005. RNA amplification for successful gene profiling analysis. *Journal of translational medicine*. 3:28.

Wang, Y., D.E. Prosen, L. Mei, J.C. Sullivan, M. Finney, and P.B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic acids research*. 32:1197-1207.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Publishing Group*. 10:57-63.

Welburn, S.C., and M. Odiit. 2002. Recent developments in human African trypanosomiasis. *Current opinion in infectious diseases*. 15:477-484.

WHO. 1998. Control and surveillance of African trypanosomiasis. Report of a WHO Expert Committee. *World Health Organization technical report series*. 881:I-VI, 1-114.

WHO. 2012. Research priorities for Chagas disease, human African trypanosomiasis and leishmaniasis. *World Health Organization technical report series*:v-xii, 1-100.

Woo, P.T. 1970. The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis. *Acta Tropica*. 27:384-386.

Xiang, C.C., M. Chen, L. Ma, Q.N. Phan, J.M. Inman, O.A. Kozhich, and M.J. Brownstein. 2003. A new strategy to amplify degraded RNA from small tissue samples for microarray studies. *Nucleic acids research*. 31:e53.

Xong, H.V., L. Vanhamme, M. Chamekh, C.E. Chimfwembe, J. Van Den Abbeele, A. Pays, N. Van Meirvenne, R. Hamers, P. De Baetselier, and E. Pays. 1998. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell*. 95:839-846.

List of Figures

List of Tables