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

Smiths Sengkwawoh Lueong

born in Mbouda, Cameroon

Oral Examination: October 16th 2014

***Trypanosoma brucei*: Protein Expression Microarrays and Circulating miRNA during Infection.**

Referees: Prof. Dr. Christine E. Clayton

Prof. Dr. Holger Sültmann

Declaration of Authenticity

I, the undersigned, Smiths Sengkwawoh Lueong, hereby declare that the work described in this thesis is authentic, was done and written by the undersigned without assistance of any kind from third parties. I further confirm that no sources other than those listed in the thesis itself have been used for its preparation. Where other people's work is described, proper accreditation is given.

The work is not being, and has not been presented elsewhere for the award of any certificate whatsoever

Heidelberg, the 10th August 2014

Smiths S. Lueong

Related Publications

Smiths S. Lueong, Syafrizayanti, and Jörg D. Hoheisel. (2014). Production of protein expression arrays from cDNA products for functional applications *Journal of Proteomics (In prep).*

Erben ED, Fadda A, Lueong S, Hoheisel JD, Clayton C (2014) A Genome-Wide Tethering Screen Reveals Novel Potential Post-Transcriptional Regulators in *Trypanosoma brucei*. *PLoS Pathog* 10(6): e1004178.

Syafrizayanti , Smiths S. Lueong, and Jörg D. Hoheisel. (2014). Personalized protein *in situ* microarrays. Nature Methods (*Submitted*).

Lueong, S., Simo, G., Jamonneau, V., Bucheton, B., Hoheisel, J.D. & Clayton, C. (2013). The miRNA and mRNA signatures of peripheral blood cells in humans infected with Trypanosoma brucei gambiense. *PLoS ONE* 8, e67312.

Lueong SS, Hoheisel JD, Saeed Alhamdani MS (2013). Protein Microarrays as Tools for Functional Proteomics: Achievements, Promises and Challenges. *J Proteomics Bioinform* S7: 004

DEDICATION

This work is dedicated to:

God the Father Almighty, for giving me the breath of life and making it possible for me to see this day

My mother, Margaret Pupongfeh Lueong, My wife Carelle Matene Fotsing and son Benaiah-Bright Kiembouo whose love, inspiration, and concern have lifted me up socially, and to whom I shall remain grateful for giving me the true meaning of life's endless struggle.

ACKNOWLEDGEMENTS

This work could not have been achieved without the help of a number of persons. This is the occasion for me to express my sincere gratitude to them, especially to:

Dr Jörg D. Hoheisel, who in the first instance, gave me the opportunity to pursue my Ph.D in his division and whose relentless efforts, corrections and advices enabled me to successfully complete this thesis. I am really indeed indebted to you.

My Advisors Prof. Dr. Christine E. Clayton and Prof. Dr. Holger Sültmann whose timely interventions, advices and counselling avoided any distraction from my focus. I must admit, that without your distinguished contributions and support, this work would not have been realized. Your efforts trained me as a person and most importantly gave this work the shape that it has taken. For all these, I say thank you.

Prof. Dr. Kuiate Jules Roger, Dr. Gustave Simo, and Dr. Tume B. Christopher who, from the beginning were there as mentors in my early days. You are the backbone engineers. I really appreciate your contributions in my life.

My sister Glory Manambowoh Leboh, who is and remains a sister indeed especially in difficult moment. May the Lord bless and prosper you to higher heights.

Dr Syfrizayanti, who was not only a colleague, but a friend indeed. Her input into my work is immense and I am indeed grateful.

Former and current members of the Functional Genome Analysis, especially, Marie Christine Leroy-Schell, Anke Mahler, Martin Ziegler, Katrin Hufnagel and Anke Eiben for a conducive working environment.

Dr Christian Betzen and Martin Ziegler, whose German language proficiencies were highly indispensable for the German text of the summary. I really do appreciate this input of yours in my thesis.

Members of the Clayton's Lab, especially Dr. Esteban Erben, Claudia Hartmann, Igor Minia, Diana and Elisha Muchunga for their kind assistance with in-vivo experiments.

I am especially thankful to the DAAD who financed my study here in Germany. Without this financial support, the realization of this work would have remained a dream.

The group of Dr Jammoneau Vincent at the CIRDES for sample collection and characterization and the the Genomic core facility of the DKFZ (Dr Frank Schwarz and

Mrs. Kersten Mohr) for Y2H and Dr Melanie Bewerunge-Hudler and Sabine Henze for mRNA/ miRNA profiling

Finally, I’m grateful to all those who assisted me in one way or the other during the course of this exercise either by reading and corrections or in any form, may you find herein my sincere gratitude.

Table of Content

List of Figures

Part I

Part II

List of Tables

Part I

Part II

List of Abbreviations

AAT	Amino Acid Transporter
ARE	AU-rich Element
AT	African trypanosomiases
BCA	Bicinchoninic Acid Assay
Bla	Blasticidine
CAT	Chloramphenicol Acetyl transferase
CATT	Card Agglutination Test for Trypanosomiasis
cDNA	Complementary DNA
CLN	Cervical Lymph Node
CNS	Central Nervous System
co-IP	co-immunoprecipitation
COX	Cytochrome Oxidase
CSF	Cerebrospinal Fluid
DAPA	DNA Array to Protein Array
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
ESAG	Expression Site Associated Gene
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GST	Glutathione-S Transferase
HAT	Human African Trypanosomiases
HAT-PCR-OC	Human African Trypanosomiases-Polymerase Chain Reaction Oligochromatography
IFA	Immunofluorescence Assay
IFN	Interferon
IgM	Immunoglobulin sub type M
IL	Interleukin
IPA	Ingenuity Pathway Analysis

IPTG	Isopropyl β -D-1-thiogalactopyranoside
KSRP	KH-type Splicing Regulatory Protein
LAMP	loop-mediated isothermal amplification
LED	Light-emitting diode
MAD	Median Absolute Deviation
MIST	Multiple Spotting Technique
MST	Microscale Thermophoresis
NAPPA	Nucleic Acid Programmable Arrays
NASBA	Nucleic Acid Sequence-based Amplification
NECT	Nifurtimox-Eflornithine Combination Therapy
NF-kB	Nuclear Factor Kappa B
Ni-NTA	Nickel Nitrilotriacetic Acid
NP40	Ninodet P-40
OC	Oligochromatography
OE	Optical Elipsometry
OIRD	Oblique-Icidence Reflectivity Difference
ORF	Open Reading Frame
PABP	PolyA-Binding Protein
PCR	Polymerase Chain Reaction
PGK	Phosphoglycerate Kinase
PISA	Protein *in situ* Arrays
PMSF	Phenylmethanesulfonyl Fluoride
PTB	Polypyrimidine Tract Binding
PVDF	Polyvinylidene Difluoride
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RBP	RNA-Binding Protein
RBS	Ribosome Binding Site
RISC	RNA-Induce Silencing Complex
RNA	Ribonucleic Acid
RNAi	RNA Interference
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SE	Standard Error
snRNA	Small Ruclear Ribonucleic Acid

SPR	Surface Plasmon Resonance
Tbg	*Trypanosoma brucei gambiense*
Tbr	*Trypanosoma brucei rhodensiense*
TgsGP	*Trypanosoma brucei gambiense* surface Glycoprotein
THT	Trypanosome Hexose Transporter
TIF	Tagged File Format
TLR	Toll-like Receptor
TNF	Tumor Necrosis Factor
TRAF	TNF Receptor-Associated Factor
tRNA	Transfer RNA
U	Units
µM	Micro Meter
UBP	U-Binding Protein
UTP	Uridine Triphosphate
UTR	Untranslated Region
VSG	Variable Surface Glycoprotein
WHO	World Health Organization
Y2H	Yeast two-Hybrid
ZFP	Zinc Finger Protein

Summary

The last decades have been galvanized by efforts to reduce the ever widening gap that exist between functional genomics and proteomics. Assigning molecular functions to, and breaking through the complex networks in which each and every individual protein of the cellular proteome is involved is the breathtaking task that needs to be overcome in order to understand the molecular and physiological basis underlying health and disease. System-wide analysis is an approach that can permit a better understanding of the proteome, and there is need for robust and reliable platforms for such analysis to be developed. To contribute to current efforts, I have developed and optimized methods for the production of functional protein microarrays in a miniaturized form from cDNA products and genomic DNA as template source. I have further shown that such arrays are very useful and reliable in various applications such as protein-protein interactions, protein-RNA interactions, Kinase substrate identification and antibody selection. I have validated some data obtained from these arrays in vivo in a model organism, *Trypanosoma brucei*. This confirms that the platform can well contribute to the already existing proteomic tools in generating reliable biological data. Moreover, using cDNA products allow for the analysis of disease-related and rare transcripts as well as other spliced variants.

Current diagnostic tools for human African trypanosomiases are very invasive and in some cases are not sensitive enough. This is compounded by a highly heterogeneous seropositive patient population that is difficult to classify. To address this, I have equally analyzed the miRNA and mRNA expression pattern in the peripheral blood of patients in search for new markers. Thirteen differentially expressed miRNAs were identified, three of which (miRNA-199a-3p, miRNA-27b and miRNA-126*) were highly selective (>95%). These miRNAs have also been reported to be differentially regulated in other diseases and miRNA-199a-3p for example is used as a diagnostic biomarker. They are therefore not suitable as specific biomarkers in sleeping sickness. I have however shown that there is deregulation of miRNA expression following *T. brucei* infection, and most of the differentially regulated miRNAs are related to host immune responses to infectious agents and other inflammatory responses that may be influencing disease outcome.

Zusammenfassung

In den vergangenen Jahrzenten wurden erhebliche Anstrengungen unternommen, die wachsende Kluft zwischen funktioneller Genomik und Proteomik zu schließen. Jedem einzelnen Protein molekulare Funktionen zuzuordnen und die komplexen Netzwerke, in die jedes Protein des zellulären Proteoms eingegliedert ist, zu entschlüsseln, ist die wesentliche Aufgabe, die erfüllt werden muss, um die molekulare und physiologische Basis von Gesundheit oder Krankheit zu verstehen. Die systemweite Analyse ist ein Ansatz, der es erlaubt, das Proteom besser zu verstehen, daher ist es nötig robuste und verlässliche Plattformen für solche Analysen zu entwickeln. Um zu aktuellen Bestrebungen beizutragen, entwickelte und optimierte ich Methoden zur Produktion funktioneller Protein-Microarrays mit cDNA-Produkten oder genomischer DNA als Matrize. Desweiteren zeigte ich, dass solche Arrays zuverlässig für verschiedenste Anwendungen wie Protein-Protein-Interaktionen, Protein-RNA-Interaktionen, Kinase-Substrat-Identifikationen und Antikörperselektion benutzt werden können. Ein Teil der generierten Daten wurde im Modellorganismus *Trypanosoma brucei* validiert, was bestätigt, dass die entwickelte Plattform sich nahtlos in die etablierten Werkzeuge der Proteomik zur Erzeugung verlässlicher biologischer Daten einfügt. Darüberhinaus erlaubt die Verwendung von cDNA-Produkten die Analyse krankheitsassoziierter und seltener Transkripte, sowie von Splicevarianten.

Aktuell verfügbare diagnostische Methoden für die afrikanische Schlafkrankheit beim Menschen sind nicht hinreichend sensitiv und die äußerst heterogene und daher schwierig zu klassifizierende serumpositive Patientenpopulation erschwert dies noch. Die Bestimmung des Krankheitsfortschritts erfolgt invasiv und belastet den Patienten, zudem werden hochtoxische Arsenverbindungen zur Therapie verwendet. Um die Situation zu verbessern, werden zuverlässigere und weniger invasiv bestimmbare Biomarker benötigt. Zu diesem Zweck habe ich gleichermaßen miRNA- und mRNA-Expressionsmuster im peripheren Blut von Patienten untersucht. Dreizehn differentiell exprimierte miRNAs konnten identifiziert werden, wovon drei (miRNA-199a-3p, miRNA-27b and miRNA-126*) hochselektiv (>95%) differentiell exprimiert waren. Es ist bekannt, dass diese miRNAs auch bei anderen Krankheiten dereguliert sind und miRNA-199a-3p wird beispielsweise bereits als diagnostischer Biomarker genutzt. Daher sind sie nicht als spezifischer Biomarker für die Schlafkrankheit geeignet. Ich konnte jedoch zeigen, dass infolge einer Infektion mit *T. brucei* Deregulierung der miRNA-Expression auftritt und dass der Großteil der deregulierten miRNAs mit der Immunantwort gegen infektiöse Erreger und andere Entzündungsreaktionen, die den Krankheitsverlauf beeinflussen, zusammenhängen.

Part I:
Production of Protein Expression Microarrays from cDNA Products for Functional Proteomic Analysis in Model Organism: *Trypanosoma brucei*

1 Introduction

As genomic databases are being filled with the complete genome sequences of several prokaryotic and eukaryotic organisms, the breathtaking challenge of assigning functional and molecular annotation to the thousands of predicted gene products is becoming unprecedentedly evident[1]. Although genes are involved in the modulation of cellular function, they only do so indirectly via the proteins for which they code for. Hence, proteins are the chief players controlling the highly complex, dynamic but well-structured and integrated metabolic, signaling and cellular networks that sum up to form organisms. System-wide analysis of the molecular interactions of proteins can allow for a comprehensive understanding of the intricate signal transduction networks in which proteins are involved. This will improve our view of the complex processes that support living systems and make sense out of health and disease states. Unlike previously thought, the one gene-one protein concept is long outdated, and the genome of some organisms such as humans code for as much as about one hundred thousand individual proteins[2]. Added to this already highly complex genomes, is the phenomenon of post-translational modifications that leads to even a higher level of complexity with more proteoforms and therefore added molecular functions for a single protein[3]. While proteins are responsible for the control and maintenance of normal cellular homeostasis, deregulations in protein expression and function can lead to several diseases such as diabetes, cancer, prions diseases and other neurological disorders[1].

Taken together, it is evident, that complete analyses of the cellular proteome is an inevitable task, and conventional proteomic tools such as Enzyme-Linked Immunosorbent Assays (ELISA) and western blots are less adapted for system-wide analysis of the proteome. This highlights the need for the development of robust but reliable tools for the global analysis of protein expression and function. Such tools will allow for parallel analysis of protein function, thereby bridging the gap that has long existed between genomics and proteomics, and help improve our understanding of the mechanistic basis of several biological processes. Although it may sound conceptually trivial, considering that such tools already exist for genomics, it is however important to note, that proteomic is far from being the protein equivalent of genomics. The highly heterogeneous nature of the building blocks of proteins and the physico-chemical properties of individual proteins add several strata of complexity to proteomics[4]. In fact, unlike nucleic acids, there is yet no protein-based analogue of the polymerase

chain reaction for the amplification of individual proteins or the post-translational modifications they undergo. More so, there can be a great disparity in the physico-chemical properties of functional elements within the same protein molecules. Furthermore, proteins do not have well defined high affinity and high selectivity binding partners that can easily be used for their immobilization on array platforms as is the case with nucleic acids[5]. The development of high-throughput tools for proteomics will therefore need to take all of these into consideration and address them individually for every protein of interest. This has been one of the major drawbacks that have retarded progress in the development of thigh throughput tools for proteomics. Nonetheless, during the last two decades, some significant efforts have been made to such ends with encouraging results[1].

Since their first conception about two decades ago in the multianalyte microspot assays[6], protein microarrays have constantly evolved and have been at the center proteomics both in basic and biomedical research where they have helped fuel growth in these fields. These miniaturized protein chips offer far-reaching potentials and advantages, such as multiplexing and parallel analysis of several analytes, high sensitivity, cost and time efficiency, and the need for relatively small sample volumes[5, 7]. The latter is very valuable in biomedical research where sample material is often very limited. Besides, protein microarrays allow for the analysis of direct contact between molecules, therefore providing the opportunity to study weak and transient interactions that would otherwise be difficult to predict or analyze with other methods[8]. Owing to its assets, protein microarrays have extensively been exploited in diverse domains in basic and clinical research. As such, they have been used for interactome mapping[9], analysis of protease-associated networks in malaria[10], patient screening[11], analysis of post-translational modifications[12, 13], identification of RNA/DNA-binding proteins[14], study of host-pathogen interactions[15], biomarker and drug discovery as well as antibody selection[16-18]. Despite the aforementioned, and even more potentials this platform may hold, its development and exploitation is relatively still at a very premature level. This is in part due to basic technical problems such as the maintenance of a well folded 3-dimensional structured protein, large scale production of pure but functional proteins, the lack of major post-translational modification and well-adapted surface chemistry that provide for high affinity and oriented protein immobilization.

An ideal situation would be one in which protein are immobilized on high affinity surfaces; that allow for proper immobilization of the proteins in their native 3-dimensional structure, is inert to non-specific immobilization, permit proper orientation of proteins, thus exposing all active sites, and do not necessitate pre-purification[19]. This is however only a hypothetical ideal situation which is not met in real life situations, and there is therefore no universal platform for protein microarrays yet. Nonetheless, several strategies have been adopted for the development of protein microarrays. Protein prints and *in situ* synthesis are the most popular, with the former being a method of choice for analytical as well as functional protein microarrays.

Regarding functional protein microarrays, protein printing has been used to generate high-density functional arrays[20-22]. The advantage of this approach is that the exact amount of protein on each spot can be evaluated, and the proteins are relatively of high purity. Moreover, the expression of eukaryotic proteins in eukaryotic host such as yeast, may allow for some post-translational modifications. However, it suffers from the challenges that are related with expression and purification of proteins in large amounts, and also the fact, that some proteins are toxic to host cells and tend not to be expressed. Besides, it requires that all genes of interest be cloned into plasmids and fused to affinity tags for purification, which is neither cost, time nor labor efficient. The size of some tags may actually be problematic to the proper functioning of the entire protein[19]. Exploring other avenues to circumvent these problems, cell-free expression has been seen to be one of the most attractive options, and has extensively been used to produce protein microarrays using several strategies. In principle, cell-free transcription/translation systems, allow for the direct synthesis of proteins from cDNA templates, which can either be cloned into expression plasmids or generated as PCR products[23, 24]. Proteins can therefore be synthesized more easily, without the need for cloning and spotted onto microarray support platforms or directly on solid surfaces. This has the advantage of being cheap, time and labor efficient, but also allow for the synthesis of otherwise toxic proteins and has been exploited for the production of high density protein microarrays[25, 26]. To add to the relatively already simplified approach, *in situ* cell-free expression has the added advantage, that proteins are synthesized *in situ,* and therefore help resolve the problem of protein availability. Although this approach is very promising, it however falls short in terms of protein purity, therefore limiting their applications only to some applications, More so, the fact that only approximations can be made with regards to protein concentration could linit

its applications. Besides, since protein expression is not uniform across the entire array, there is a great need for normalization to avoid bias.

Protein *in situ* Arrays (PISA), one of the most recent and popular strategies employed for the development of protein microarrays was introduced about a decade ago[27]. It has been emulated with more technical development and miniaturization (**Fig.1**). The Nucleic Acid Programmable Arrays (NAPPA)[28], the DNA Array to Protein Array (DAPA)[29], puromycine capture[30], the HaloTag[31], and the Multiple Spotting Technique (MIST)[32] are different versions of PISA developed to date. With the exception of the NAPPA which recently produced arrays of thousands proteins for functional screens[33], most of the others have remain very less exploited or at least only at a low throughput level since they were first reported. It is plausible to speculate, that issues related to less optimal performances of these arrays as well as other technical problems could, in part explain their limited applications. In effect, if PISA chips must be integrated into system-wide analysis of protein function, they should not only be robust, but should be able to generate large but biologically meaningful data that is synchronizable with those obtained from other platforms. Nonetheless, given the numerous challenges facing protein microarray production and exploitation, reliability and reproducibility is not always guaranteed in every situation. The aforesaid therefore suggest that if much is expected from protein microarrays, much should be invested, not only into its development and evaluation, but also enhance its availability.

The present study was therefore designed, to optimize, and evaluate the multiple spotting technique, for the production of PISA, and to investigate its functional performance and scope of application, based on a model organism: *Trypanosoma brucei*.

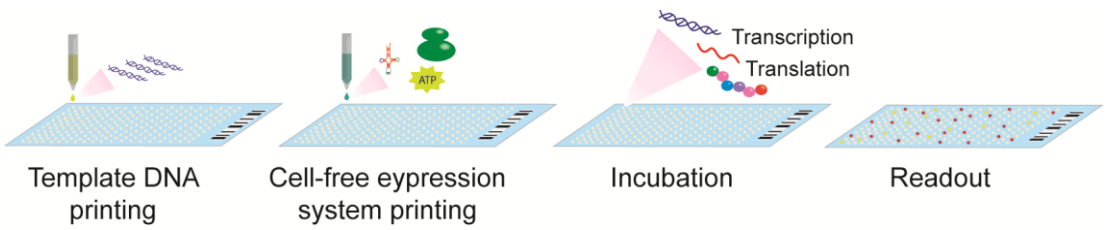


Figure 1: Schematic representation of the multiple spotting technique (MIST).

Aim

This work was aimed at optimizing and producing functional protein microarrays by the multiple spotting technique and evaluating the scope of applicability and reliability of such arrays for proteomic analyses.

Objectives

To arrive at the above-mentioned aim, the following objectives were set forth:

- Evaluation of cDNA products as template source for the production of PISA by MIST.
- Optimization of technical parameters for optimal array quality and performance.
- Establishment of reliable protein detection method.
- Evaluation of the scope of applicability of MIST-based protein microarrays.
- Data validation by complementary approaches.

1.1 Proteomics tools: State of the Art

Following the completion of the human genome project in 2003, the development of DNA microarrays accelerated and they fast became very robust and reliable tools with unprecedented analytical performance. This has earned this platform wide-spread applications ranging from pharmacogenomics[34], single nucleotide polymorphisms[35], gene expression analysis[36], biomarker discovery[37] and identification of transcription factor binding sites[38] among others that marked the genomic revolution. Despite these major breakthroughs, the contribution of DNA microarrays to the understanding of the mechanisms that underlie the fundamental basis of life and making the difference between health and disease is very limited. This is mainly as a result of the fact, that DNA in most cases, is only an indirect actor in the control of several cellular processes via the protein for which they code, and the great disparity between mRNA and protein abundance within a given cell[39]. To complement functional genomic data and bridge the gap between genomics and proteomics, the post-genomic era would have to develop tools with better or comparable performance as the DNA microarray. The proteomic era came into play, with the rapid development/optimization of high throughput proteomic tools such as phage/ribosome display, yeast two-hybrids, mass spectrometry, and protein microarrays.

Although microarrays in general, and protein microarray in particular can be potential source of many false positives[40], they also can generate high accuracy data[41], are very sensitive, allowing for parallel analysis of multiple analytes and require highly reduce sample volumes[42]. They also can provide direct evidence for a physical contact between two or more analytes, and allow for the analysis of transient or weak interaction, which would otherwise be left out[8]. Meanwhile techniques such as pull-down, co-Immunoprecipitation (co-IP), and display may allow for in vivo analysis, they however suffer from inherent limits that circumscribe their scope of applications. For example, pull-down and co-IP are not only labor intensive and biased in favor of high affinity interactions, but do not provide any information about any physical interactions[43]. Display on the other hand, is also labor intensive, requiring effort to be invested into cloning, and fusing proteins to the relatively large surface protein for display. This can greatly alter the folding, and hence three-dimensional structures of the protein and its functionality[44]. Moreover, all these methods do not allow for simultaneous multiple analyte analysis, and therefore not well suited for present day proteomics. Although yeast two-hybrid has been developed for high-throughput applications, it equally cannot be applied in all settings. For example, transcriptional repressors are not suitable for this method; auto-activating baits could potentially lead to very high false positive rates, while proteins that do not localize to the nucleus may lead to false negatives. More so, if an intermediary protein brings the prey and bait to very close proximity, this could also lead to false positive interactions[45]. Though at its very early stage, protein microarrays appear to be the most suited for current day proteomics. This becomes even more evident, when one compare the numerous applications for which proteins microarrays have been used in recent years while considering its relatively developing stage with respect to other well established methods.

1.1.1 Protein Microarrays

Based on their applications, two major types of protein microarrays can be distinguished: analytical protein microarrays and functional protein microarrays. The former are generally composed of well-characterized molecules with specific binding activities such as antibodies. These arrays are often used to characterize the content of highly complex biological samples such as serum and cell lysates and therefore can be used for profiling. As such, they have been widely exploited both in basic and applied research for the identification of posttranslational modification[46], biomarker

discovery[47, 48], proteomic profiling[49], network analysis[50] and disease proteomics[51]. Since there is yet no available technology for *in situ* synthesis and assembly of entire antibodies, production of analytical arrays still rely on protein prints.

Functional protein microarrays, consist of a collection of individual protein molecules that are immobilized on solid planar surfaces in a well-organized fashion, such that positional information allow for the identification of individual proteins. In general, these arrays are used to query the biochemical properties and activities of the arrayed proteins. With the advent of *in-situ* cell-free expression, it is now possible to synthesize virtually all proteins of a given organism, not considering proteoforms that result from post-translational modification. Protein printing has also been widely used for the production of functional protein arrays, although some toxic and membrane proteins cannot be produced in this manner.

1.1.2 Production of Protein Microarrays

Protein printing and *in situ* synthesis are the currently available methods used for the production of functional protein microarrays. In fact, high throughput protocols for the expression and purification of proteins have been developed in yeast and bacteria. In yeast for example, using homologous recombination, a total of about 6000 proteins have been cloned into GST-fusion vectors, expressed and purified in 96-well plate model as GST-tagged N-terminal fusion proteins. Following this strategy, it is possible to purify as much as 1152 proteins per day[20]. In bacteria, an automated procedure for expression and purification of proteins as 6×His-tagged recombinant proteins has been developed and has permitted the production of about 4000 individual proteins in a single day[21]. These proteins are then printed onto solid surfaces using contact or non-contact printers to generate corresponding arrays.

Meanwhile these protocols have greatly enhanced the throughput of protein expression and purification, cell-free expression has done even better. Using the wheat germ extract, more than 13,000 human proteins have been produced simultaneously for array purposes. This clearly out performs the results obtained from all high-throughput *in vivo* expression procedures developed to date and highlight the potentials held by this approach[25, 26]. Despite efforts geared toward the exploitation of this strategy for the production of protein microarrays, most, if not all of the protein *in situ* arrays developed to date still remain with a limited scope of application. In fact, since its first description

more than a decade ago[52], diverse successions of strategies based on this method have been developed including:

1.1.2.1 The Nucleic Acid Programmable Arrays (NAPPA)

Developed about a decade ago[28], this method uses biotinylated plasmids containing ORFs coding for the target protein as a GST-fusion protein for template source. The plasmids are immobilized side-by-side with capture antibodies (anti GST antibodies), and incubated with a cell-free expression system (Mammalian reticulocyte lysate) for the expression of the target protein. After successful transcription and translation, the nascent peptide is then captured by the by-waiting antibody. This method was later developed and used for the production of protein microarrays of about a thousand proteins[33]. As one of the most exploited PISA, NAPPA has extensively been used in domains such as immune response tracking[53], antibody profiling in cancer[54, 55], assessment of bacterial immunogenicity[56] and biomarker discovery[57].

Based on a similar approach, the HaloTag NAPPA array was developed, and used for the analysis of protein-protein interactions for 12,000 Arabidopsis (Promega). This is one of the most interesting large-scale exploitation of this array type arrays since it was conceived.

1.1.2.2 The DNA Array to Protein Array (DAPA)

The DAPA uses a sandwich model for repeated synthesis of “near pure” protein microarrays from DNA microarrays[29]. A permeable membrane is saturated with cell-free expression mixture, which is then laid between a template-containing array, and another slide containing an affinity capture agent that immobilizes the nascent peptide. The advantage of this technology is the fact, that the resulting arrays contain individual proteins at a “near pure” state, and that the original DNA microarrays can be repeatedly used to produce protein arrays on-demand. More so, it uses PCR-derived fragments as templates, therefore circumventing labor intensive and time-consuming cloning procedures[58]. Though recently optimized for performance[59], DAPA arrays have remained with limited, if at all any functional application despite the advantages it brings to high-throughput proteomics.

1.1.2.3 Puromycine Capture from RNA Arrays

This technique is also based on the PISA strategy, but with a different capture agent. Here, a ribosome is installed at the end of an RNA template, which permits the immobilization of nascent polypeptides by a puromycin moiety that is grafted to an oligonucleotide immobilized on a solid surface[30]. This certainly allows for a well-oriented immobilization of the nascent peptide, and purification. Just like the DAPA, since its description nearly a decade ago, this strategy has not recorded any significant functional application.

1.1.2.4 The Multiple spotting technique (MIST)

The MIST combines cell-free expression and *in situ* protein synthesis for the generation of high density protein microarrays. The major difference between DAPA and MIST is the fact that instead of saturating a membrane with cell-free protein expression system, it is spotted in a subsequent spotting on the top of the template DNA using non-contact printers[32]. This procedure does not only reduce reaction volumes and therefore is cheap, it also uses unpurified PCR-generated templates. It is therefore considered a miniaturized version of the PISA technique. If affinity tags are incorporated into the expression template via PCR, this eliminates the need for cloning, and allows for the use of affinity capture for the immobilization of the nascent peptide. This enhances the purity of the captured proteins, and makes them more suitable for functional applications. Although this technique is cost, time, and labor efficient, just like the DAPA and puromycine capture arrays, it has recorded very little, if any application since it was first developed.

1.1.2.5 Other Techniques

As each of the above-mentioned technological developments contributed to the alleviation of some of the drawbacks associated with protein in-situ array, some stones were left unturned. For example, the need to generate tag-free native proteins, recovery of protein-coding DNA or RNA and the need to enhance purity while maintaining functionality at high-throughput among others, have motivated the development of other systems for the fabrication of protein microarrays. As such, the Intein-based approach, photo cleavage and the human protein factories have been developed[60-63]. Meanwhile these approaches offer the possibility to produce protein microarrays for customized applications; they also retain some bottle-necks including the need for

cloning, the use of modified amino acids, and the involvement of multiple processing steps with potential risk of material loss.

The above-mentioned strengthen the fact, that the route to an ideal protein microarray is not a nearby destination. However, the state of the art, though not the best, is already good enough to take proteomics to the next level. Ideally, considering protein microarrays as a routine tool for proteomics, one would expect that its production should be relatively simple and cheap. It should also have reasonable complexity and long shelf half-life with simple and straightforward methods for the detection and analysis of the immobilized proteins. In this light, it is possible that protein *in situ* arrays may hold the key into "next-generation" proteomics.

1.1.3 Applications and Impact of Protein *Microarrays*

In the course of the past decade, microarray technology has evolved rapidly both in terms of technological development, and scope of applications. Thus, it has moved from the tool of basic analytical research, as it was first seen to be, to a more sophisticated platform. It has been exploited in diverse applications in fields such as protein expression profiling, molecular interaction mapping, drug discovery, disease diagnosis, toxin detection and vaccine development[1].

1.1.3.1 Applications in Basic Research

Functional protein microarrays have marked its presences in due course, and have found applications in both basic and biomedical research. Most successful have been the expression-based protein arrays, but protein *in-situ* arrays have also recorded some basic applications. In basic research, both array formats have been used for the analysis of protein-protein interactions, wherein, HaloTag NAPPA has been used to map protein-protein interactions of transcription factors in Arabidopsis (http://www.promega.de/). Meanwhile, expression-based arrays have been used for several such applications including interactome mapping[9]. In immunoprofiling, PISA just like expression-based arrays have recorded diverse applications ranging from biomarker discovery[64], autoimmune disease profiling[55], analysis of immune responses[65] and immunogenicity studies[56].

A Protein-Protein Interactions

PISAs have successfully been used for the study of protein-protein interactions[66]. It is estimated that almost every protein performs its biological function by interaction with other proteins or biomolecules. The physiological and pathological events that underlie health and disease are all equally protein interaction-driven processes[4]. It is therefore crucial to understand the interactions that exist between proteins and other biomolecules, as this gives more insights into signal transduction pathway mapping, protein function determination, identification of enzyme substrates, understanding of metabolomics and systems Biology[67, 68]. Besides, protein interaction studies are key to drug discovery and personalized medicines. Protein microarrays has been applied for early drug target identification, target validation, mechanisms of drug action, specificity profiling and lead optimization studies[69]. Contributing in this area, PISA have been used to analyze the interaction of proteins that are involved in DNA replication[28], and mapping of protein interaction in cancer-related proteins such as c-Jun, TP53, and c-Fos[70].

B Toxin Detection

One of the most fascinating applications of PISA has been its usages in toxin detection. Considering that some substances have the potentials to inhibit protein synthesis, protein *in situ* arrays have been used to evaluate the inhibitory effects of some antibiotics on protein synthesis[71]. Such an approach could be useful in the elucidation of the mechanism of drug action, profiling of off-targets effects and toxin identification. This application has the great benefits of rescaling the current efforts and resources that are being invested into drug discovery.

It is worth noting that with the advent of protein *in situ* arrays, some otherwise unattainable goals in high throughput proteomics can be envisaged. For example, despite the potentials held by membrane receptors in drug targeting and signal transduction, their exploitation is barely reported in literature. Nonetheless, a reasonable percentage of the few that have so far been studied are currently being used as drug targets[72]. The toxicity and insolubility of most membrane proteins is the major bottleneck explaining the difficulties encountered during their characterization. Cell-free protein expression has offered an attractive alternative to such problems, and has been used to successfully synthesize functional membrane proteins[73].

C Protein-DNA/RNA Interactions

Beyond the above enumerated applications, protein microarrays have also been exploited in other field of biology and biochemistry. As such studies aiming at understanding the regulation of gene expression have successfully been undertaken using protein microarrays. Gene expression control can be mediated by cis-acting and/or trans-acting elements that are capable of modulating the fate of every gene and its product. Steroid hormones are general modulators of gene expression, and protein-lipid interactions have been made possible in high throughput formats thanks to protein microarrays. In yeast, several lipid-protein interactions have been identified using an array of about 5800 proteins[22]. Such discoveries are not only essential for the understanding of gene expression control, but also for the understanding of the mechanism of active or facilitated transport aided by lipid-bound membrane protein.

DNA-binding proteins such as transcription factors, nucleases and polymerases and histones as well as RNA-binding proteins are key modulators that control diverse cellular and molecular functions, by virtue of their capacity to bind to nucleic acids, and influence transcription and translation[74]. Aided by protein microarrays, several novel, but also other well-characterized protein-DNA/RNA interactions have been identified and characterized[75-78]. The role of RNA/DNA-binding proteins in the development and progression of genetic disorders such as cancers is now well-established[78, 79]. The characterization of RNA and DNA-protein interactions could foster the understanding of the mechanisms underlying tumorigenesis, DNA damage and repair, RNA stability and decay metastasis and even the complex mechanisms at the basis of cancer development and cancer cell survival.

D Analyses of Post Translational Modifications

Proteins are most often the targets of a plethora of post-translational modifications, whose entire scope is not well established. It is estimated, that the scope of this dynamic process could increase the complexity of the cellular proteome to as high as 10^6 fold[80]. As a result of their dynamic nature, post-translational modifications can greatly alter the basic properties, and hence the molecular functions of a protein. This can result for example in swapping its function from either from activator to inhibitor or from active to inactive state. This has as consequence, the triggering of cellular pathways, controlling cell fate and making vital decisions of life and death for a given cell[1]. Functional arrays have also been used in this domain for

the identification and characterization of diverse post-translational modifications such as glycosylation[81], ubiquitination[82], acetylation[83], nitrosylation[84] and protein phosphorylation[12]

E Analysis of Host-Pathogen Interactions

Most obligate parasites and other infectious agent rely on the hijack of host metabolic pathways for their survival. Viruses are well known for this as well as some bacteria [85]. It has been shown that the Epstein–Barr virus kinase is capable of activating an upstream mediator of DNA damage response, the histone acetyltransferase TIP60. It is also capable of using the chromatin remodeling function of the same protein during a positive feedback loop to enhance the expression of own genes necessary for viral replication[19]. Functional protein microarrays have been used for the analysis of host-pathogen interaction in viral infections[15]. Exploring such avenues could help identify host genes that are vital for viral replication, and provide a framework for the development of antiviral agents.

1.1.3.2 Applications in Biomedical Research

A Biomarker Discovery

The need for reliable biomarker has been on a constant increase, as they have been shown to be very informative in biomedical research. Biomarkers allow for early identification of disease states, treatment monitoring, and following disease prognosis[86]. To this end, PISA arrays have been valuable tools, and thus have been used extensively. NAPPA-based arrays have been used for the detection of tumor-specific autoantibodies, and multiplexed detection of serum antibodies that are crucial for diagnosis and monitoring of immune responses[54, 87].In the field of vaccine development, PISA, and precisely NAPPA have been used for the identification of vaccine candidates in *Plasmodium falciparum*, with more than seventy highly immune-reactive proteins being identified[88]. Furthermore, NAPPA have been used for successful identification of diagnostic biomarkers for Q-fever, and for the analysis of immunogenicity of the outer membrane of *Pseudomonas aeruginosa*[56, 89].

B Drug Discovery

A reasonable proportion of all validated drug targets are proteins and most reported side-effects of most drugs are due to undesired interaction of the active

substance or other compounding components of a drug formulation with biomolecules, especially proteins[90]. Proteins have now become one of the major focuses for the design and development of drugs that can modulate the biological function of cellular pathways. However, without sophisticated screening tools, identification of biologically active molecules from huge libraries, on complex systems such as the cellular proteome can be a very daunting task[1]. Functional protein microarrays have started to be seen as one of the tools that could alleviate such challenges, and reduce the present disproportionate consumption of resources and efforts. As such, methods have been developed[91], and used for screening protein microarrays with small molecules[91]. Besides, biotinylated Small Molecule Inhibitors of Ripamycin (SMIR) have been used to screen a yeast proteome in microarray format. This has led to the identification of novel candidate targets of SMIR, and help established the binding profile of such molecules on the yeast proteome[92]. This, if performed on other proteomes such as the human proteome, could identify novel drug targets, but also help identify the ground causes of some observed side effects.

Despite these significant applications, the exploitation of PISA still lags behind that of expression-based arrays, which for example have been used for the analysis of posttranslational modification[13], study of enzyme kinetics[12], drug discovery[93], host-pathogen interactions[15] and the identification of RNA-binding proteins[14].

1.1.4 Current Challenges facing protein microarrays development and exploitation

Despite recent developments, several drawbacks have impeded effective application of protein *in situ* microarrays in several fields. Most crucial of these, include, protein immobilization (appropriate surface chemistry). In fact, the success or failure of every microarray experiment, depends on the suitability of the properties of the surface chemistry used[94]. Protein purity, functionality, large scale expression of functional proteins and cheap methods for template generation are some major issues to be dealt with. Besides, although the current detection strategies available might be of relatively good resolution with up to single molecule detection[95], they might simply be just too expensive for routine usage, or may require analyte manipulations that can put the functionality of the proteins in jeopardy.

1.1.4.1 Immobilization and Surface Chemistry

There is still no universal platform applicable to all array types and screen, and the right surface chemistry for every application needs to be determined empirically[5]. Some surfaces perform well for one application, resulting in an improved signal to noise ratio, and for another application, they are not suited. For example, FAST slides have been shown to perform well for the analysis of protein-DNA interactions and acetylation assays but not for phosphorylation[96, 97]. Nickel chelated slides are well adapted for acetylation, but not phosphorylation[97]. Ideally, a surface coating should provide, not only for specific and oriented capture with high affinity, but should be able to yield a good signal to noise ratio and descent spot morphology with long shelf life[98]. All such properties are however difficult to have on one surface type. For example, three-dimensional matrices such as agarose and polyacrylamide have been developed and used to coat glass slides for microarray fabrication. These surfaces have the advantages of retaining protein functionality, with high binding capacities. They also exhibit limited lateral diffusion as compared to PVDF, nitrocellulose and polystyrene, They however require highly sophisticated equipment for high density loading of the nanowells[19]. Non-covalent adsorption allows for high capacity binding, with little impact on the protein structure, but the amount of protein and its orientation is not controlled. As such, reaction efficiency, reproducibility and accuracy are variable and it has been found, that protein activity is reduced up to 10-fold when proteins are immobilized randomly[99]. Oriented capture, especially affinity capture seems to be an option of choice. However, the introduction of large affinity tags may affect protein folding and hence functionality, meanwhile some of the available surfaces for affinity capture yield very poor signal to noise ratios.

1.1.4.2 Data Readout

A Label-Dependent Detection

Several methods exist for microarray readout, and are grouped under two main principles: label-dependent and label-free detection[5]. In label-dependent detection methods, radioisotopes and fluorescent dyes are often used to detect molecules bound to the microarray. The use of radioisotopes is one of the most ancient detection strategies that have been used in several biological applications. It has the advantage over fluorescent dyes that it is more sensitive and reliable, but need the manipulation of

radioactive substances. Most protein microarray-based enzyme assays to date have relied on radiolabelling [18, 83, 97]. Fluorescent dyes, such as Cy3/5 and other homologues are now becoming very popular for probe detection. They circumvent the need to manipulate radioactive substances, but are less sensitive than radioisotopes. Their relatively large size could influence the outcome of an experiment, or even make it unsuitable for probes with low molecular weight. Fluorescent dyes, also have the advantage of reducing cost, as they can be used for simultaneous detection and direct comparison of different samples, thereby avoiding chip-to-chip variations[5]. To overcome the problem of sensitivity, other label-dependent, nonradioactive detection methods have been developed. Semiconductor quantum dot labeling[100] and enhanced chemiluminescence[101] have also been applied for sensitive and stable detection on protein microarrays. Signal amplification strategies have also been developed for the detection of samples with low abundance even down to the femtomole level. Rolling cycle amplification[102], and tyramide amplification are now available for more sensitive detection of labeled probes[103]. Moreover, fluorescence-linked immunosorbent assay has been applied for the detection of exposure to tuberculosis toxin at the single molecule level. This increases the sensitivity that can be achieved with label-dependent techniques, but there is need for sophisticated equipment that is not always available.

Other indirect detection methods have also been established, especially for the detection of protein-protein interactions using specific antibodies to recombinant bait proteins[104]. Meanwhile this does not require the manipulation of the protein structure; it however demands that the bait protein be fused to an epitope tag, which could affect the protein folding and function. Besides, it is also possible, that during protein folding, the antibody-binding epitope is localized close to the binding domain of the bait protein. After interaction of both proteins, stearic hindrance could affect antibody detection either by reducing the signal, or excluding it entirely leading to false negative results.

B Labe-Free Detection

Since label-dependent methods always require a modification of the native structure of the probe molecule, label-free strategies have been introduced. These methods allow for real-time monitoring of a binding event using a probe in its native state. Label-free detections generally rely on the measurement of a change in some

optical parameters during the course of a reaction. Imaging Surface Plasmon Resonance Spectroscopy (SPR)[105], Imaging Optical Elipsometry (OE)[106] and Reflectometric Interference Spectrometry[107] are label-free detection methods, that rely on the measurement of the optical dielectric response on a thin film such as changes in the physico-chemical properties of the film with time.

A more sensitive variant of elipsometry; the Oblique-Icidence Reflectivity Difference (OIRD), has also been developed and used for real time analysis of DNA hybridization and protein-protein interactions. This detection method demonstrated its potentials for high throughput application, with a detection limit of 14 fg and a resolution time of about 20 µs[108]. Mass spectrometry has also been applied for microarray readout with the advantage that it is simple, fast, and requires relatively small amounts of samples and is applicable to complex samples such as urine, plasma and cell lysates[109]. Another label-free detection method, the Atomic Force Microscopy, measures topological changes on the microarray surface such as increase in the height of a protein/antibody spot[110].

Label-free detection methods have largely enhanced the sensitivity of microarray readout, and circumvent the need for sample manipulation that can jeopardize the native structure of some proteins. However, the simple fact that most of these methods require very sophisticated equipment and specialized training for their exploitation make them inaccessible to the global scientific community. It is therefore evident that the need for more simple, reliable and cost efficient detection methods for microarray readout is still of prime importance and remains a bottle neck to be overcome.

1.1.4.3 The Issue of False-Positives and False-Negatives

In most high throughput experiments, several samples are analyzed simultaneously and multiple hypothesis are tested. Faced with multiple hypotheses testing, it is possible to make type 2 statistical errors resulting in false-positives, or type 1 error, resulting in false-negatives. There is therefore a need for methods to control the probability of making erroneous conclusions[111]. Moreover, in microarray experiments, molecules that would otherwise not come into contact due to cellular compartmentalization are brought together, and it is possible that such molecules have affinities to one another or simply just stick to each other. Adequate positive and negative controls and the analysis of interaction events as a function of compartmentalization are needed. This is a difficult task, as some molecules can transit

from one compartment to another and some unknown proteoforms of a given protein might specifically localized in unexpected cellular compartments. Large-scale complementary *in vivo* experiments may be helpful in such cases. However, the design and execution of such experiments could be extremely challenging.

1.2 Gene Expression and Control in Trypanosomes

Trypanosomes are unicellular protozoa parasites that are responsible for human and animal diseases in tropical areas especially Africa and South America. Unlike other eukaryotes where transcriptional and post transcriptional regulation of gene expression is well developed and ensures a balance in the steady state of RNA levels, most of their gene expression control is at the posttranscriptional level[112]. Just like other kinetoplast such as Leishmanias, their genetic organization is unusual in that each polymerase II promoter controls the transcription of multiple open reading frames down stream of it. Individual transcripts are then processed and regulated individually by trans-splicing and polyadenylation[113]. It is therefore almost impossible to exert transcriptional regulation for the expression of individual genes with such genomic organization. Besides, some protein-coding genes such as the GPEET procyclic surface coat protein are under the control of polymerase I[112]. Without any transcriptional regulation of gene expression, one would expect, as it is the case, that the combined effects of *cis*- and *trans*-acting elements such as RNA-binding proteins that are involved in RNA processing, translation and degradation fine tune gene expression and determine the fate of individual RNA or subset of RNA molecules. Most of the trans-acting elements bind to *cis*-elements in the 3′ UTR and 5′ UTR, but some are known to bind coding regions even if the later seem awkward regarding translation[114]. It is estimated, that between 100–200 different proteins encoded in the trypanosome genome have RNA-binding properties[115], and about 70 proteins are known to possess RNA recognition motifs (RRM)[116]. These proteins are sub classified according to their functions or RNA recognition domains and play distinct roles.

The ALBA domain proteins, for example, are known to interact with polyribosomes, proteins of the translation machinery such as the cap-binding protein eIF4E4 and polyA-binding proteins [117]. These observations fuel the speculation that even if they may have some other function, they are most probably involved in RNA stability. Another example, the Pumilio domain proteins, a class of ten proteins known in most organisms to be involved RNA decay, RNA transport, rRNA processing and

translational repression by sequence-specific binding to target RNA molecules[118]. In trypanosomes however, the only characterized member of this protein family, the PUF9, has been shown to stabilize a subset of mRNAs [119]. It is well possible, that added to the already known functions of these proteins, they exert other functions that are either cell cycle-specific or that depend on other pressures exerted on the cell. Zinc finger proteins, another class of RNA-binding proteins are known to play diverse roles. Some, such as ZFP1 and ZFP 2 are involved in stage differentiation, meanwhile others such as ZFP3 and ZC3H11 are known to stabilize some RNA transcripts under stress conditions[112]. The polyA-binding proteins, PABP1 and PABP2 binds to the polyA tail of mRNA and stabilizes the mRNA by interaction with the initiation factor eIF4 located at the 5′ cap of mRNA transcripts forming a loop structure that is necessary for efficient translation. Other proteins with RRMs such as the U-binding proteins do not seem to have sequence specific functions, even if they have been shown to be essential[112]. The polypyrimidine trac binding proteins (PTB), yet another class of RNA-binding proteins, are involved in *cis*- and *trans*-splicing, as well as stabilization of some mRNA transcripts and like other mammals, they regulate RNA metabolism from synthesis to degradation[120].

Besides proteins that are involved in RNA processing, translation and stability, others, by their RNA-decay properties, conjugate their efforts to ensure the balance between synthesis and abundance. They are proteins involved in RNA degradation and generally tend to destabilize transcripts and in most cases, this starts with transcript deadenylation. In trypanosomes just as with other eukaryotes, the CAF/NOT complex is responsible for deadenylation. This leads to the dissociation of the PABP and hence release of the loop structure exposing the polyA tail to degradation by exosomes. Meanwhile, removal of the 5′-cap structure following the release of the loop-structure exposes the 5′ end which can then be degraded by 5′ exoribonuclease XRNA[121]. It seems therefore that the action of the CAF/NOT complex in many cases is likely to be a key step into RNA degradation and may also be the limiting step. In trypanosomes, the CAF-NOT complex unlike in other eukaryotes does not have the NOT4 protein, which is known to be an ubiquitin ligase [122, 123]. The complex is made up of the following subunits CAF1, NOT1, NOT2, NOT5 NOT9, NOT10, and NOT11 all of which are most likely, but not exclusively involved in RNA degradation. In addition to its role in RNA decay and ubiquitination, it has recently been shown to have more functions than

previously expected. The CCR4-Not complex has been shown the play an indispensable role in miRNA-mediated silencing by facilitaiting miRNA-dependent decapping and deadenylation[123]. A novel role has also been described for the complex, involving its participation in the regulation of cellular proteostasis and protein quality control by translational arrest[124, 125].

Considering the central role played by proteins and their involvement in the regulation of cellular homeostasis and the subsequent implications in health and disease, understanding the intricate networks underlying such functions is very important. This could allow not only for system-wide analysis of the extent and mechanisms of protein involvement in diverse cellular and physiological conditions, but also improve the outcome of clinical interventions. With the very complex nature of the cellular proteome, reliable and cost efficient methods for simultaneous analysis of protein structure and function need to be developed and validated.

2 Materials and Methods

2.1 Materials

Antibodies

Product	Catalogue N^{o}	Manufacturer
AntiPhosphoserine/Threonine/ Tyrosine Antibody	MA1-38450	Thermo Scientific
Monoclonal Anti-V5, Cy3 conjugate	V 4014	Sigma-Aldrich
Penta·His Alexa Fluor647 Conjugate	35370	Qiagen

Kits

Product	Catalogue N^{o}	Manufacturer
BCA protein assay Kit	23225	Thermo Scientific
EasyXpress Insect Kit II	32561	Qiagen
Magnetic RNA-Protein Pull-Down Kit	20164	Thermo Scientific
miScript II RT Kit	218160	Qiagen
miScript SYBR® Green PCR Kit	218073	Qiagen
PCR Extract Mini Kit	2300600	5 Prime
pET101/102 Directional Topo cloning kit	K101/102-01	Invitrogen
PureLink PCR Purification Kit	K3100-01	Invitrogen
PURExpress™ *In Vitro* Protein Synthesis Kit	E6800S	NewEngland Biolabs
PureYield Plasmid Miniprep System	A1222	Promega
QIAprep Miniprep Kit	27106	Qiagen
RNA 3′ End Biotinylation Kit	20160	Thermo Scientific
RNA 3′ End Desthiobiotinylation Kit	20163	Thermo Scientific
S30 T7 High-Yield Protein Expression Kit	L1110	Promega
TNT T7 Quick for PCR DNA Expression Kit	L5540	Promega
TranscriptAid T7 High Yield Transcription Kit	K0441	Thermo Scientific

Chemicals

Product	Catalogue N°	Manufacturer
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	HN77.4	Carl Roth
Acetic acid	27221	Sigma-Aldrich
Acrylylamide/Bisacrylamide	3029.1	Carl Roth
Agar-Agar,Kobe I	5210.3	Carl Roth
AirPore Tape Sheets	19571	Qiagen
Ammonium per sulfate	A3678	Sigma-Aldrich
Amplicillin (Sodium salt)	M3110	Genaxxon
Bacteria Cell Spreader	89042-018	VWR International
Betaine Monohydrate	B2754	Sigma-Aldrich
Biozyme LE Agarose	840004	Biozyme
Bovine serum Albumine	05470	Sigma-Aldrich
Butane	106-97-8	Camping Gaz
Chloroform	288306	Sigma-Aldrich
Cy3-UTP	PA53026	GE Healthcare
Cy5-UTP	PA55026	GE Healthcare
DL-Dithiothreitol (DTT)	R0861	Thermo Scientific
dNTP set, Sodium salt	M3015.4100	Genaxxon
Ethanol, absolute	24102	Sigma-Aldrich
Ethidium bromide	M3178.0010	Genaxxon
Ethylenediaminetetraacetic acid	E9884	Sigma-Aldrich
GeneRuler DNA marker	SM 0311	Thermo Scientific
Glutathione agarose	16100	Thermo Scientific
Glycerol	G5516	Sigma-Aldrich
Guanidine Hydrochloride	6009.2	Carl Roth
Halt™ Protease and Phosphatase Inhibitor	78443	Thermo Scientific
Hellmanex II	320.001	Hellma GmbH
Heparin sodium salt from Porcine Intestinal Mucosa	H4784	Sigma-Aldrich
HisPur™ Ni-NTA Resin	88221	Thermo Scientific
Hydrochloric acid (HCl), (37%)	85848.290	VWR International

Imidazole	3899.2	Carl Roth
Isopropanol	W292907	Sigma-Aldrich
Isopropyl- β -D thiogalactopyranoside	M3198.1001	Genaxxon
Kanamycine	K4378	Sigma-Aldrich
L-Glutamic acid potassium saltmonohydrate	G1501	Sigma-Aldrich
LightCycler® 480 Multiwell Plate 384, white	04729749001	Roche
Magnesium acetate tetrahydrate	M0631	Sigma-Aldrich
Methanol	322415-2L	Sigma-Aldrich
Millex-GP, 0.22 µm filter	SLGP033RS	EMD Millipore
MultiScreen $PCR_{\mu 96}$ Plate	LSKM PCR 50	EMD-Millipore
N,N,N′ ,N′ - Tetramethylethylenediamine (TEMED)	T9281	Sigma-Aldrich
NEXTERION® Slide E	1066643	SCHOTT
Nickel Chelate slides	VNC 000 10	Xenopore
Nitrocellulose slides	SMN	Arrayit corporation
Nuclease-Free Water	AM9939	Life Technologies
PCR tubes,8 strip 0,2ml	AM12230	Life Technologies
Peptone from Casein	70171	Sigma-Aldrich
Phenylmethylsulfonyl fluoride	36978	Thermo Scientific
Polystyrene cell culture tubes	187261	Greiner bio-one
Ponceau S solution	33427.01	Serva
Potassium chloride (KCl)	P5405	Sigma-Aldrich
Q5 High-Fidelity DNA polymerase	M0491L	New England Biolabs
RNaseOUT™ RibonucleaseInhibitor	10777-019	Invitrogen
Sodium Chloride (NaCl)	S9888	Sigma-Aldrich
Sodium dodecyl sulfate (SDS)	71725	Sigma-Aldrich
Sodium phosphate dibasic(Na_2HPO4)	S9763	Sigma-Aldrich
Sodium phosphate monobasic(NaH_2PO4)	S9638	Sigma-Aldrich
SpectraMulticolor Broad range protein ladder	26623	Thermo Scientific
Streptavidin-coated slides	MPC 000 10	Xenopore
SYPRO® Ruby Protein Gel Stain	S4942	Sigma-Aldrich
Syringes50 ml	SS03L2138	Terumo
Tergitol® solution	NP40S	Sigma-Aldrich
Triton X-100	T8787	Sigma-Aldrich

Trizma® Base	T1503	Sigma-Aldrich
Trizma® HCl	T3253	Sigma-Aldrich
TRIzol Reagent	15596-018	Invitrogen
tRNA from *E.coli*	10109541001	Roche Diagnostics
Tween® 20	P2287	Sigma-Aldrich
Yeast Extract Biotechnology Grade	1830,1000	Gerbu
Zinc chloride ($ZnCl_2$)	208086	Sigma-Aldrich
β -mercaptoethanol	M6250	Sigma-Aldrich

Labware

Product	**Catalogue N°**	**Manufacturer**
Adhesive PCR Seal	600208	Biozyme
Amicon Ultra-15 Centrifugal Filter Units	UFC900308	EMD Millipore
Cell culture 96 well microplate,U-bottom	650160	Greiner Bio-One
Centrifugal Ultrafiltration device	VC1002	Sartorius
Conical bottom falcon tubes(15 ml/ 50ml)	227261/210261	Greiner Bio-One
Eppendorf Safe-Lock tubes (0.5 ml, 1.5 ml and 2 ml)	0030121694/ 0030121597/ 0030121570	Eppendorf
Gloves, Latex Medical Examination	BM11228-PF-AV	Blossom
Gloves, Nitril Freeform SE	FFS-700	Microflex
Hydrogel slides	SLDF	Arrayit corporation
Nunc 96 well plates	M9410	Sigma-Aldrich
Nunc-Lab-Tek petri dishes	4021	Thermo Scientific
Parafilm® M	P8505	Labomoderne
Thin wall standard 96-well PCR plate	82-0600-A	Peqlab
Whatman® UNIPLATEmicroplates 384 well, V-bottom	7701-5101	GE Healthcare Life Science

Equipment

Name	**Manufacturer**
Arrayscan 4000XL	Pekin Elmer
Biofuge, Pico	Hereaus Instruments

Electronic Balances, Kern 434,440–45	Kern & Sohn GmbH
Heating Block	Grant Instrument
Hoefer TE 70 Semidry Transfer Unit	GE Healthcare
Ice maker	Scotsman
Laminar Hood	Hereaus instruments
Microcomputer electrophoresis power supply	Renner GmbH
Microwave oven	Bosch
Mixmate Microplate Mixer	Eppendorf
Nanodrop Spectrophotometer N1000	Peqlab
Nanoplotter 2	GeSIM
Orbital Shaker	Edmund Bühler
Power Scanner	Tecan
Refrigerated centrifuge 2k15	Sigma
Rocky shaker	Fröbel Instruments
RS-TR05	Phoenix Instruments
Centrifuge R5810	Eppendorf
Thermocycler	MJ Research, LifeEco
TKA MilliQ watter supply	Millipore
Ultraspec 2000 Spectrophotometer	Pharmacia Biotech
Ventillated Oven	Kendro Instruments
Vortex Mixer	Neolab
Water bath	Grant Instruments

Buffers and media

Name	Composition
Binding buffer 5×	25 mM Hepes pH 7.9, 50 mM Glutamte potassium salt, 1 mM DTT, 8 mM Magnessium acetate, 10% v/v glycerol
Ethidium bromide solution	0.5 µg/ml Final concentration
Laemmlibuffer	30.1 g Tris Base, 144.2 g Glycine, 50 ml SDS (20%), add 1 l dH2O
LB-Agar	LB-Medium + 1.5% (w/v) Agar
LB-Medium (1 liter)	10 g Tryptone/Pepton, 5 g yeast extract, 5 g NaCl, pH 7.2
PBS 10× (1 liter)	80 g NaCl, 2 g KCl, 26.8 g Na_2HPO_4, 2.4 g KH_2PO_4, pH 7.4

TBE 10× (1 liter)	108 gTris, 55 g Boric acid, 40 ml 0.5 M Na_2EDTA, pH 8
TBS 10× (1 liter)	50 mM Tris, 150 mM NaCl with HCl, pH 7.5
TBST	1× TBS/0.05% Tween-20
Transfer buffer	150 mM Glycine, 25 mM Tris-base, 20% Ethanol
PBST 1× (1 liter)	1× PBS/0.05% Tween-20
1M HEPES-KOH pH 7.5	Dissolve 238.30 g HEPES in 1 l H_2O. Use KOH Stock solution to adjust pH to 7.5
0.5 M EDTA pH 8.0	Dissolve 186.1 g $Na_2EDTA.2H_2O$ in 800 ml diH_2O. Adjust pH to 8.0 with NaOH (~20 g of NaOH pellets). EDTA will dissolve at pH 8.0. Adjust volume to 1 liter with diH_2O. Sterilize by autoclaving and store at room temperature.
1 M Tris-HCl pH 6-8	12.1 g tris base in 100 ml H_2O, adjust pH with concentrated HCl
10 M NaOH	40 g NaOH in 100 ml H_2O
1 M NaH_2PO_4 (monobasic)	138 g $NaH_2PO_4.H_2O$ in sufficient H_2O to make a final volume of 1 liter
1M Na_2HPO_4 (dibasic)	142 g of Na_2HPO_4 in sufficient H_2O to make final volume 1 liter
6× Gel loading buffer	25 mg bromophenol blue and 4 g sucrose, make up volume to 10 ml with diH_2O, store at 4°C
6× Gel loading buffer with Ficoll	25 mg bromophenol blue, 25 mg xylene cyanol FF and 1.5 g Ficoll (Type 400;Pharmacia) in 10 ml diH_2O, store at 4°C
Gluthathione free acid reduced	Sigma-Aldrich Chemie GmbH, Germany

2.2 Methods

2.2.1 Template Generation for Protein Microarray Production

Template DNA for PCR-based generation of protein expression constructs was obtained from a cDNA library cloned in lambda zap CMV. The library was constructed from procyclic mRNA of *Trypanosoma brucei* and directionally cloned into the phagemid with a complexity of 150,000 pfu/ml. Individual clones were obtained by plaque lifting as described in the manufacturers. Briefly, XL1bleu E.coli cells were grown to an optical density of 0.6, and then inoculated with about 1.5×10^2 pfu from the phage library. The bacteria-phage mixture was then incubated at 37°C for 30 minutes for phage attachment. After attachment, 1.8ml of top agar containing 10mM $MgSO_4$ and 0.2% Maltose maintained at 48°C was immediately added to the phage-bacteria mixture and then poured onto NZY agar plates in 150×25mm polystyrene culture dishes (Becton Dicson). The plates were allowed at room temperature for 30 minutes, then inverted and incubated overnight at 37°C. Plaques were chilled at 4°C for 1 hour and lifted using sterile toothpicks. Individual plaques were lifted into sterile U-shaped 96 well plates (Cellstar) containing 200 µl of milliQ water. The plates were sealed and heated at 80°C for 15 Minutes in a ventilated oven. Clone individuality was checked by PCR and the heat-killed bacteriophage was then used for the generation of cell-free expression templates. For Open Reading Frames (ORF), primers were designed and synthesized (Biomers) to amplify specific genes directly from gDNA. All gene-specific primers carried 5′ and 3′ overhangs which were later used as priming sites for the generation of expression constructs.

All expression primers were similar in their basic structure, but their complementary sequences varied depending on template source (**Index 1**). The basic primer consisted of an upstream spacer, T7 promoter internal spacer, a ribosome binding site/Kozak sequence and a downstream spacer. For ORFs, the downstream spacer was followed by a start codon, a 6×His epitope and the complementary region. Meanwhile, for the cDNA products, the downstream spacer was directly followed by the complementary sequence. PCRs were carried out in V-bottom 384 well clear PCR plates in a total volume of 25 µl containing 1× Q5 reaction buffer, 1× Q5 high GC enhancer, 0.25 µM forward and reverse primers, 0.2 mM dNTP, 5 µl of heat-killed

bacteria plaques and 2.5 U of Q5 high fidelity DNA polymerase. Cycling was done following the manufacturer's instructions and 5 µl of amplicons was randomly checked on 1.2% agarose gels.

2.2.2 *In Situ* Cell-Free Protein Expression

The Multiple spotting technique was used for on-chip protein expression, and several cell-free expression systems were tested including: the T7 TNT quick for PCR DNA (Promega), the EasyExpress insect kit II (Qiagen), PureXpress In vitro protein synthesis kit (New England Biolabs), and the S30 T7 High yield protein expression system (Promega). Expression constructs were mixed with 0.5 M betaine (final concentration) in V-bottom 384 Uniplate (whatman). The plate was briefly spun down and mixed on a micromate mixer for 30 seconds at 1600 rpm. Approximately 0.9 nl of template was spotted using a non-contact Nanoplotter 2 (GeSIM) onto the glass surface followed by 3.6 nl of the cell-free expression mixture with or without transcend-biotin-lysyl tRNA (in the case of streptavidin slides). The slides were then incubated in deep humid hybridization chamber for protein expression. Many expression temperatures and incubation periods were tested on diverse surface coatings. Therefore, spotting was either done on epoxysilane slides (Nexterion slide E), Ni-NTA coated slides (Millipore), Streptavidin-coated slides (Millipore) or Nitrocellulose slides (Arrayit corporation), and incubated in a ventilated oven: at room temperature, at 37°C for two hours, at 37°C overnight, at 37°C for two hours plus 30°C overnight and at 37°C for one hour and 30°C overnight oven. Slides were removed from the incubation chambers and stored at –20°C for at least 24 hours before use.

2.2.3 Detection of Expressed Proteins

Two strategies were adopted for the detection of expressed proteins, depending on whether or not the expressed proteins were fused to epitope tags. For proteins expressed from ORFs, Alexa Fluor 647-conjugated monoclonal anti 6xHis was used for detection. Therefore, slides were mounted on slide frames and blocked at room temperature for one hour in blocking buffer containing 2% BSA in PBST on an orbital shaker. The slides were washed three times 5 minutes each in PBST before incubation with a 1:1000 dilution of the antibody in blocking buffer. Incubation was done at room temperature for one hour on an orbital shaker and the slides were washed in PBST three

times 5 minutes each. Finally, the slides were rinsed in milliQ water, air-dried at 30°C for 30 minutes and scanned in a Tecan power scanner. Images were saved as tagged image files (TIF) and analyzed with GenePix Pro.6.0. Meanwhile, proteins expressed from cDNA products were stained with SyproRuby protein gel stain. For this purpose, proteins were fixed on slides in a solution of 50% methanol and 7% acetic acid at room temperature for 30 minutes. Then, the slides were incubated with SyproRuby protein gel stain for 90 minutes at room temperature on an orbital shaker. Afterwards, slides were washed in a solution of 10% methanol and 7% acetic acid for 30 minutes at room temperature, rinsed in milliQ water, air-dried at 30°C for 30 minutes and scanned on a ScanArray 5000 microarray scanner. Images were saved as TIF files and analyzed as above.

2.2.4 Test of Functionality and Protein Quantification

Since the main aim of this study was the production of functional array, we equally evaluated the functionality of protein expressed on the platform. Therefore, the Green Fluorescent Protein (GFP) was used. Purified recombinant GFP was spotted in serial dilution, and on the same slide, GFP was expressed by the multiple spotting technique. The slides were incubated together for the expression of GFP. One slide was scanned directly for the auto fluorescence of GFP while the other was treated with 6 M guanidine-hydrochloride before scanning. Moreover, the trypanosome trypanothione reductase was expressed in the S30 T7 high yield system and assayed for its activity using Trypanothione (TS_2) and NADPH

2.2.5 Expression and Purification of Recombinant Bait Proteins and Enzymes

The DNA sequences of the Bait proteins and enzymes were amplified from gDNA and cloned into the Champion pET 101 directional Topo expression vector (Invitrogen) or in pDEST 15 following the manufacturer's protocol. Briefly, PCR products were purified using Purelink PCR purification kit (Life technologies), and quantified using Nanodrop. About 75 ng of purified PCR products were mixed with 15 ng of pET plasmid in the presence of 0.6 mM NaCl and 0.03 mM $MgCl_2$. The reaction was allowed at room temperature for 15 minutes and 3 µl of the cloning reaction was used to transform Top10 chemically competent cells and later plated onto LB ampicillin

(100 µg/ml) agar plates. Single colonies were picked from the agar plates and grown in LB medium containing 100 µg/ml Ampicillin for 8 hours and plasmid DNA was isolated using the Qiaprep plasmid isolation kit (Qiagen) following the manufacturers protocol. The presence of the insert was confirmed by PCR using vector and insert-specific primers. Positive clones were sequenced and used for transformation of BL21 star *E.coli* cells. Cells were transformed with 50 ng of the plasmid and grown in 10ml LB medium containing 100 µg/ml of ampicillin. The culture was grown until an optical density of 0.6 and test protein expression was initiated by the addition of 250 mg/l of IPTG. After 6 hours of culture, the cells were harvested by centrifugation at 4000 rpm for 30 minutes, proteins extracted and a western Blot was performed.

For pilot expression, a 10ml of overnight culture was transferred into a 1liter culture flask containing 2YT medium. The culture was grown until an optical density of 0.6 and pilot protein expression was initiated by the addition of 250 mg/l of IPTG. After induction, the cells were grown for further 4–16 hours and harvested by centrifugation at 5000*g* for 15 minutes in an RC-5C super speed centrifuge (Sorvall). The pellet was washed in ice-cold PBS (pH 7.5). Proteins were extracted using the B-PER reagent supplemented with lysozyme (1mg/ml) according to manufacturer's protocol and recombinant proteins purified using Ni-NTA agarose (Invitrogen). The concentration of the purified protein was measured using the BCA protein assay Kit (Pierce) and stored in HEPES buffer (pH 7.5), containing 150 mM NaCl, and 25% glycerol.

2.2.5 RNA preparation

The genes coding for candidate RNA were amplified by PCR and cloned downstream of the T7 promoter. The resulting linear construct was used as template for the synthesis of RNA using the Transcript Aid T7 high yield in vitro transcription kit (Fermentas) according to the manufactures protocol. For the synthesis of fluorescently-labeled RNA probes, this protocol was modified by using a ratio of 1:3 of UTP and Cy3-UTP (GE Healthcare) respectively. The reaction was incubated at 37°C for 6–8 hours, and RNA was purified by sodium acetate precipitation, re-suspended in Nuclease-free water containing 20mM EDTA and stored at −80°C. Biotinylated probes were synthesized in a similar manner, but without Cy-Dye UTP and labelled using the Pierce™ RNA biotinylation/destiobiotinylation kit following manufacturer's instructions. Therefore, 25% DMSO was added to the synthesized RNA and heated at 90°C for 5 minutes on a heating block and immediately placed on ice. The ligation

reaction, containing 1× RNA ligation reaction buffer, 40U of RNase inhibitor, and 40U of T4 ligase, 1 nmol biotinylated cytidine (bis) phosphate and 15% polyethylene glycol in a total volume of 30 µl was prepared. The denatured RNA was added to the reaction mixture and incubated at 16°C for 8 hours. After ligation, the reaction volume was completed to 100 µl, and labelled RNA was purified by isoamyl acohol:chlorofoam extraction. RNA oligos that could not be synthesized due to size constraints were commercially synthesized and labelled as desired.

2.2.6 Protein-Protein Interactions (PPI)

Protein microarrays were thawed directly in blocking buffer (20 mM Tris-HCl, pH 7.9, 0.1% Triton X-100, 0.2 mM PMSF,5 mM $CaCl_2$, 5 mM $MgCl_2$, 100 mM NaCl 1mM DTT, 1× Halt protease and phosphatase inhibitor cocktail, 20% glycerol and 1% BSA) and blocked for one hour at room temperature on an orbital shaker. The arrays were washed three times 5 minutes each in wash buffer (20 mM Tris-HCl, pH 7.9, 0.1% Triton X-100, 0.2 mM PMSF, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, 1× Halt protease and phosphatase inhibitors). Bait proteins were fluorescently labeled with Dylight 532/647 and purified on Zeba spin columns. Bait proteins were diluted as appropriate (25 µg/ml) in blocking buffer and incubated on the arrays overnight at 4°C. Afterwards, arrays were washed again three times 10 minutes each, and then rinsed in milliQ water. All arrays were air-dried and scanned on a Tecan scanner. Data was extracted using GenePix 6.0 Pro and analyzed with the Limma package in R.

2.2.7 Protein-RNA Interactions (PRI)

For PRIs, protein microarrays were screened with either 0.1 µM concentrations of short oligomers, or 2 µg for longer RNA species such as 3′ UTRs, polyA RNA and total RNA. The arrays were thawed in Blocking buffer (25 mM Hepes-KOH, 50 mM Glutamic acid potassium salt, 8 mM magnesium acetate, 100 µM zinc chloride, 20 µg/ml heparin, 50 µg/ml *E. coli* tRNA, 0.1% Triton X-100, 40U RNaseOUT RNase inhibitor, 1× Halt protease and Phosphatase inhibitor cocktail and 1% BSA) and blocked for 1 hour at room temperature with gentle shaking (50 rpm). The arrays were washed in wash buffer (25 mM Hepes-KOH, 50 mM Glutamic acid potassium salt, 8mM Magnesium acetate, 100 µM Zinc chloride, 20 µg/ml heparin, 50 µg/ml *E. coli* tRNA, 0.1% Triton X-100, 40U RNaseOUT RNase inhibitor, 1× Halt protease and

Phosphatase inhibitor cocktail) three times 5 minutes each, and then probed with labelled RNA probed in Binding buffer (blocking buffer containing 0.4 mg/ml *E. coli* tRNA and 40 µg/ml heparin) overnight at 4°C without shaking. Arrays were again washed three times 15 minutes with gentle shaking at room temperature, air-dried and scanned as above.

2.2.8 Minimal Sequence Mapping

Selected fragments of the 3′-UTR of 6 genes were analyzed on protein microarrays. To investigate if the observed binding was sequence-specific, minimal sequence mapping was performed. Therefore, the trypanosome hexose transporter gene 3′-UTR was used. Fragments were generated via PCR to have 100nt difference in length, and subsequently fused to a T7 transcription construct. The fragments were transcribed as above and used to screen the arrays in triplicates as before. *In silico* folding of the fragments was performed using RNAfold web server[126]. A stem loop conserved across all fragments was commercially synthesized and biotin-labeled. This stem loop was then used to rescreen the arrays as described previously

2.2.9 Kinase Substrate Identification

Substrates of the *T. brucei* Aurora Kinase B (Centrin II and Histone H3) were expressed using the MIST technique and arrays were used for protein kinase assay. To this end, slides were blocked in superblock supplemented with 0.1% Triton X-100 for one hour at room temperature on an orbital shaker. The arrays were then washed twice in Kinase buffer (25mM Hepes-KOH pH 7.9, 150mM NaCl, 10mM $MgCl_2$, 0.1% Triton X-100,1× Halt protease and phosphatase inhibitor cocktail and 25% glycerol) for 5 minutes at room temperature. The kinase reaction containing 50 µg of Aurora kinase B from *Trypanosoma brucei*, and 33 mM ATP in kinase buffer was incubated at 30°C for 2 hours, then washed in kinase buffer 3 times 5 minutes. A 1:100 dilution of mouse anti phosphoaminosereine/threonine/tyrosine antibody (SPM 101) was made in kinase buffer containing 1% BSA and incubated on the slides for 90 minutes. The arrays were again washed as previously, and a 1:5000 dilution of Cy5-conjugated anti mouse secondary antibody was incubated on the slides for one hour. Finally, the slides were washed three times 10 minutes and then rinsed briefly in milliQ water, air-dried and scanned.

2.2.10 Antibody Selection

To check if specific proteins could be expressed from cDNA products, arrays made from cDNA products were screened with a pool of *T. brucei* protein-specific antibodies. In this instance, arrays were blocked in 1% BSA in PBST at room temperature for 1 hour, then washed in PBST three times and incubated with 1:1000 dilution of the antibody pool. The slides were washed again, and then incubated with a 1:5000 dilution of fluorescently labelled secondary antibody for one hour at room temperature. The arrays were washed as above; air-dried and read out using a Tecan power scanner. Images were stored as Tiff files and analyzed using GenePix pro 6.0.

2.2.11 Microscale Thermophoresis

To evaluate the binding affinity between some protein-RNA and protein-protein pairs, microscale thermophoretic measurements were made. To this end, RNA oligonucleotide probes were fluorescently labeled at the 3′-end with Cy5. Proteins of interest were purified and quantified by using the BCA protein assay kit (Pierce). Dilutions of the RNA probe were made to cover a range between the concentrations used on the microarrays ± 2 fold. These dilutions were used for a capillary scan to determine the appropriate dilution for thermophoresis. Sixteen serial dilutions of the protein were made, starting with undiluted sample. An equal volume of the RNA probe was added into each serial dilution and filled into standard capillaries. A capillary scan was performed to evaluate the fluorescence distribution and the binding reaction was monitored for all capillaries using 20% and 40% MST and 90% LED power. Date was analyzed using the NTAnalysis software and exported into excel sheet

2.2.12 Cell Culture

Blood stream form cells were used for all experiments, and cells were grown and maintained in HMI medium supplemented with 10% fetal bovine serum, 0.2µg/ml of phoemycine, 1.3 mM of L-cystein in ß-mercaptoethanol at 37°C under 5% CO_2 conditions. Bloodstream*1312514* cell were used for the RNAi and *in situ* tagging experiments and grown in hygromycine or Blasticidin-containing medium, for the RNAi plasmid and for epitope tagging plasmid respectively. Meanwhile bloodstream form *13132277* and *13131991* cells were used for the CAT assay and were grown in hygromycine-containing medium (5µg/ml).

2.2.12.1 Candidate Gene Knock Down

Double stranded RNA interference was used to knock down some genes that showed significant binding to RNA. For this, 400 bp fragments of each gene starting from the start codon were amplified by PCR using Taq DNA polymerase. The resulting amplicon was purified using the PureLink PCR purification kit and quantified by Nanodrop. These fragments were clones into the p2T7 plasmid after digestion with Eam1105I restriction enzyme to generate TA cloning sites. The plasmid carries two T7 polymerase promoters each of which is under the control of a bacterial tet repressor and a hygromycine resistance gene as a selection marker. After successful cloning, plasmids were digested with Not1, and precipitated with sodium acetate. Afterwards, 5 µg of digested plasmid was transfected into the parasites. Parasites were allowed to grow overnight, and the selection marker was added and cells plated into 24 well plates in serial dilutions and incubated for five day. Three positive clones from the most diluted wells were transferred into hygromycine-containing medium and allowed to grow. Cells were then diluted and reseeded in 10 ml cultures in duplicates, one of which was induced with 0.75 µl of tetracycline (5mg/ml) and the other served as a control. Cells were counted for six successive days during which they were counted before and after dilution.

2.2.12.2 Epitope Tagging and RNA Pull-Down

Genes were endogenously tagged using a one-step PCR-based strategy as previously described[127] with some modifications. To this end, pair of 90-nt fragments of each ORF starting with the start codon downstream and the 5' UTR upstream was commercially synthesized. These fragments were made to carry overhangs that permit them to anneal to the blasticidine-V5 cassette of Bla-V5 vector. The resulting construct generated therefore contains 90-nt immediately upstream of the start codon of the gene, followed by the bla-resistance marker, an intergenic region, a V5 epitope and the first 90-nt of the ORF. The entire construct was used to transfect parasites which were later selected in blasticidine-containing medium. Positive clones were isolated and allowed to grow in 50 ml cultures. 7.5×10^6 cells were lysed in lamini buffer and used for western blot to check for correct epitope tagging. To pull-down the tagged proteins and associated RNA, 50µl (per sample) of agarose immobilized anti-V5 antibody was washed five times in in 1× PBS, followed by equilibration in immunoprecipitation buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% NP40 and 1× Halt™ Protease

and phosphatase cocktail) .Approxiamtely 1×10^9celles were centrifuged at 2000 rpm for 10 minutes, and the cell pellet was washed in PBS. The pellet was the transferred into 1.5 ml Eppendorf tubes and kept on ice where 100 µl of lysis buffer (20 mM Tris-HCl, 10 mM NaCl, 0.1% NP40 and 1× Halt™ Protease and phosphatase inhibitors) was added and cells passed through a 21-gauge needle 10 times. Cell debris was removed by centrifugation for 15 minutes at 15k at 4°C and the supernatant was transferred into a new tube. As control, wild type cells, and cells expressing c-terminal c-*myc* were used. The resulting protein extract was divided into two fractions. One was used for input RNA extraction and the other for immunoprecipitation.

For immune precipitation, cell lysate was incubated at 4°C with the immobilized antibody for 5 hours with gentle rotation. The beads were washed five times in IP buffer and RNA was extacted from the beads using Trizol and precipitated in sodium acetate. One fraction of the beads was boiled in lamini buffer and analysed on a 4–12% gradient SDS gel. One microgram of the input and immunoprecipitated RNA were used for cDNA synthesis using Protoscript cDNA synthesis kit. The resulting cDNA was used as template to amplify fragments of the 3′ UTR of genes that were previously analyzed on protein microarrays. Amplicons were resolved on 2% agarose gels and stained with ethidium bromide.

2.2.12.3 Immunofluorescence

Following successful *in situ* tagging of candidate protein, their cellular localization was verified by immunofluorescence microscopy. For this, cells were grown to mid-log phase and 2.5×10^9 cells were harvested by centrifugation as above and washed in 0.5 ml PBS. A 0.01% solution of polylysine was prepared in water and used to coat 4-chamber cell/tissue culture slides. The slides were incubated at room temperature for 15 minutes without shaking and washed 4 times with 1ml of water. The cell pellet was resuspended in 25 µl of PBS and 0.5 ml of freshly prepared 4% paraformaldehyde in PBS was added to the cells and incubated at room temperature for 18 minutes. Cells were collected by centrifugation at 6000 rpm for 2 minutes, and washed twice in 0.5ml PBS. Fixed cells were resuspended in PBS and dispensed onto the polylysine modified slides and incubated at 4°C overnight. Unattached cells were removed by suction, and the slides were incubated at room temperature with 0.2% Triton X-100 gentle shaking for 20 minutes for permeabilization. Excess Triton was removed by washing with PBS three times five minutes. Slides were blocked with 0.5%

gelatin in PBS for 20 minutes, and then incubated with a 1:200 dilution of mouse Anti-V5 antibody for 60 minutes at room temperature. The slides were washed twice with PBS, and a Cy3/5-conjugated anti-mouse antibody in a 1:500 dilution was added including 100 ng/ml of DAPI nuclear staining dye and incubated at room temperature for 60 minutes with gentle shaking. Slides were washed with excess PBS, air-dried, mounting medium added and sealed with cover slips using nail polish. Cells were observed under 100× oil objective on a CellR Olympus camera and images taken.

2.2.12.4 Pair-Wise Yeast Two Hybrids

Preliminary validation of protein interaction by yeast two-Hybrid screening was performed (by the Genomic and proteomic core facility at the DKFZ). Therefore, genes coding for selected candidate were amplified from gDNA by PCR and cloned into bait (pGKBT7) and prey (pGADT7) vectors between EcoRI and BamHI. A Single colony of each was allowed grown over night in 2 ml selective medium at 30 °C and 200 rpm. Thereafter, a volume of 500 µl of bait and prey were mixed, spun down at 2800 rpm for 5 min at room temperature. Two milliliters ofmYPDA/PEG mating medium was then added. Cells were thoroughly mixed in mating medium and mating was achieved by incubating them at 30 °C and 100 rpm for 3 h. Cells were harvested at 2800 rpm for 3 min, then washed with 5 ml of ΔLW medium and resuspended in 1.5 ml ΔLW. Diploids were grown for 2 days in liquid ΔLW at 30 °C without shaking and then plated on selective agar plates (ΔLWH). In the case where the bait and prey protein interacted, visible colonies could be seen 2 days later. Stringency was adjusted by adding an appropriate amount of 3-AT to the selective agar plates.

3 Results and Discussion

3.1 Protein Expression Microarrays

3.1.1 Template Construction

Genomic DNA and cDNA products were used for the construction of protein expression templates. The use of gDNA was possible since trypanosomes, unlike other eukaryotes lack intron and open reading frames can directly be amplified. Primers were designed for the amplification of 384 individual genes. About 97% of the target genes were amplified successfully with the right size, most of which gave single amplicons (**Fig. 2a**). Expression constructs were generated in a second PCR step using a primer pair carrying sequences complementary to the overhangs of the gene-specific primers. In most cases, such a construct allowed for protein expression. However, constructs containing a start codon 6 nucleotides after the Ribosome Binding Site (RBS) followed by an N-terminal 6×His tag led to enhanced expression. A start codon alone downstream of the RBS preceding the primer annealing site was less performant, but did better than constructs without supplementary start codons (**Fig. 2c**) The basic expression construct consisted of a 6 nucleotide upstream spacer, a T7 RNA polymerase promoter followed by a spacer, a ribosome binding site and another spacer. Omission or replacement of the upstream spacer significantly decreased expression by more than 50%. The reverse primer in some cases contained a V5 or c-myc epitope for detection of the full-length protein. The c-myc epitope constructs were seen to yield less protein expression than the C-terminal V5 epitopes constructs.

A phage cDNA library (provided by Prof. Isabel Roditi University of Bern) was used to transform bacteria with an efficiency of about 150 pfu/µl. Individual plaques were lifted, and clone uniqueness could be confirmed by PCR. A total of about 14,000 plaques were lifted, of which more than 75% were PCR positive containing single inserts. Expression constructs generated from cDNA products were closed at both ends, and protein expression relied exclusively on gene-specific translation start from full-length cDNA. Two primer pairs were designed to generate expression constructs from cDNA products. One primer pair annealed directly next to the cloning sites while the second annealed about 50-nt further. Both primers allowed for the amplification of the inserts. The first pair had an extremely high melting temperature (68°C), requiring the

use of PCR additives and yielded highly specific products (**Fig. 2b**). The second pair annealed at a relatively low temperature, but was less specific then the first. Besides, it was impossible to achieve significant protein expression using this primer (**Fig. 2d**). In both cases, the size distribution of the amplified constructs was analyzed on 1.2% agarose gel. A very significant number of the constructs were about 1 kb in length.

Templates generated from gDNA were expressed and full-length proteins could be detected using C-terminal epitope tags. Constructs generated from cDNA products using primer pair I equally allowed for protein expression. It was impossible to say if the correct or even full-length proteins were expressed. Protein-specific antibodies however reacted with some of the expressed proteins (**Fig. 2e**). The corresponding constructs were sequenced and verified.

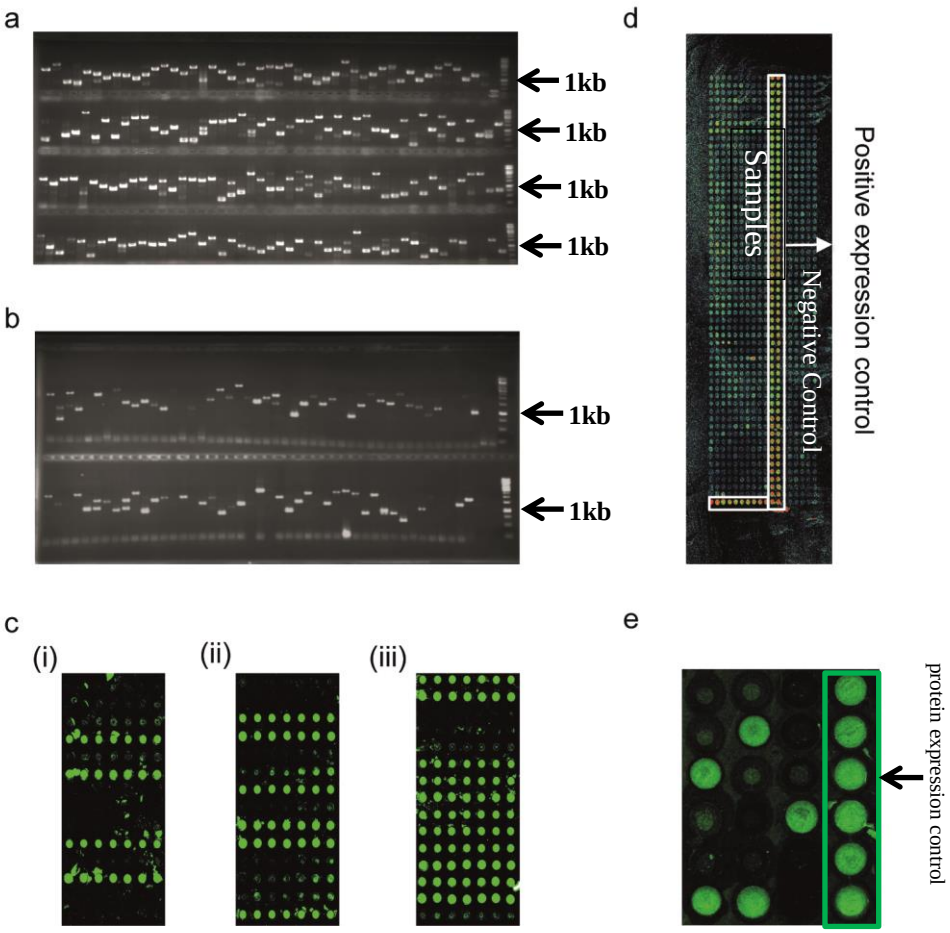


Figure 2: DNA template construction for full-length expressed protein on the array. (**a**) PCR products from gDNA (**b**) PCR products from cDNA library (**c**) Full length protein stained with antibody to c-terminal V5 tag. DNA template constructions with no supplementary ATG (i), with supplementary ATG alone (ii), with supplementary ATG followed by 6xHis (iii). (**d**) Slide stained with SyproRuby following expression. Primer used for expression has a 56-nt distance from RBS (**e**) Protein-specific antibody detection of proteins with antibodies to tubulin, treparedoxin, trypanothione reductase and XRND.

The distance between the RBS and the start codon is crucial for successful protein expression[128]. The less expression observed for the second primer par could actually be attributed to the distance between the putative start codon and the RBS. This is compounded by the fact, that *trans*-splicing in trypanosomes add a 39 nucleotide splice leader sequence to the 5′end of mRNA transcripts which is meant to play the role of the cap[129]. The enhanced expression observed for constructs carrying N-terminal 6xHis is indeed interesting. This epitope is often used for affinity purification and immobilization and can help insure protein availability which is in most cases problematic. It is possible, that a 6× repeat of histidine offers an easy translation start, thereby favoring the expression of the host template. This is however a simple speculation, and the exact cause-effect relationship still needs to be established. Expression construct generation from gDNA may seem to be an attractive option that circumvents gene synthesis. It offers template flexibility, allowing the user to decide which portion of the open reading frame is to be expressed.

Furthermore, when there are undesired sequences such as signal peptide, it is possible to eliminate such sequences. This is however only valid for organisms lacking introns and only possible for well annotated genes. Besides, generating large primer sets with similar melting temperature could prove to be challenging. cDNA products on the other hand have the major disadvantage of being less flexible as a result of upstream sequences and a stop codon. More so, one can expect only about a third of the construct to xpress the correct protein. There is always the need for protein-specific antibodies or complementary methods for detection of the expressed protein. They are however cheap template source, and most importantly, can allow for the study of rare or unknown splice variants especially in disease states. They have for long been used as template source for protein expression[130].It is therefore possible to generate protein arrays from both cDNA products and open reading frames derived from diverse sources. A minimal distance must be kept between the RBS and the initiation codon to ensure proper and optimal expression. A sequence upstream of the T7 RNA polymerase promoter is very essential to avoid stalling of the transcription complex.

3.1.2 On-Chip Cell-Free Protein Expression

The multiple spotting technique was used to produce protein microarrays from PCR-generated templates. The performance of several cell-free expression systems, including the TNT T7 for PCR (Promega), the Easy express insect Kit II (Qiagen), the

PurExpress (New Englan Biolabs) and the S30 T7 High yield protein expression system (Promega) was evaluated. PurExpress, TNT T7 for PCR and the EasyXpress insect kit II were unsuitable for the application. The all yielded relatively low protein expression levels, sometimes beyond detection limit and reacted with detection antibodies. Additionally, the TNT T7 was very viscous and could not easily be spotted easily. The insect kit was uncoupled and required an additional spotting step. All these led to the discontinuation of these expression systems, and the S30 T7 high yield from Promega was continued.

As previously reported[131], there was no need for expression template cleanup, and purified PCR products yielded less protein than their corresponding unpurified counterparts (**Fig. 3**). There was an moderate relationship between gene size and expression level and different proteins were expressed at different levels with larger proteins showing better expression than smaller ones (R^2<0,6).

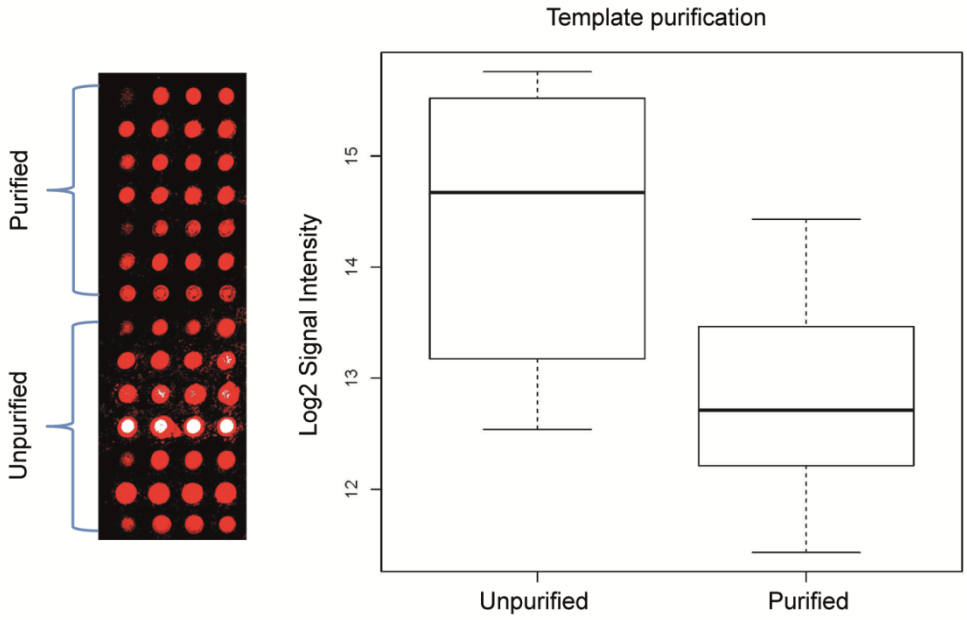


Figure 3: DNA template preparation for protein microarrays. Seven *T brucei* proteins (from top row, U70k, U24k, U1C, U1A, TFIIA, TRF4 and SNAPc) expressed on single slide using unpurified PCR and purified PCR products. Each protein was spotted across the row. Array stained with SyproRuby for expression ***(left)*** and quantification boxplot ***(right).***

Nikel coated (Xenopore), Epoxysilane (SCHOTT Nexterion), Streptavidin-coated (Xenopore) and Nitrocelluloseslides and hydrogel (Arrayit Corporation) were tested. Protein expression was detectable on all except the nitrocellulose and hydrogel slides. Nickel chelate slides were not used for the immobilization of proteins expressed

from cDNA construct. When used, they produced well-rounded spot morphologies and in some cases tail-like features could be seen attached to the spots (**Fig. 4a**).

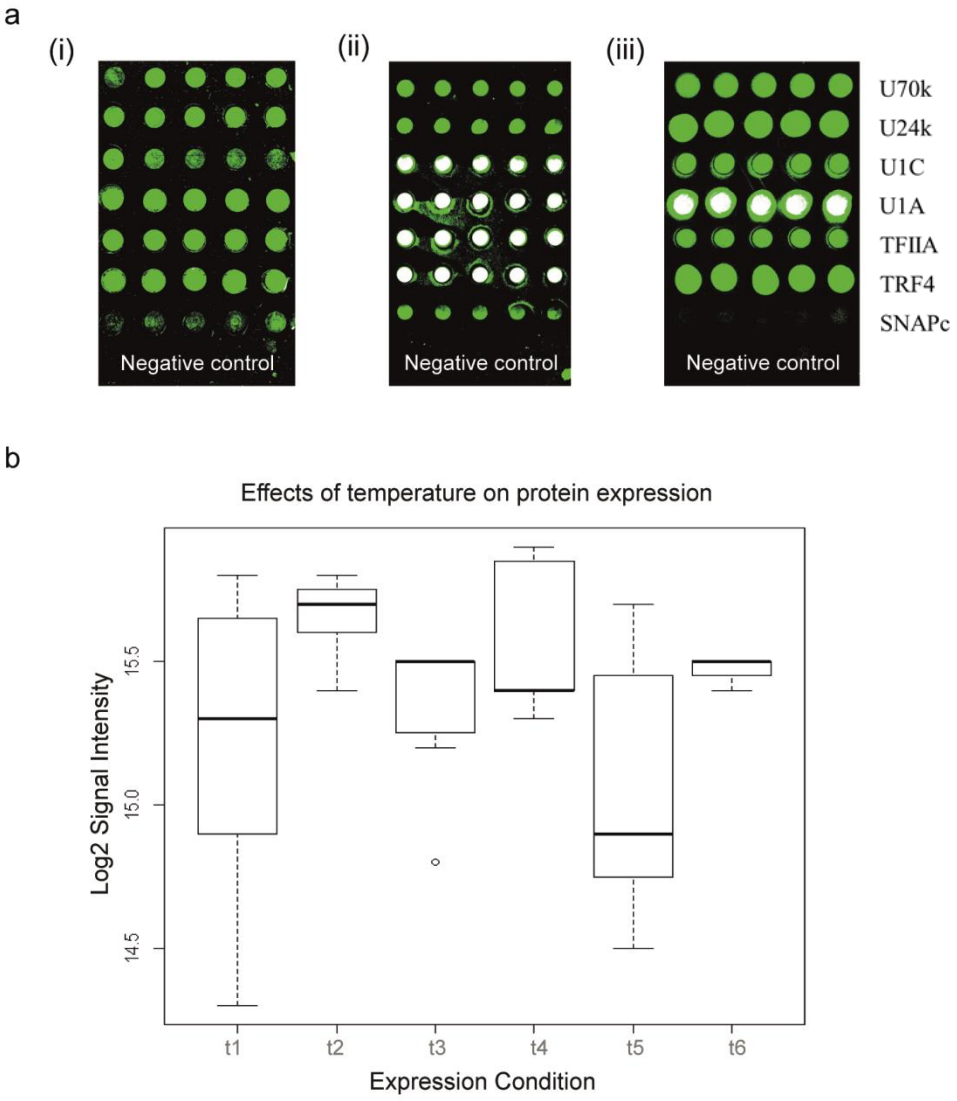


Figure 4: Optimal condition for protein expression microarrays. (**a**) Expression of seven different *T brucei* proteins (from top U70k, U24k, U1C, U1A, TFIIA, TRF4 and SNAPc) on nickel chelate (i), streptavidin-coated slides (ii), and the epoxy slides (iii). Expressed proteins were detected with antibody against c-terminal V5 epitope. (**b**) Comparison of incubation temperature for optimal protein expression. Slides incubation was performed at: t1) room temperature; t2) 37°C overnight; t3) 37°C 2Hour and 30°C overnight; t4) 37°C 2 hours room temperature overnight; t5) 37°C 2 hours; and t6) 37°C 1 hour and 30°C overnight.

Streptavidin coated slides equally gave similar spot morphology as the Nickel chelate slides (**Fig. 4a**), but with more background than the nickel slides but without tail-like features. Epoxysilane substrate gave a large spot, which at times diffuses, towards adjacent spots. However, it later could be controlled by adjusting the humidity in the incubation chambers (**Fig. 4a**). More so, there were no tails seen for this surface

coating, but some proteins whose expression could be detected on other surfaces could not be detected on epoxy substrate. Of all the surfaces, the epoxysilane was the only that gave very high signal to noise ratio, but the spot morphology was not as good as the other slide surfaces.

The incubation temperature and duration was equally a crucial factor affecting protein expression. By fine-tuning these parameters, the level of expressed protein for different proteins could be controlled. Incubation at room temperature after spotting the expression system yielded the least expression level of all proteins with a great variation in the level of expression for different proteins (**Fig 4b**). Incubating at 37°C overnight favored the expression of certain proteins, meanwhile others were not expressed under the same conditions. Furthermore, incubation for two hours at 37°C followed by incubation at 30°C overnight led to low and variable expression, and this was ameliorated when 37°C incubation was followed by incubation at room temperature. Finally, incubating at 37°C for 1 hour followed by incubation at 30°C overnight gave the best, though not highest expression. The expression level for all proteins was within a well-defined range.

Regarding template and expression system spotting volumes, a maximum of 0.9nl of template, but not less than 0.6nl gave the highest achievable expression (**Fig. 5a**). About 3.6nl of the S30 T7 High-Yield Protein Expression mixture was the optimal volume when 0.9nl were spotted, and 3nl was optima for 0.6 nl template (**Fig. 5b**). More so, higher volumes could not be supported while keeping good spot morphology and a high array spot density. Following optimal parameters, it was possible to express more than 85% of all amplified constructs, this, both for cDNA generated constructs (about 14000 constructs) and gDNA-derived constructs (384 constructs) (**Fig. 6**).

The MIST is a well-miniaturized version of the PISA and requires extremely low amounts of starting material for protein expression. This mean, that all conditions must be optimal to insure optimal protein expression. The failure to detect expressed proteins from the PurExpress, EasyXpress and TNT T7 for PCR could result from suboptimal expression. In fact, the PurExpress is said to produce between 10 and 200 µg/ml of protein per ml of lysate, while the EasyXpress produces just up to 50 µg/ml of lysate.

The TNT T7 for PCR on the other hand does not provide any information about performance. Its viscosity greatly limits its use in such a system, coupled with difficulties associated with detection of expressed protein, as it stains with most of the

commonly used antibodies. If only 3.6 nl of expression system is used, one would logically expect a yield of up to 50 and 200 picograms of proteins for more than 3 hours of incubation for the EasyXpress and pure express respectively.

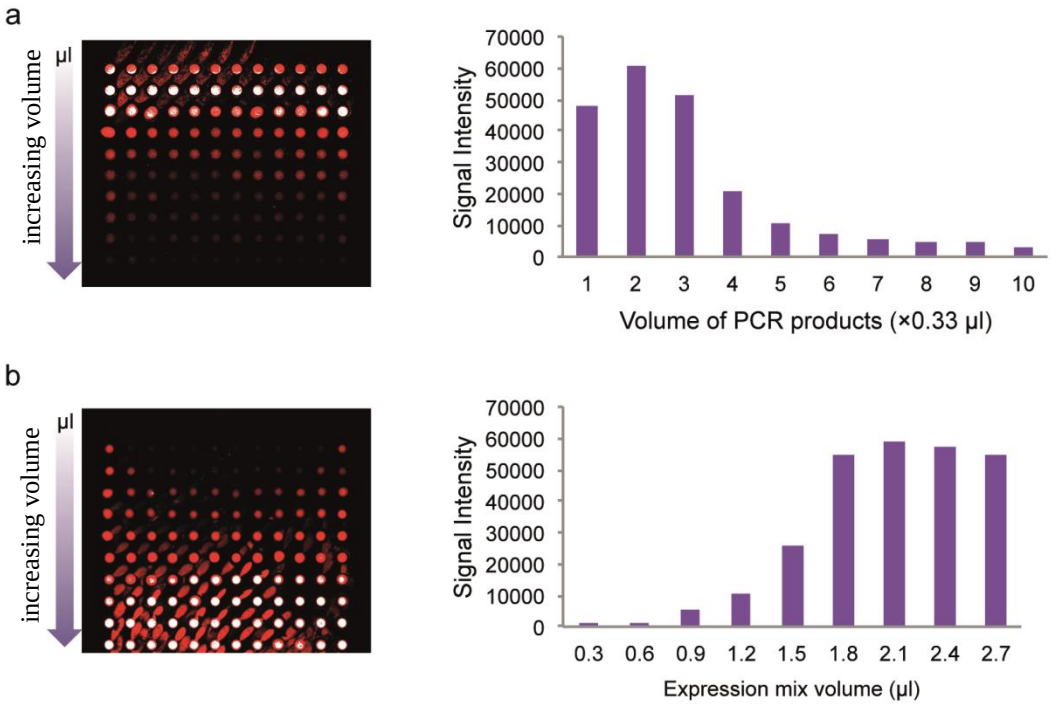


Figure 5: The optimal ratio of PCR products and expression mix volume. (**a**) Increasing DNA template (U1A as template) volume with constant expression mix quantity. (**b**) Increasing cell-free expression volume with constant PCR products quantity. microarray scanned image (***left***) and signal quantification (***Right***)

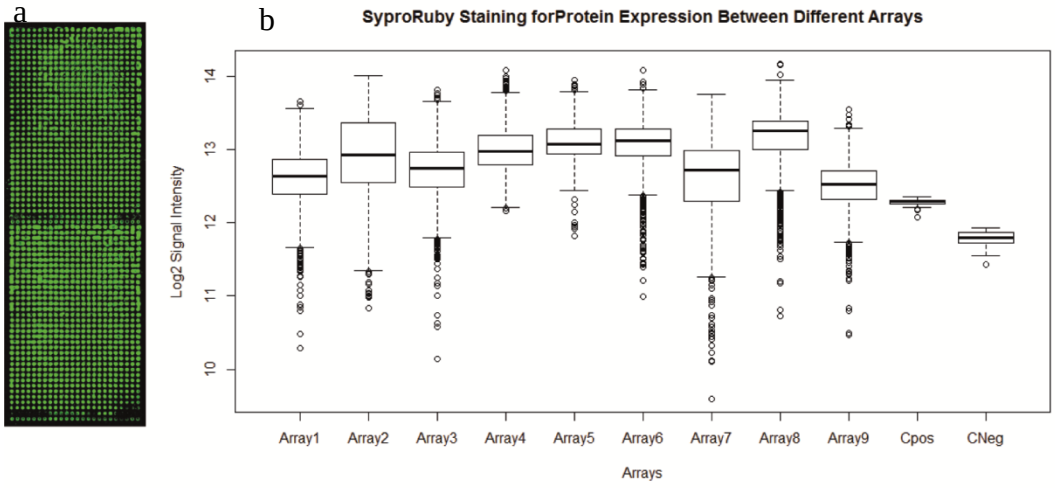


Figure 6: Protein expression microarray from *T. brucei* cDNAs as template. (**a**) The false color image of one out of nine slides of protein microarrays. (**b**) Signal intensity from more than 14,000 protein of *T. brucei*, in nine different protein microararys chips, each contains about 1600 spots representing individual proteins.

These are however only estimations for expression from plasmid DNA, which produces about 10 fold more protein than linear templates. It is therefore possible that the numbers are much lesser. Besides, the uncoupled nature of the EasyExpress makes it even less suitable for PISA technology. The S30 T7 expression system produces relatively high amount (500 µg/ml) in just one hour and this increases with incubation time. This might explain why overnight expression can allow for synthesis of detectable amount of proteins following the MIST.

There is currently no universal surface chemistry for the production of protein array and this must be determined empirically depending on array type and intended application[5, 97]. Protein expression on epoxysilane substrate was most consistent, but antibody-based detection was not suitable for the detection of all proteins. Spot morphology was also not uniform, and tends to smear at higher humidity. Poor detection on this surface could have resulted from random covalent immobilization, making some epitopes inaccessible to the antibodies. Smearing and morphology are certainly intrinsic problems related to the contact angle of the epoxy surface, which is about 105°. Nonetheless, epoxy slides were seen to give the best signal to noise ratios both for functional and analytical arrays[98]. Random immobilization on this platform might actually interfere with downstream applications as a result of modifications on some amino acid residues that are involved in protein activity. However, some functional applications performed so far, suggest that even if a considerable amount of the active sites might not be accessible, there is still enough for screening purposes. Nickel chelate and streptavidin coated surfaces produce uniform spots, with little or no smearing. They also provide for oriented immobilization and this could be interesting in cases where full protein functionality is required. However, since these surfaces are less inert than the epoxy substrate, a low signal to noise ratio is characteristic of them, and Nickel chelate surface bind with weaker strength and easily saturated leading to tail-like features. Such surfaces are known to be suitable for acetylation studies, but not phosphorylation[97]. It still unclear, why nitrocellulose and hydrogen surfaces failed on this platform since they are routinely used in most proteomic applications.

Protein abundance can very much influence the outcome of an assay. Without any handy and straightforward method to control the level of expression of the several thousands of proteins on a PISA and hence avoid abundance biases, fine-tuning expression parameters might be a useful but indirect method. Discrepancies observed in expression with different incubation parameters might be related to the relative stability

of individual transcripts, and ease of expression. Meanwhile, an incubation period of 1 hour at 37°C and 30°C overnight could be helpful in normalizing expression levels. Although it is not possible to prescribe a unique surface for the production of MIST-based arrays; protein expression levels can be controlled. Depending on the desired application, a suitable surface can be selected between epoxy, Nickel and streptavidin substrates. More than 80% expression can be achieved from cDNA-derived templates. However, whether or not these are full-length proteins, or if they are in the correct frame cannot be guaranteed for all expressed proteins. Nonetheless, one would expect a third of the expressed proteins to be in the right reading frame. This is of prime importance when investigations into disease-related proteomes are intended. Finally, the combined use of cDNA products and miniaturized sample volumes required make the MIST a cost efficient and available high throughput proteomic tool.

3.1.3 Protein Detection

Generation of protein microarrays from open reading frames has the advantage of flexibility over cDNA products. This actually allow for the addition of epitope tags for protein detection. Without information of all proteins expressed from cDNA products, and the lack of protein-specific antibodies for all expressed proteins, complementary detection strategies must be adopted. Two strategies were used to detect proteins expressed on the microarrays. For protein expressed from ORFs, antibody staining was used to detect the N-terminal 6×His epitope tag. Detection of full-length proteins using C-terminal V5 epitopes was also used in when this was possible. Arrays Generated from cDNA products on the other hand could not be detected in a similar manner. As such, total protein staining was optimized for protein detection.

The SyproRuby protein gel stain was used to stain expressed protein, with comparable outcomes (**Fig. 7**). SyproRuby staining gave a more uniform result across several proteins, while antibody detection was highly variable. However, there was a near background signal for negative controls with antibody detection, while SyproRuby gave a relatively strong signal, which was clearly less than the signals from expressed proteins. Much caution is however needed when interpreting such resuts in order to avoid erroneous conclusions and care must be taken to avoid over staining.

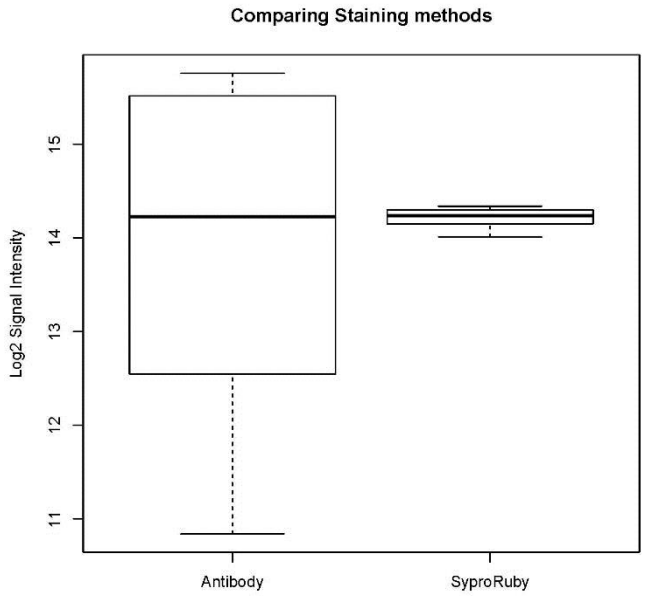


Figure 7: Comparing terminal tag-antibody detection and SyproRuby staining of seven expressed proteins (U70k, U24k, U1C, U1A, TFIIA, TRF4 and SNAPc).

SyproRuby is one of the most recommended protein staining dyes due to its high sensitivity[132]. Being a total protein stain, it was expected to raise signals against the negative control, which contains proteins from the PCR reaction and the cell-free expression system. The fact therefore, that cDNA products cannot be fused to epitope tags for detection cannot rule out the utilization as template source for protein expression. They are however not the best for whole proteome screening, but most intendend when the need to screen the cellular proteome under a given condition is desired.

3.1.4 Protein Quantification and Functionality

Since functionality was an aspect not to be left behind, we evaluated the possibilities, that functional proteins could be made by cell-free expression systems, and that *in situ* expression will maintain functionality if any. Trypanothione reductase was amplified from gDNA and used to generate an expression construct. The construct was expressed in a tube format and its activity was measured by measuring the absorption of NADPH at 340nM as a function of time (**Fig. 8a**). A linear decrease in the absorbance of NADPH could be observed confirming its depletion as it is used to for the reduction of trypanothione in the crude mixture. On chip protein functionality was tested using green fluorescent protein.

For this purpose, purified recombinant green fluorescent protein (GFP) was spotted in serial dilutions, and on the same array, GFP was also expressed using the MIST. One array was scanned at 488 nM for the auto-fluorescence of GFP, and another slide was first treated with 6 M guanidine hydrochloride at 80°C before scanning. Signals were obtained for both the expressed and spotted standard GFP (**Fig. 8b**). Meanwhile the guanidine hydrochloride-treated slide gave a blank image. A standard plot was made from the spotted GFP, from which the approximate amount of the expressed GFP could be extrapolated. About 4 fmole of GFP was synthsized, corresponding to about 800 pgram of protein.

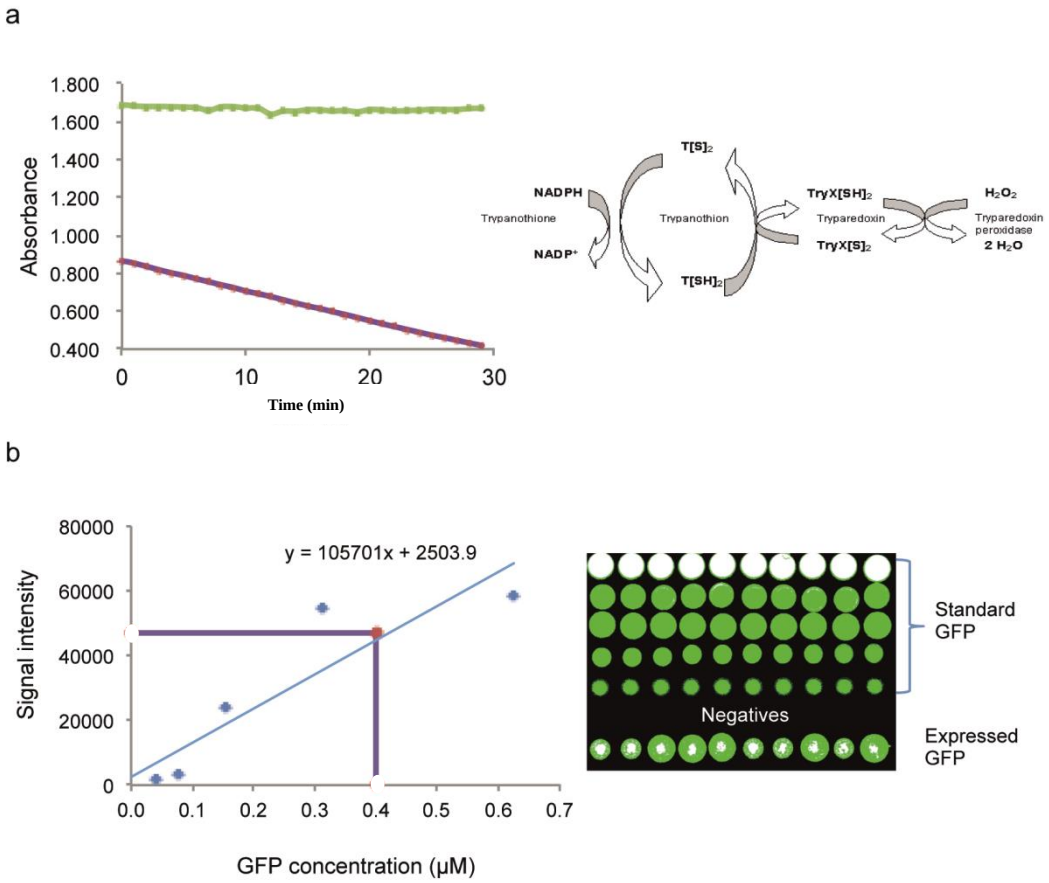


Figure 8: Function test of cell-free expressed protein. (**a**) *In vitro* protein functionality test of cell-free expressed (unpurified mixture) Trypanothione reductase and Scheme of TR reaction (*purple line*) and control (*green line*) Absorbance of NADPH is measured at 340nm as it depletes. (**b**) Auto-fluorescence of standard GFP (◆) and cell-free expressed GFP (●). TR reaction adapted from Patricia Paglini-Oliva and Hector Rivarola (September 2002)

3.2 Functional Applications

3.2.1 Protein-Protein Interactions

We performed protein-protein interaction screen on the cDNA and gDNA-derivedy arrays. For this purpose, the *Trypanosoma brucei* CAF40 (Tb927.4.410), a member of the CCR-NOT complex was expressed as a 6×His fusion protein and purified by affinity chromatography (**Fig. 9a, b**). The protein was fluorescently labelled and screened against proteins expressed on the arrays. Two proteins, the *T brucei* Aurora kinase B and *T brucei* polo-like kinase were used as background control to fish out sticky proteins. Previously reported, as well as novel interactions were found pending in vivo validation (**Fig. 9c**). For example, it has been reported[133], that CAF40 can interact with the Nascent polypeptide-associated complex as well as the 60s ribosomal protein S3[134]. These interactions as well as other novel interactions could be found. The same protein was screened against another targeted protein microarray, leading to identification of other binders. Novel interactions including hypothetical SWAP domain proteins and the DHH1 helicase were identified. The ATP-dependent RNA helicase DHH1 is also known to interact with CAF 40 in *Saccharomyces cerevisiae*[135, 136]. In trypanosomes, it is known, that CAF 40 interacts with the NOT1 scaffold protein of the CCR4-NOT complex[137]. This interaction could not be confirmed in the screen with the targeted arrays containing two fragments of the NOT1 protein.

Interactions of *T. brucei* CAF 40 with NOT1, which was shown to exist by pairwise yeast two-hybrid, could not be confirmed[137]. Using the N-terminal fragment of NOT1 as a positive control, for a pair-wise screen, this interaction was not found. (See yeast two-hybrid data). It is possible, that this fragment alone could not allow for a strong binary interaction on the array and was therefore washed away, or it might not have folded properly. The CCR4-NOT complex has been shown to play key roles in cellular proteostasis, by controlling protein quality[124, 125]. The interaction of CAF 40 with components of the NAC complex could imply several roles including proteostasis control. The NAC has been shown to be involved in the control of cellular proteostasis[138], and it is therefore possible, that the CCR4-NOT not only play roles in RNA turnover as it has previously been known, but also recruit chaperon for other functions.

In yeast, the NOT4 component of the CCR4-NOT complex is responsible for the ubiquitination of the NAC. However, its *T. brucei* homologue does not exist, or at least is not known to date. DHH1, play a role in mRNA decapping, and is known to interact with dacapping and deadenylase complexes[139]. The CAF-NOT complex in trypanosomes is responsible for mRNA deadenylation among others. Interaction between the DHH1 and CAF 40 could well be indeed a functional interaction. Confirmations of these interactions via complementary methods are currently being pursued.

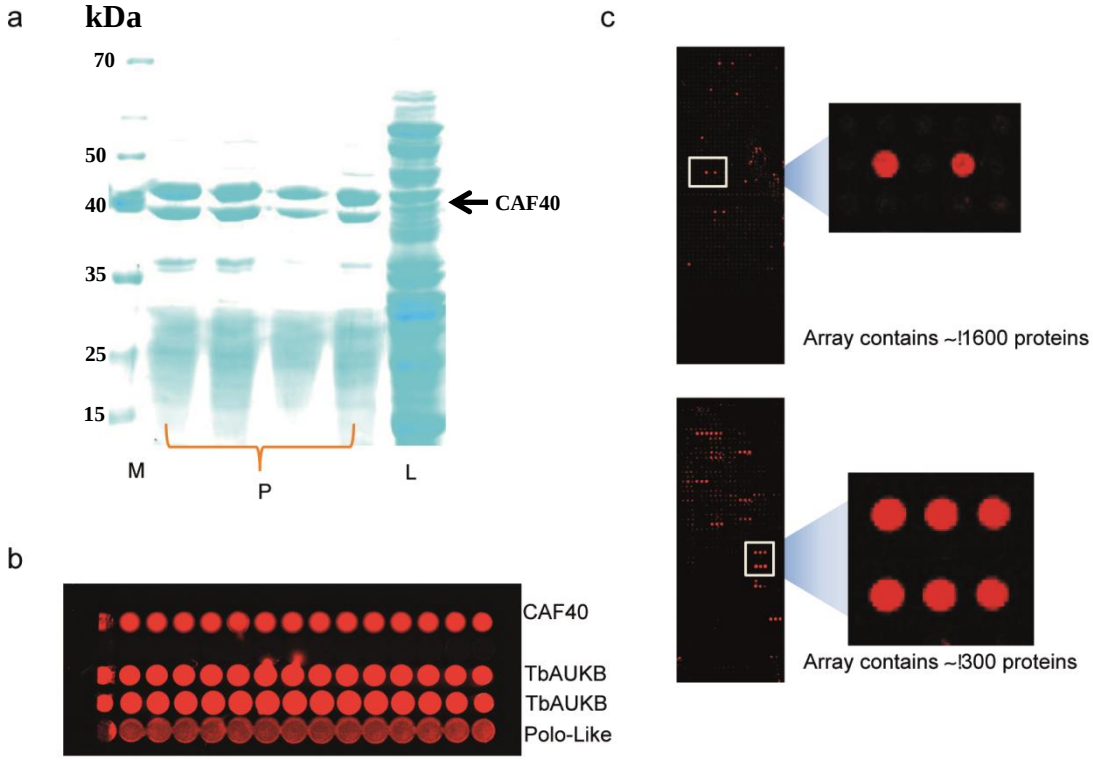


Figure 9: Protein-protein interactions. (**a**) Coomassi staining for purification of CAF 40 by IMAC. M=Mw Marker, P= partially purified protein, L= Lysate (**b**). Detection of spotted (Each protein is spotted across the row) purified protein with anti 6×His antibody (**c**) Protein-protein interactions for CAF 40 on large array and mini array examples of triplicate signals are indicated.

3.2.2 Protein-RNA Interactions

Protein-RNA interactions were also performed on targeted arrays initially as proof of concept and functional screens. Well-characterized interactions between specific RNA sequences or molecules and known protein could be reproduced. In this case, the U1 snRNA of *Trypanosoma brucei*, involved in the splicing of the polyA polymerase was screened against the U1 proteins. It was reported, that the U1-70K protein can specifically bind the U1 snRNA [140]. We used this as a model for our

protein-RNA interactions, and were able not only to reproduce this interaction, but could identify a binding effect between the U1 snRNA and the U1A protein which has equally been reported elsewhere[141] to have an affinity for the U1 snRNA (**Fig. 10a**). Other reported protein-RNA interactions were tested, and the arrays could differentiate a wild type from a mutant TbZC3H11 (**Fig. 10b**) which was recently shown to bind to the 3′UTR AU-rich element (ARE)[142]. Single mutation in the Zinc finger domain of this protein was swufficient to abolish binding on a shift assay; but a triple mutation was needed on the microarray (**Fig. 10c**).

Table 1: Proteins bound by both the classical and non-classical ARE

Gene ID	Name	ZScore/FD Classical		Z-Score/FD Non classical		Function
Tb927.11.530	RBP3	11,5	9,7	16,4	16,5	RNA binding
Tb927.11.500	UBP1	2,5	2,5	6,1	6,9	RNA binding
Tb927.5.810	ZC3H11	0,8	2,3	1,4	2,5	RNA binding
Tb927.9.6870	RBSR1	1,9	3,0	1,2	2,4	RNA processing
Tb927.7.1390	RBSR2	2,8	3,7	1,7	2,9	RNA processing
Tb927.11.2250	hypothetical	4,1	4,6	3,5	4,5	
Tb927.7.4660	hypothetical	7,2	7,2	3,7	4,7	
Negative Control		-0,3	1	-0,4	1	

FD= Fold difference with Negative control

A targeted array screen with the classical and non-classical ARE sequences identified other binders, including the ZC3H11 included on the array as a positive Control (**Fig. 10d).** Both sequences shared common binding proteins such as the ZC3H11 (**Table 1**), but equally had proteins that exclusively bound just one of the sequences. The non-classical sequence bound with low affinity to than the classical.

As background control, a polyU, a polyC and a mutant of the ARE were screened as well (**Fig. 11a**). Other sequences were screened against the targeted array, including a sequence reported to bind to TbPUF9 and 3′ UTR fragment for the trypanosome hexose transporter (THT), the amino acid transporter (AAT1) expression site associated gene 9 (ESAG9), Phosphoglycerate Kinase (PGK), Cytochrome Oxidase (COX) and the EP procyclin genes. The number of binding proteins increased with RNA probe sequence length which varied from 23nt (ARE) to more than 600nt (THT) Most of the sequences screened were found to bind UBP1 and RBP3 except the polyC sequence which bound to all protein at background level. However, the PUF-binding sequence reproducibly bound a hypothetical protein but not to other proteins on the array (**Fig. 11b**).

This sequence was expected to bind PUF9, which was equally expressed on the array, but it failed to. The PGKCR, the EP1 and to a lesser extent the COX reproducibly bound to a set of identical proteins including the ALBA domain protein and the DRBD proteins (**Table 2**) all of which are involved in RNA stability and translation[117, 143]. These sequences were seen to bind to a significant set of exosome complex proteins and like the other to many other hypothetical proteins. ESAG9 UTR and the PGKCR bound to a relatively higher number of proteins; meanwhile COX bound a rather restricted number, although they were all of identical lengths. The THT UTR, as a result of its size was fragmented and analyzed in a minimal sequence-mapping assay.

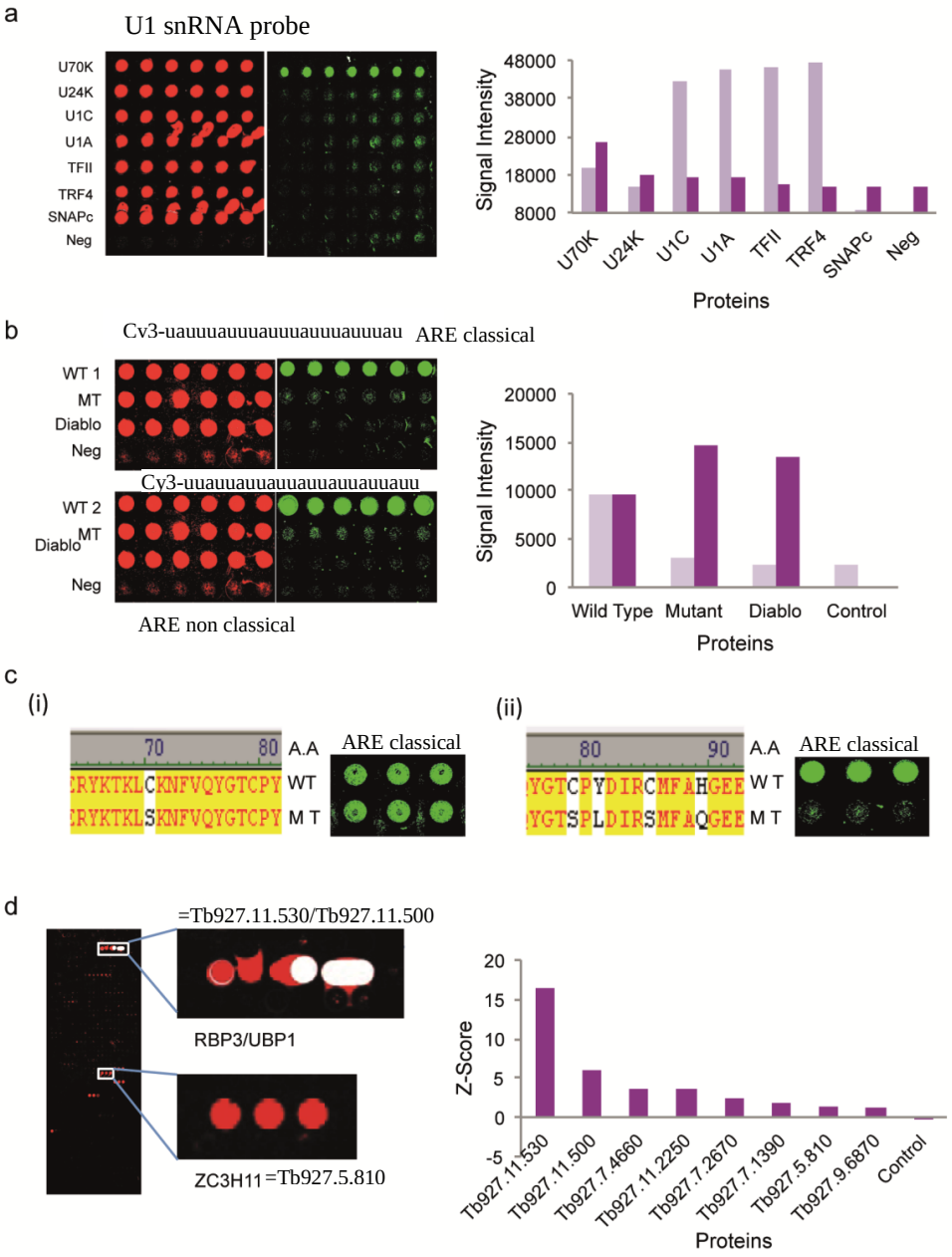


Figure 10: Protein-RNA interaction for AU-rich element and U1 snRNA in proof of concept. (**a**) False color image, confirmation of expressed proteins (***red*** array) and detection of U1-protein interactions (***green*** array). Plot of signal intensities of expressed proteins and it interaction with snRNA U1 (***light*** and ***dark*** color bar, respectively). (**b**) Expression and the interaction of the classical and non classical ARE with TbZC3H11 WT and Mutants (triple mutant) together with expression control (Diablo Human mitochondrial Protein). (**c**) Comparison of single and triple mutation of CCCH domain on the array. (**d**) Interaction of ARE on mini array and plot of *Z*-score of protein binders of classical ARE.

AA= amino Acid, WT=wild Type, MT= Mutant Type

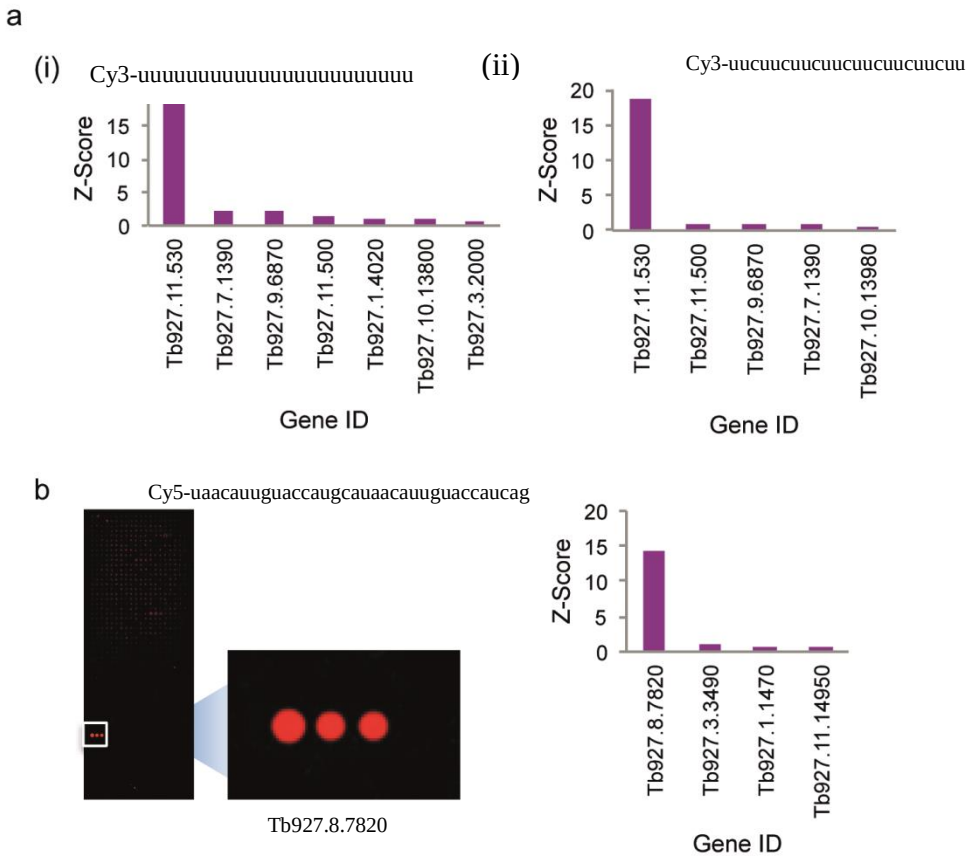


Figure 11: Protein-RNA interaction on mini array. (a) Plot of poly U binding proteins (i) and plot of UUC binding proteins (ii). (**b**) Mini array was screened with PUF-binding element (*left* panel). One protein was reproducibly bound to the array. Plot of PUF-binding protein (*right* panel).

In addition to U1-70K which is known to bind the U1 SnRNA in trypanosomes, we also found the U1A binding the same molecule, though with weaker strength. U1A in humans bind to the U1 SnRNA more strongly than U1-70K[140]. The weak signal could either result from less active sites exposed on the epoxy surface, or that there is simply transient interactions between both. Interactions between ZC3H11 and the ARE is known to regulate heat shock protein 70 (HSP70) mRNA in bloodstream forms[142]. Both the classical and non-classical ARE were shown to bind wild type ZC3H11 but not its mutant. A single mutation of a cysteine residue was sufficient to abolish binding, but only a triple mutation of two cysteine residues and a histidine residue of the CCCH domain could abolish binding on microarrays.

The single mutant could actually have bound the RNA, but the complex, due to its weak interaction simply might have dissociated as it ran through the gel. This was not the case on microarray platform. The PUF domain used was reported to bind PUF9 and regulate the stability of a subset of mRNA[119]. This sequence did not yield any

significant binding on the arrays, either as a result of low expression or folding on the epoxy surface. It is conceivable, that the 3′ UTR of the EP1, PGKCR and COX contain *cis*-acting elements involved in mRNA stabilization, since most of these sequences bound significantly to the DRBD protein family. The polypyrimidine tract binding proteins are known to regulated RNA stability in trypanosomes[144].

Table 2: Common proteins binding the EP1, PGK, and COX UTR fragments (Z-score≥1,5)

Gene ID	Symbol	Function
Tb927.11.10020	short-chain dehydrogenase. putative	Enzyme
Tb927.7.2440	P5CR	Mitochondrial pathway
Tb927.9.8740	DRBD3 (PTB1)	RNA binding
Tb927.7.320	RBP8	RNA binding
Tb927.8.710	DRBD17	RNA binding
Tb927.3.3960	DRBD6A	RNA binding
Tb927.4.400	DRBD7	RNA binding
Tb927.5.3750	PPCT1	RNA binding
Tb927.7.1180	RBP19	RNA binding
Tb927.11.3340	RBP34	RNA binding
Tb927.11.12120	RBP9	RNA binding
Tb927.7.7280	RNA binding protein. putative. SWAP domain	RNA binding
Tb927.3.1340	CBP66.	RNA binding
Tb927.8.2810	XRNC	RNA degradation
Tb927.8.3220	exonuclease. putative	RNA degradation
Tb927.2.2180	exosome protein PH subunit RRP41B	RNA degradation
Tb927.3.4210	hypothetical protein. conserved	
Tb927.10.14150	hypothetical protein. conserved	
Tb927.10.13980	hypothetical protein. conserved	
Tb927.5.1600	hypothetical protein. conserved	
Tb927.5.4450	hypothetical protein. conserved	
Tb927.6.3420	hypothetical protein. conserved	

Samples shown have fold difference of ≥2 for all probe considered

Besides, the EP1, the COX and the PGK all share a common 26-mer element that is said to control RNA stability and translation[145, 146]. It is therefore not astonishing, that these sequences all show preference for the polypyrimidine binding proteins and the ALBA domain proteins, which are all involved in mRNA stabilization and translation. Exosomes are major machinery in RNA degradation following deadenylation of the 3′ extremity. Finding exosomes binding to 3′ UTR elements is

therefore not surprising. However, the fact that some sequences, but not others bound to a relatively high number of exosome complex proteins is unusual.

3.2.2 Minimal Sequence Mapping of the THT1

A 600-nt fragment of the THT UTR was fragmented to generate fragments differing by 100nt in length. Fragments were fused to a T7 RNA polymerase construct for transcription and subsequently labelled at the 3′ end with biotin. RNA-protein interaction was identified by incubation with fluorescently labelled extravidin. Fragments bound between 40 and 106 different protein with the number increasing with probe length. The most reproducible results were obtained for the 100, 200 and 300-nt long fragments. Beyond this size, there was little consistency in the binding patter. Analysis of the folding patterns of the different fragments was done using RNAfold web server, and this led to the identification of a stable stem loop across all fragment and with all parameters applied (**Fig. 12a** and **b**).

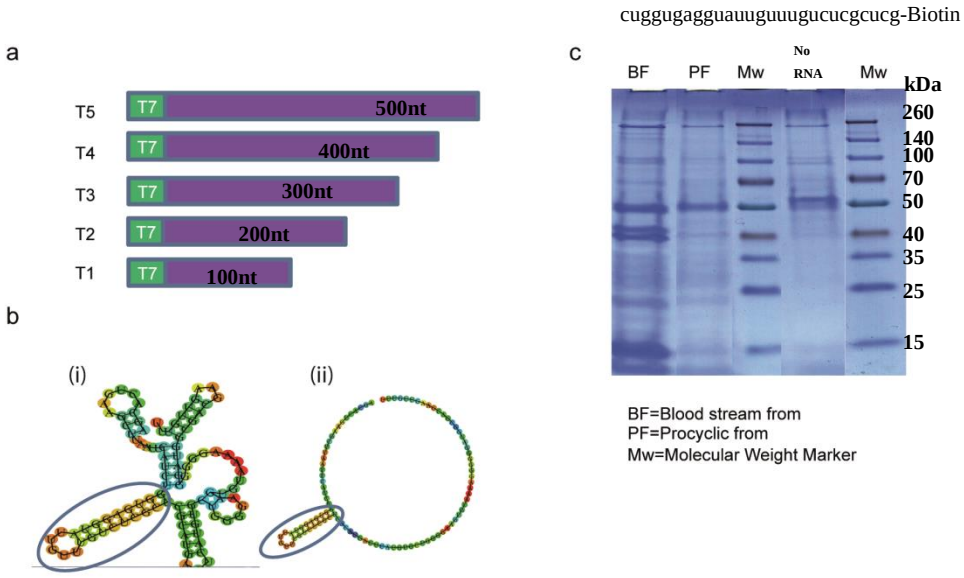


Figure 12: Identification of potential sequences element for THT interaction. (**a)** Minimal motif mapping of the Trypanosome Hexose Transporter TH1 3′ UTR. Fragment length increases from the 5′ end of the 3′UTR from T1 to T5. (b) RNAfold in Silico analysis of THT1 fragments. (i) Minimal fold energy (ii) Centroid secondary structure. Colors indicate the probabilities of base-pairing, increasing from green to red. (**c**) SDS-PAGE gel for enrichment of RBP using THT oligomer**.** Eluate from beads was stained with coomassi

The stem loop, which was 26nt long, was synthesized and biotin-labelled as with the others. Screening this oligomer against the array gave about 80% of the proteins that reacted with the parent fragment (**Table 3**). The fragment was then labelled with desthiobiotin, and used to pull-down proteins from blood stream form and procyclic cell

lysate. There was an important protein enrichment in the blood stream lysate compared to the bead control, and procyclic (**Fig. 12c**).

Longer probes tend to give more unspecific binding than their short counterparts due to the possibility of multiple loop structures. It is well possible, although this must be confirmed otherwise, that the conserved stem loop across all fragments might contain the element needed for binding. Bloodstream form express about 40 fold more of the THT1 than THT2. THT2 is expressed in a glucose-dependent manner in procyclics[147].

Cell lysate enrichment is of course not an ideal procedure for the identification of RNA-binding proteins, but could provide an insight into RBP enrichment between lysates or conditions. The data generated from such enrichment must therefore be analyzed with much caution. Nonetheless, using such an approach as a complementary method is worth the pain.

Table 3: Shared binders of THT1 oligomer and precursor fragment (Samples Z-score≥1,5)

Shared between Both	Oligomer alone	Description
Tb927.10.11270	Tb927.10.11270	RBP23
Tb927.10.12210	Tb927.10.12210	ribulose-5-phosphate 3-epimerase. putative
Tb927.10.12330	Tb927.10.12330	ZC3H34
Tb927.10.14150	Tb927.10.14150	hypothetical protein. conserved
Tb927.10.1860	Tb927.10.1860	hypothetical protein. conserved
Tb927.11.10020	Tb927.11.10020	short-chain dehydrogenase. putative
Tb927.11.13970	Tb927.11.13970	Pab1p-dependent poly(A) ribonuclease subunit PAN3
Tb927.11.14100	Tb927.11.14100	DRBD4 (PTB2)
Tb927.11.14520	Tb927.11.14520	polypeptide deformylase. putative
Tb927.2.2180	Tb927.11.8290	exosome complex exonuclease RRP44
Tb927.3.1340	Tb927.11.950	DRBD14
Tb927.3.2930	Tb927.2.2180	exosome protein PH subunit RRP41B
Tb927.3.3960	Tb927.2.4710	TRRM1
Tb927.3.4210	Tb927.3.1340	ZC3H7=CBP66. nuclear cap-binding complex
Tb927.6.3420	Tb927.3.2930	RBP6
Tb927.6.850	Tb927.3.3960	DRBD6A
Tb927.7.2440	Tb927.3.4210	hypothetical protein. conserved
Tb927.7.7280	Tb927.6.3420	hypothetical protein. conserved
Tb927.9.8740	Tb927.6.4270	hypothetical protein. conserved. NADH dehydrogenase
	Tb927.6.850	NOT complex subunit NOT2
	Tb927.7.2440	pyrroline-5-carboxylate reductase. putative (P5CR)
	Tb927.7.7280	RNA binding protein. putative. SWAP domain
	Tb927.9.10280	ZC3H48
	Tb927.9.8740	DRBD3 (PTB1)

Samples have fold differences between 2.5 and 7.5 with respect to negative control

3.2.3 Kinase Substrate Identification

As proof of concept, the identification of kinase substrates was done using the *T. brucei* Aurora kinase 1. Two of its substrates Centrin II and Histone H3 were expressed by MIST, and incubated with purified kinase. Two other unrelated proteins were equally expressed together as phosphorylation control. All proteins were well expressed (**Fig. 13),** but only the two substrates of the kinase showed significant phosphorylation following incubation with the enzyme. Protein phosphorylation was detected by incubation with anti phosphoserine/threonine and tyrosine antibody. Histone H3 gave a stronger signal compared with Centrin II and the other proteins gave background signals. Protein microarrays have been used for the identification of kinase substrates[12]. However, this has only been done with printed arrays and there are no reports of the performance of protein *in situ* arrays regarding such an application. Besides, all reported cases have been based on radioactivity. Combining PISA, protein-protein interactions and kinase substrate identification could lead a revolution into the understanding of complex signaling cascades.

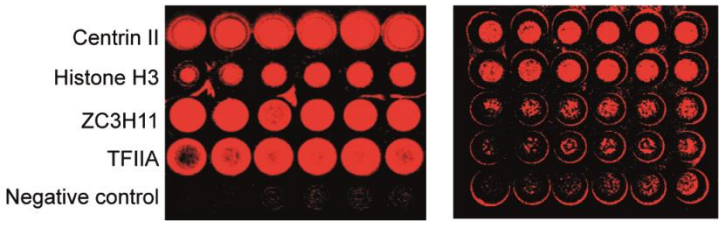


Figure 13: Proof of concept for on-chip protein phosphorylation. Confirmation of expressed TbAUK B substrates by c-terminal anti-6xHis antibody (***left array***) and on-chip phosphorylation detected by antiphosphoprotein antibody (***right array***). All proteins are replicated across the row.

3.2.4 Antibody Selection

Microarrays made from cDNA products were screened with protein specific antibodies in a pool format. This was meant first to verify if some of the expressed proteins were correct in their primary structure, and secondly to find out if the array could be useful for the selection of antibodies. Antibodies to *T. brucei* proteins such as Trypanothione reductase, Tryparedoxin alpha tubulin and XRND were tested. Antibody binding was detected using the corresponding anti mouse antibody conjugated to Cy5. Some of these antibodies could specifically bind to their antigens, and this was later confirmed by sequencing (**Fig. 14**). Tubulin was found to be highly abundant with 3 out of five sequences corresponding to tubulin. These findings were interesting, since this

will mean that such array can be used to identify antigens for antibody libraries, or reactivity to protein isoforms or simply to check for antibody quality.

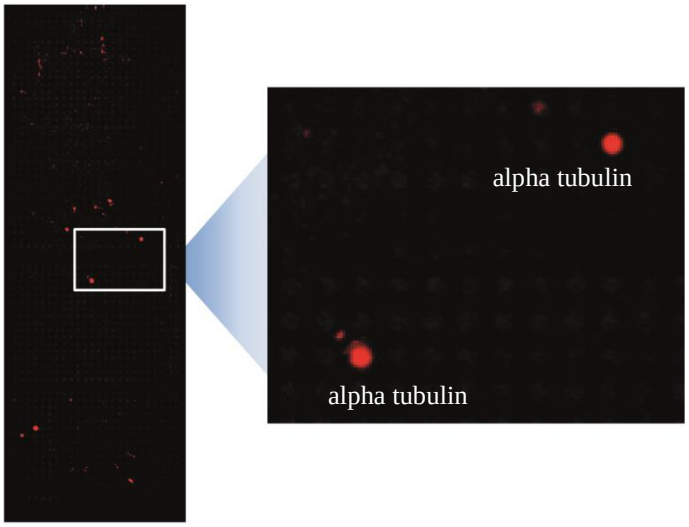


Figure 14: Pooled antibody screen on large array using a pool of 21 *T brucei* protein-specific antibodies. Antibody binding was confirmed by detection with Cy5-conjugated antimouse antibody. Only one array shown here on the right is an enlarged portion of the array.

3.3 *In Vivo* Studies

There are often many discrepancies between results obtained from high throughput platforms when compared with data generated *in vivo*. However, since in vivo techniques are rather less adapted for high throughput application, it is important that high throughput platforms be validated to make them more reliable. Besides, any biological data will only be useful, if it reflect a real life situation. For this reason candidate proteins were chosen for *in vivo* validation.

3. 3.1 Gene Knockdown

Fragments of the coding region of the genes were cloned into the vector p2T7, digested with NOT1 and transfected into BF 1313514 cells for double stranded RNAi and selected in Hygromycin-containing medium. RNAi was induced with 0.75 µl/10 ml of tetratcycline continuously for 6 days. Knock down of two of the four genes led to a lethal phenotype. One of the lethal phenotypes occurred as early as 48 hours post induction (**Fig. 15a**) and the other 6 days post induction (**Fig. 15b**). Of the other two genes, one led to mild growth defects meanwhile the other had no major difference with control cell (**Fig. 15c** and **d**). The lethal phenotype observed could suggest that these genes are essential genes in bloodstream form of the parasite.

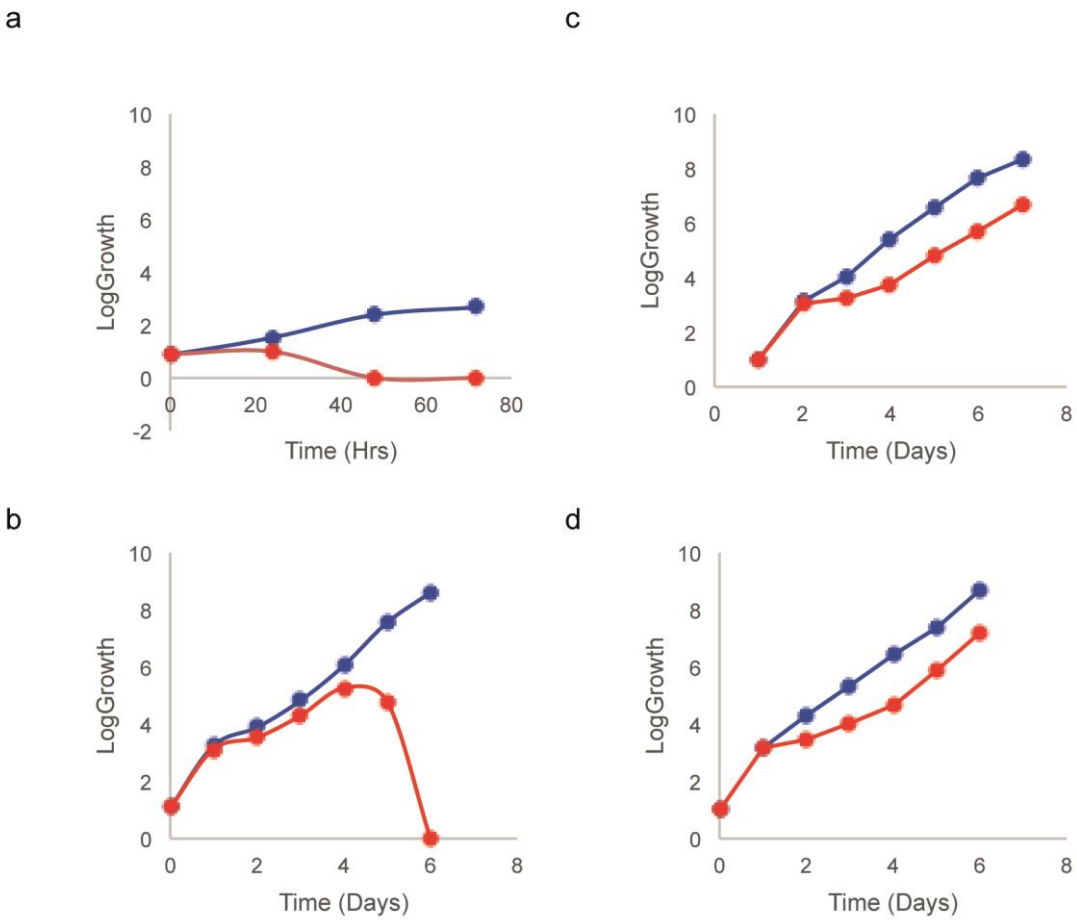


Figure 15: Cummulative growth curves for dsRNAi knockdown of candidate genes. (**a**) Knockdown of Tb927.10.2190. (**b**) Knockdown of Tb927.10.14150. (**c**) Knockdown of Tb927.10.12330. (**d**) Tb927.5.1990 Knockdown. Tet− (*blue*) and Tet+ (*red*).

3.3.2 *In Situ* Tagging, Immunofluorescence and Pull-Down

In situ epitope tagging was performed for all four genes using PCR-generated blasticidin-V5 cassettes. The constructs were transfected into BF 1313514 cells and selected in blasticidin-containing medium. Three of the four genes were successfully tagged, while the other failed in five successive trials. Successful tagging of the genes could be confirmed by western blot using anti-V5 antibodies (**Fig. 16a**).

Wild type BF1313514 cells and BF13132277 cells carrying an N-terminal c-myc tagged GFP were used as negative controls. Proteins were extracted from all cells including the controls, and 40 µg of protein was immunoprecipitated from each lysate. The precipitating protein fraction was used for RNA extraction, and part was analyzed on SDS PAGE. It was possible to amplify fragments of some of the UTR used in the

microarray experiments. All the fragments of the THT1(T1 to T6) could be amplified strongly; meanwhile the PCR for the other UTR was positive with weak bands at background level. The PCR efficiency was variable between all three genes. One of them (Tb927.10.2190) gave very weak bands, while the others (Tb927.10.12330 and Tb927.10.14150) stronger bands. There was low amplification both for the wild type control and the GFP control (**Fig. 16b**). The amount of bound RNA as revealed by PCR did not correlate with the binding intensity from the microarray experiment for one of the proteins (Tb927.10.2190). Immunofluorescence microscopy revealed, atleast two of the proteins localized to the cytoplasm, although the images are not the best (**Fig. 17**).

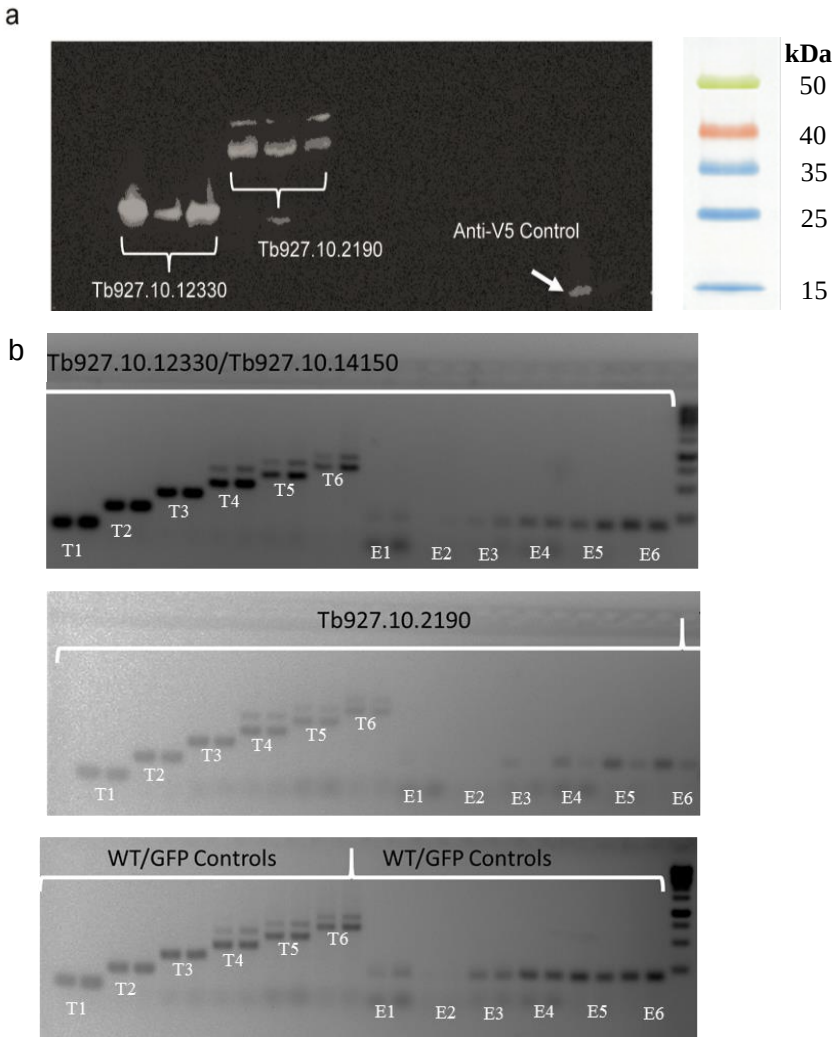


Figure 16: Epitope tagging and RNA IP for candidate genes. (**a**) *In situ* epitope tagging confirmed by Western Blot via N-terminal V5 epitope tag. (**b**) RT-PCR of UTR from RNA IP. Primers used to amplify the initial fragments were used for PCR. THT1 UTR fragments T1 through T6 increases in length by about 100nt and loaded in duplicates. E1to E6 are UTRs of the AAT1, PGK, COX, ESAG9, EP1 and Gim5 respectively.

Unfortunately, the GFP and wildtype controls did bind RNA, even if the amplification of THT 1 RNA bound to two of the proteins (Tb927.10.12330 and Tb927.10.14150 was higher. This may be interesting, but must be interpreted with caution, since the THT1 might have been abundant. However, if this would be the case, one would expect to have similar amplifications on all samples, which is not the case. This data could therefore be preliminary interesting data that needs to be quantified. Moreso, if a protein destabilizes RNA, it is very unlikely to find RNA binding to it, or at least in significant mount. One would therefore speculate that the weak bands obtained for one of the proteins could imply its involvement in RNA decay or such like. Meanwhile, the other might be involved in translation and mRNA stabilization. Polysome analysis could shed more light on this. This is being investigated currently via CAT reporter assays.

The current data from the immune fluorescence assay are not very conclusive. Although one could tell that some of the proteins localize to the cytoplasm, it will be necessary to repeat the assay with better resolution to be able to draw proper conclusions. If in effect these proteins turn out to be cytoplasmic, their potential role in RNA metabolism could be more evident.

In all it is very likely, that the RNA-binding obtained on the arrays could actually be real pending more validation. Interpretation of the data must however be taken with caution especially when dealing with very long probes as there is a tendency, that they form secondary structures that help them bind even strongly to some proteins.

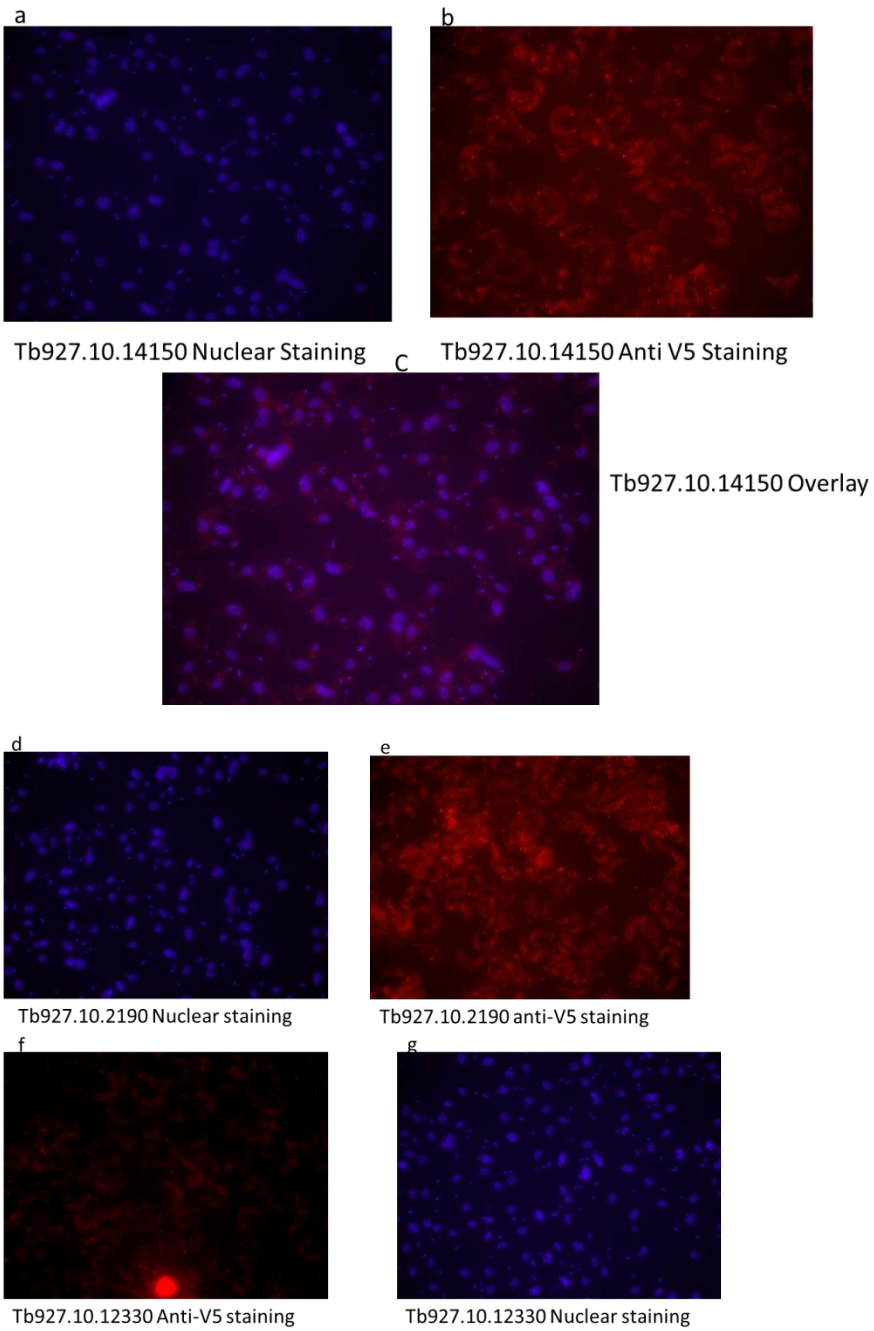
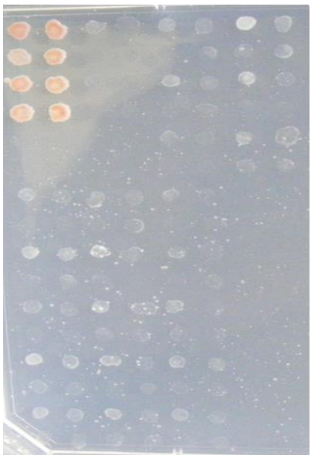


Figure 17: Immunofluorescence assay for epitope tagged endogenous proteins. Nuclear staining for Tb927.10.14150, Tb927.10.2190 and Tb927.10.12330 respectively (a,d and g). Anti-V5 immunostaining for Tb927.10.14150, Tb927.10.2190 and Tb927.10.12330 (b,e and f). Overlay of Nuclear and epitope staining,for Tb927.10.14150 (C).

3.3.3 Yeast-Two Hybrid

Initial pair-wise yeast two hybrid experiments were performed (By the genomic and proteomic core facility). One samples gave a positive result with the CAF40 as bait (**Fig. 18**). This is the alpha subunit of the NAC which was one of the binders from microarray experiments. The fragments of the NOT1, which has been shown to bind to CAF40 failed to bind in this assay. These results were however only obtained for one to one bait-prey screen where one protein was bait and the others prey. Some interactions could be lost if the role of the proteins (as prey and bait) are interchanged. It therefore will be wise to do further validations of these data to make them more conclusive.

a



b

NAC-α	NOT5	eIF4	Vector
NAC-α	NOT5	eIF4	Vector
NOT10	C2ORF29	UFPM	Vector
NOT10	C2ORF29	UFPM	
NOT1-C	CAF40	DNAj	
NOT1-C	CAF40	DNAj	
NOT2	CAF1	NAC-β	
NOT2	CAF1	NAC-β	

Figure 18: Y2H protein-protein interaction for candidate genes from large array. TbCAF40 was used as bait in pGKBT7 and screened against candidate genes as prey in pGADT7. Agar plate image (***left***) and sample annotation (***right***)

A microarray experiment generally translates binary interaction and enriches transient interaction [8]. Meanwhile there could be high chances of false positives as well; this would not be true for an entire data set. The decision of the prey and bait constructs can also be of key importance, since some proteins tend to be weak activators. Interaction between CAF 40 and the NAC has also been reported in *S. cerevisaie*[148]. Failure to identify an interaction between CAF 40 and NOT1 could depend on several factors including misfolding and low expression.

3.3.4 Affinity Measurements

Microscale thermophoresis (MS) was used to measure the binding kinetics between some protein-RNA pairs. Bovine serum albumin and TbAUK1 were used as negative controls. The binding of ZC3H11 and another hypothetical protein (Tb927.7.4660) with the classical ARE were also measured. The binding reaction was performed using the well characterized classical ARE RNA conjugated at its 5` end with Cy5 as a probe. The dissociation constant (Kd) was for ZC3H11 between 10-14 nM (**Fig. 19**). BSA and TbAUK1 both gave K_d values beyond 20,000 nM. The K_d values for ZC3H11 and the hypothetical protein correlated well with the signal strength on the array. Although this correlation may not be strong enough, it however indicate that the signal intensity observed could be indicative of the affinity between the interacting species. This must however be confirmed for a reasonable number of proteins.

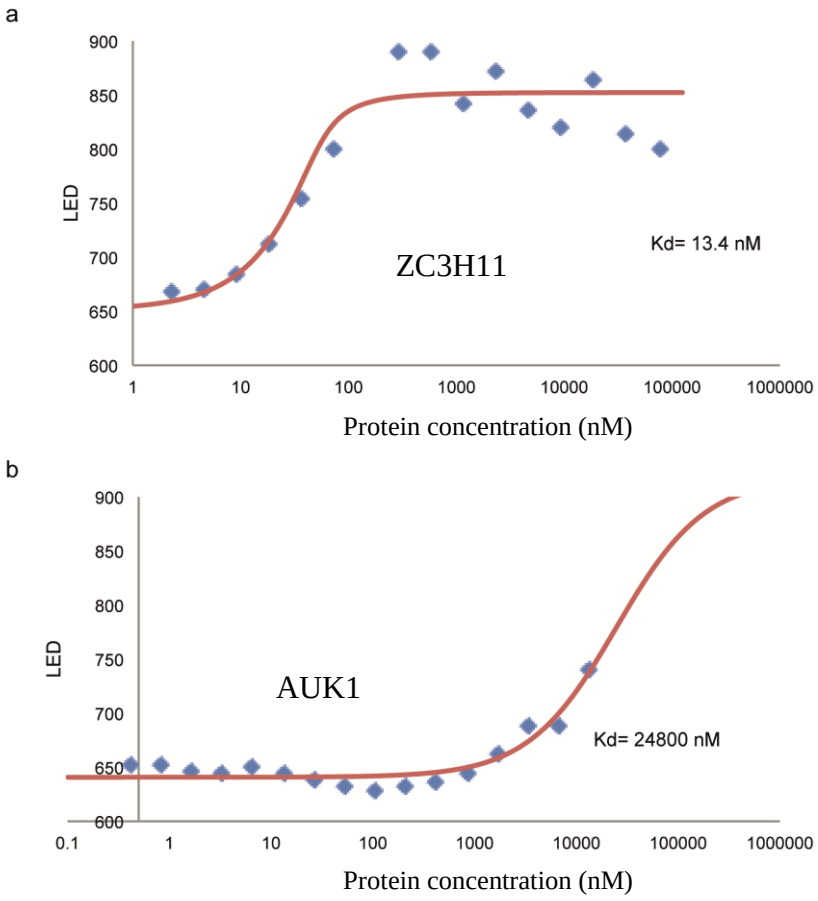


Figure 19: Kinetic results of protein-RNA interactions. (**a**) Plot of MST data for ZC3H11 binding to the classical ARE. (**b**) Plot of MST data for TbAUK1 interaction with the classical ARE. Majority of the data points are horizontal and further extrapolation is approximative for AUK1.

4 Conclusion

We set out to produce functional protein microarrays from cDNA products and evaluate the usability of such arrays in a model organism (*Trypanosoma brucei*). We have optimized and produced a protein microarray from a cDNA library using the multiple spotting technique. For optimal performance, primers must be designed to minimize the distance between the ribosome binding site and the translation initiation codon as much as possible. When open reading frames are used, the addition of an N-terminal 6×His tag can greatly enhance protein expression. Syproruby protein staining dye can be reliably used to detect protein expression when it is not possible to fuse expression construct with epitope tags for detection. Purification of PCR generated expression template is not needed and protein expression across the array can be controlled by controlling the incubation temperatures and time. It is possible to generate expression constructs directly from gDNA with the advantage of generating template flexibility and it will be advisable to normalize cDNA libraries to avoid redundancy of overrepresented transcripts. Arrays generated from gDNA-and cDNA-derived templates can be used for functional applications with great reliability.

We have shown for example, that such arrays could be useful for protein-protein interaction, protein-RNA interactions, identification of Kinase substrates and antibody selection. It is quite possible, that the scope of application of such arrays be wider if optimized. As with other protein microarray platforms, there is no universally suitable platform for MIST-based array. However, when signal-to-noise ratio is fundamental factor to consider, epoxysilane substrates perform better in this light. Streptavidin and Nickel chelate slides show good spot morphology compensating for high background signal. Some data obtained from our microarrays have been validated for protein-RNA interaction and protein-protein interaction to an extent. Combining cost efficiency, reliability, robustness and ease of use, this tool will be a plus in the arsenal of proteomic tools, and will bring protein microarray technology to almost everyone interested. Using cDNA products as template source offer the advantage of not only being cost effective, but important when studying disease conditions. It also offers the chance of screening rare and unknown transcripts. It is therefore most important for screening cellular proteomes in well-defined conditions. Nonetheless, some templates may therefore be absent if not transcribed during a given situation or lost during template preparation. Furthermore, it is not possible to confirm if any full-length protein is expressed. It is

therefore important that procedure for cDNA synthesis be well optimized. Taken together, the MIST is a reliable technology for the development of functional protein *in situ* arrays. Protein microarrays made by the MIST are highly versatile and of high quality, capable of generating biologically meaningful data.

REFERENCES

1. Sun, H., G.Y. Chen, and S.Q. Yao, *Recent advances in microarray technologies for proteomics.* Chem Biol, 2013. **20**(5): p. 685-99.
2. Pandey, A. and M. Mann, *Proteomics to study genes and genomes.* Nature, 2000. **405**(6788): p. 837-46.
3. Pflieger, D., *et al.*, *Linking the proteins--elucidation of proteome-scale networks using mass spectrometry.* Mass Spectrom Rev, 2011. **30**(2): p. 268-97.
4. Aebersold, R. and B.F. Cravatt, *Proteomics--advances, applications and the challenges that remain.* Trends Biotechnol, 2002. **20**(12 Suppl): p. S1-2.
5. Zhu, H. and J. Qian, *Applications of functional protein microarrays in basic and clinical research.* Adv Genet, 2012. **79**: p. 123-55.
6. Ekins, R. and F. Chu, *Multianalyte microspot immunoassay. The microanalytical 'compact disk' of the future.* Ann Biol Clin (Paris), 1992. **50**(5): p. 337-53.
7. Lee, Y., *et al.*, *ProteoChip: a highly sensitive protein microarray prepared by a novel method of protein immobilization for application of protein-protein interaction studies.* Proteomics, 2003. **3**(12): p. 2289-304.
8. Fuentes, M., *et al.*, *Solid phase proteomics: Dramatic reinforcement of very weak protein-protein interactions.* J Chromatograph B, 2007. **849**(1-2): p. 243-50.
9. Huang, Y.P., *et al.*, *Global tumor protein p53/p63 interactome Making a case for cisplatin chemoresistance.* Cell Cycle, 2012. **11**(12): p. 2367-79.
10. Lilburn, T.G., *et al.*, *Protease-associated cellular networks in malaria parasite Plasmodium falciparum.* BMC Genomics, 2011. **12**.
11. Chen, J.X., *et al.*, *A Protein Microarray for the Rapid Screening of Patients Suspected of Infection with Various Food-Borne Helminthiases.* PLoS Negl Trop D, 2012. **6**(11): p. e1899.
12. Mok, J., H. Im, and M. Snyder, *Global identification of protein kinase substrates by protein microarray analysis.* Nat Protoc, 2009. **4**(12): p. 1820-7.
13. Oh, Y.H., *et al.*, *Chip-based analysis of SUMO (small ubiquitin-like modifier) conjugation to a target protein.* Biosens Bioelectron, 2007. **22**(7): p. 1260-7.
14. Scherrer, T., *et al.*, *A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes.* PLoS One, 2010. **5**(11): p. e15499.
15. Margarit, I., *et al.*, *Capturing host-pathogen interactions by protein microarrays: identification of novel streptococcal proteins binding to human fibronectin, fibrinogen, and C4BP.* FASEB J, 2009. **23**(9): p. 3100-12.
16. Robinson, W.H., *et al.*, *Autoantigen microarrays for multiplex characterization of autoantibody responses.* Nat Med, 2002. **8**(3): p. 295-301.
17. Hu, S., *et al.*, *A protein chip approach for high-throughput antigen identification and characterization.* Proteomics, 2007. **7**(13): p. 2151-61.
18. Zhu, J., *et al.*, *Protein array identification of substrates of the Epstein-Barr virus protein kinase BGLF4.* J Virol, 2009. **83**(10): p. 5219-31.

19. Zhu, H. and J. Qian, *Applications of Functional Protein Microarrays in Basic and Clinical Research.* Adv Genet, Vol 79, 2012. **79**: p. 123-55.

20. Gelperin, D.M., *et al*., *Biochemical and genetic analysis of the yeast proteome with a movable ORF collection.* Gen Dev, 2005. **19**(23): p. 2816-26.

21. Chen, C.S., *et al*., *A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli.* Nat Methods, 2008. **5**(1): p. 69-74.

22. Zhu, H., *et al*., *Global analysis of protein activities using proteome chips.* Science, 2001. **293**(5537): p. 2101-5.

23. Allen, S.V. and E.S. Miller, *RNA-binding properties of in vitro expressed histidine-tagged RB69 RegA translational repressor protein.* Anal Biochem, 1999. **269**(1): p. 32-7.

24. Murthy, T.V.S., *et al*., *Bacterial cell-free system for high-throughput protein expression and a comparative analysis of Escherichia coli cell-free and whole cell expression systems.* Protein Express Purif, 2004. **36**(2): p. 217-25.

25. Goshima, N., *et al*., *Human protein factory for converting the transcriptome into an in vitro-expressed proteome.* Nat Methods, 2008. **5**(12): p. 1011-17.

26. Crompton, P.D., *et al*., *A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray.* Proc Natl Acad Sci USA, 2010. **107**(15): p. 6958-63.

27. He, M.Y. and M.J. Taussig, *Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method).* Nucleic Acids Res, 2001. **29**(15): p.e73-3.

28. Ramachandran, N., *et al*., *Self-assembling protein microarrays.* Science, 2004. **305**(5680): p. 86-90.

29. He, M., *et al*., *Printing protein arrays from DNA arrays.* Nat Methods, 2008. **5**(2): p. 175-7.

30. Tao, S.C. and H. Zhu, *Protein chip fabrication by capture of nascent polypeptides.* Nat Biotechnol, 2006. **24**(10): p. 1253-4.

31. Wang, J., *et al*., *A versatile protein microarray platform enabling antibody profiling against denatured proteins.* Proteomics Clin Appl, 2013. 7(5-6): p. 378-83.

32. Angenendt, P., *et al*., *Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products.* Mol Cell Proteomics, 2006. **5**(9): p. 1658-66.

33. Ramachandran, N., *et al*., *Next-generation high-density self-assembling functional protein arrays.* Nat Methods, 2008. **5**(6): p. 535-8.

34. van't Veer, L.J. and R. Bernards, *Enabling personalized cancer medicine through analysis of gene-expression patterns.* Nature, 2008. **452**(7187): p. 564-70.

35. Qu, H.Q., *et al*., *Genome-wide profiling using single-nucleotide polymorphism arrays identifies novel chromosomal imbalances in pediatric glioblastomas.* Neuro Oncol, 2010. **12**(2): p. 153-63.

36. Ambra, R., *et al*., *Transcriptome analysis of human primary endothelial cells (HUVEC) from umbilical cords of gestational diabetic mothers reveals candidate*

sites for an epigenetic modulation of specific gene expression. Genomics, 2014. **103**(5-6): p. 337-48

37. Harris, C. and N. Ghaffari, *Biomarker discovery across annotated and unannotated microarray datasets using semi-supervised learning.* BMC Genomics, 2008. **9**.
38. Zhu, J.Y., Y. Sun, and Z.Y. Wang, *Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq).* Methods Mol Biol, 2012. **876**: p. 173-88.
39. Gry, M., *et al*., *Correlations between RNA and protein expression profiles in 23 human cell lines.* BMC Genomics, 2009. **10**.
40. Stiffler, M.A., *et al*., *Uncovering quantitative protein interaction networks for mouse PDZ domains using protein microarrays.* J Am Chem Soc, 2006. **128**(17): p. 5913-5922.
41. Kwon, J.A., *et al*., *High diagnostic accuracy of antigen microarray for sensitive detection of hepatitis C virus infection.* Clin Chem, 2008. **54**(2): p. 424-428.
42. Lee, Y., *et al*., *ProteoChip: A highly sensitive protein microarray prepared by a novel method of protein, immobilization for application of protein-protein interaction studies.* Proteomics, 2003. **3**(12): p. 2289-2304.
43. Miernyk, J.A. and J.J. Thelen, *Biochemical approaches for discovering protein-protein interactions.* Plant J, 2008. **53**(4): p. 597-609.
44. Dias-Neto, E., *et al*., *Next-Generation Phage Display: Integrating and Comparing Available Molecular Tools to Enable Cost-Effective High-Throughput Analysis.* PLoS One, 2009. **4**(12): p.e8338.
45. Bruckner, A., *et al*., *Yeast Two-Hybrid, a Powerful Tool for Systems Biology.* Int J Mol Sci, 2009. **10**(6): p. 2763-88.
46. Ivanov, S.S., *et al*., *Antibodies immobilized as arrays to profile protein post-translational modifications in mammalian cells.* Mol Cell Proteomics, 2004. **3**(8): p. 788-95.
47. Sanchez-Carbayo, M., *Antibody microarrays as tools for biomarker discovery.* Methods Mol Biol, 2011. **785**: p. 159-82.
48. Puig-Costa, M., *et al*., *Discovery and validation of an INflammatory PROtein-driven GAstric cancer Signature (INPROGAS) using antibody microarray-based oncoproteomics.* Oncotarget, 2014. **5**(7): p. 1942-54.
49. Sanchez-Carbayo, M., *Antibody array-based technologies for cancer protein profiling and functional proteomic analyses using serum and tissue specimens.* Tumour Biol, 2010. **31**(2): p. 103-12.
50. Voshol, H., *et al*., *Antibody-based proteomics: analysis of signaling networks using reverse protein arrays.* FEBS J, 2009. **276**(23): p. 6871-9.
51. Borrebaeck, C.A.K. and C. Wingren, *Design of high-density antibody microarrays for disease proteomics: Key technological issues.* J Proteomics, 2009. **72**(6): p. 928-35.

52. He, M. and M.J. Taussig, *Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method).* Nucleic Acids Res, 2001. **29**(15): p. e73-3.

53. Ramachandran, N., *et al.*, *Tracking humoral responses using self assembling protein microarrays.* Proteomics Clin Appl, 2008. **2**(10-11): p. 1518-1527.

54. Anderson, K.S., *et al.*, *Application of protein microarrays for multiplexed detection of antibodies to tumor antigens in breast cancer.* J Proteome Res, 2008. **7**(4): p. 1490-1499.

55. Miersch, S., *et al.*, *Serological autoantibody profiling of type 1 diabetes by protein arrays.* J Proteomics, 2013. **94**: p. 486-96.

56. Montor, W.R., *et al.*, *Genome-wide study of Pseudomonas aeruginosa outer membrane protein immunogenicity using self-assembling protein microarrays.* Infect Immun, 2009. **77**(11): p. 4877-86.

57. Wallstrom, G., K.S. Anderson, and J. LaBaer, *Biomarker discovery for heterogeneous diseases.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(5): p. 747-55.

58. Stoevesandt, O., M. He, and M.J. Taussig, *Protein microarrays printed from DNA microarrays.* Methods Mol Biol, 2011. **671**: p. 95-106.

59. Schmidt, R., *et al.*, *Optimised 'on demand' protein arraying from DNA by cell free expression with the 'DNA to Protein Array' (DAPA) technology.* J Proteomics, 2013. **88**: p. 141-8.

60. Chattopadhaya, S., L.P. Tan, and S.Q. Yao, *Strategies for site-specific protein biotinylation using in vitro, in vivo and cell-free systems: toward functional protein arrays.* Nat Protoc, 2006. **1**(5): p. 2386-98.

61. He, M. and M.J. Taussig, *Rapid discovery of protein interactions by cell-free protein technologies.* Biochem Soc T, 2007. **35**: p. 962-65.

62. Lim, M. and K.J. Rothschild, *Photocleavage-based affinity purification and printing of cell-free expressed proteins: application to proteome microarrays.* Anal Biochem, 2008. **383**(1): p. 103-15.

63. Goshima, N., *et al.*, *Human protein factory for converting the transcriptome into an in vitro-expressed proteome.* Nat Methods, 2008. **5**(12): p. 1011-7.

64. Anderson, K.S., *et al.*, *Serum antibodies to the HPV16 proteome as biomarkers for head and neck cancer.* Brit J Cancer, 2011. **104**(12): p. 1896-905.

65. Ceroni, A., *et al.*, *Systematic analysis of the IgG antibody immune response against varicella zoster virus (VZV) using a self-assembled protein microarray.* Mol Biosyst, 2010. **6**(9): p. 1604-10.

66. Tom, I., *et al.*, *Protein microarrays for identification of novel extracellular protein-protein interactions.* Curr Protoc Protein Sci, 2013. **Chapter 27**: p. Unit 27 3.

67. Predki, P.F., *Functional protein microarrays: ripe for discovery.* Curr Opin Chem Biol, 2004. **8**(1): p. 8-13.

68. Chandra, H. and S. Srivastava, *Cell-free synthesis-based protein microarrays and their applications.* Proteomics, 2010. **10**(4): p. 717-30.

69. Fang, Y., J. Lahiri, and L. Picard, *G protein-coupled receptor microarrays for drug discovery.* Drug Discov Today, 2003. **8**(16): p. 755-61.

70. Hurst, R., *et al*., *Protein-protein interaction studies on protein arrays: effect of detection strategies on signal-to-background ratios.* Anal Biochem, 2009. **392**(1): p. 45-53.

71. Mei, Q., *et al*., *Toxin detection by a miniaturized in vitro protein expression array.* Anal Chem, 2005. **77**(17): p. 5494-500.

72. Rathod, V., *et al*., *Human pregnane X receptor: a novel target for anticancer drug development.* Drug Discov Today, 2014. **19**(1): p. 63-70.

73. Schwarz, D., V. Dotsch, and F. Bernhard, *Production of membrane proteins using cell-free expression systems.* Proteomics, 2008. **8**(19): p. 3933-3946.

74. Vaquerizas, J.M., *et al*., *A census of human transcription factors: function, expression and evolution.* Nat Rev Genet, 2009. **10**(4): p. 252-263.

75. Hall, D.A., *et al*., *Regulation of gene expression by a metabolic enzyme.* Science, 2004. **306**(5695): p. 482-84.

76. Ho, S.W., *et al*., *Linking DNA-binding proteins to their recognition sequences by using protein microarrays.* Proc Natl Acad Sci USA, 2006. **103**(26): p. 9940-5.

77. Hu, S.H., *et al*., *Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling.* Cell, 2009. **139**(3): p. 610-22.

78. Bielli, P., *et al*., *The RNA-binding protein Sam68 is a multifunctional player in human cancer.* Endocr Relat Cancer, 2011. **18**(4): p. R91-R102.

79. Abdelmohsen, K. and M. Gorospe, *Posttranscriptional regulation of cancer traits by HuR.* Wiley Interdiscip Rev RNA, 2010. **1**(2): p. 214-29.

80. Tyers, M. and M. Mann, *From genomics to proteomics.* Nature, 2003. **422**(6928): p. 193-197.

81. Kung, L.A., *et al*., *Global analysis of the glycoproteome in Saccharomyces cerevisiae reveals new roles for protein glycosylation in eukaryotes.* Mol Syst Biol, 2009. **5**: p. 308.

82. Habelhah, H., *Emerging complexity of protein ubiquitination in the NF-kappaB pathway.* Genes Cancer, 2010. **1**(7): p. 735-47.

83. Lin, Y.Y., *et al*., *Protein Acetylation Microarray Reveals that NuA4 Controls Key Metabolic Target Regulating Gluconeogenesis.* Cell, 2009. **136**(6): p. 1073-84.

84. Foster, M.W., M.T. Forrester, and J.S. Stamler, *A protein microarray-based analysis of S-nitrosylation.* Proc Natl Acad Sci USA, 2009. **106**(45): p. 18948-53.

85. Li, R., *et al*., *Conserved herpesvirus kinases target the DNA damage response pathway and TIP60 histone acetyltransferase to promote virus replication.* Cell Host Microbe, 2011. **10**(4): p. 390-400.

86. Ramachandran, N., S. Srivastava, and J. Labaer, *Applications of protein microarrays for biomarker discovery.* Proteomics Clin Appl, 2008. **2**(10-11): p. 1444-59.

87. Wong, J., *et al*., *Rapid detection of antibodies in sera using multiplexed self-assembling bead arrays.* J Immunol Methods, 2009. **350**(1-2): p. 171-82.

88. Doolan, D.L., *et al*., *Profiling humoral immune responses to P. falciparum infection with protein microarrays.* Proteomics, 2008. **8**(22): p. 4680-94.

89. Beare, P.A., *et al*., *Candidate Antigens for Q Fever Serodiagnosis Revealed by Immunoscreening of a Coxiella burnetii Protein Microarray.* Clin Vaccine Immunol, 2008. **15**(12): p. 1771-9.

90. Schreiber, S.L., *Organic synthesis toward small-molecule probes and drugs.* Proc Natl Acad Sci USA, 2011. **108**(17): p. 6699-702.

91. Meng, L., D. Mattoon, and P. Predki, *Small molecule protein interaction profiling with functional protein microarrays.* Methods Mol Biol, 2009. **572**: p. 177-88.

92. Huang, J., *et al*., *Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips.* Proc Natl Acad Sci USA, 2004. **101**(47): p. 16594-9.

93. Huang, J., *et al*., *Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips.* Proc Natl Acad Sci USA 2004. **101**(47): p. 16594-9.

94. Guilleaume, B., *et al*., *Systematic comparison of surface coatings for protein microarrays.* Proteomics, 2005. **5**(18): p. 4705-12.

95. Schmidt, R., *et al*., *Single-molecule detection on a protein-array assay platform for the exposure of a tuberculosis antigen.* J Proteome Res, 2011. **10**(3): p. 1316-22.

96. Hu, S., *et al*., *Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling.* Cell, 2009. **139**(3): p. 610-22.

97. Lu, J.Y., *et al*., *Acetylation of yeast AMPK controls intrinsic aging independently of caloric restriction.* Cell, 2011. **146**(6): p. 969-79.

98. Tao, S.C., *et al*., *Lectin microarrays identify cell-specific and functionally significant cell surface glycan markers.* Glycobiology, 2008. **18**(10): p. 761-9.

99. Peluso, P., *et al*., *Optimizing antibody immobilization strategies for the construction of protein microarrays.* Anal Biochem, 2003. **312**(2): p. 113-24.

100. Zajac, A., *et al*., *Protein microarrays and quantum dot probes for early cancer detection.* Colloids Surf B Biointerfaces, 2007. **58**(2): p. 309-14.

101. Huang, R.P., *Detection of multiple proteins in an antibody-based protein microarray system.* J Immunol Methods, 2001. **255**(1-2): p. 1-13.

102. Zhou, H., *et al*., *Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements.* Genome Biol, 2004. **5**(4): p. R28.

103. Varnum, S.M., R.L. Woodbury, and R.C. Zangar, *A protein microarray ELISA for screening biological fluids.* Methods Mol Biol, 2004. **264**: p. 161-72.

104. Biesiadecki, B.J. and J.P. Jin, *A High-Throughput Solid-Phase Microplate Protein-Binding Assay to Investigate Interactions between Myofilament Proteins.* J Biomed Biotechnol, 2011: p. 421701.

105. Nelson, B.P., *et al*., *Near-infrared surface plasmon resonance measurements of ultrathin films. 1. Angle shift and SPR imaging experiments.* Anal Chem, 1999. **71**(18): p. 3928-34.

106. Wang, Z.H. and G. Jin, *A label-free multisensing immunosensor based on imaging ellipsometry.* Anal Chem, 2003. **75**(22): p. 6119-23.

107. Piehler, J., *et al*., *Label-free monitoring of DNA-ligand interactions.* Anal Biochem, 1997. **249**(1): p. 94-102.

108. Liu, S., *et al*., *Label-free, real-time detection of the dynamic processes of protein degradation using oblique-incidence reflectivity difference method.* Appl Phys Lett, 2014. **104**(16): p. 163701.

109. Evans-Nguyen, K.M., *et al*., *Protein arrays on patterned porous gold substrates interrogated with mass spectrometry: Detection of peptides in plasma.* Anal Chem, 2008. **80**(5): p. 1448-58.

110. Yan, H., *et al*., *DNA-templated self-assembly of protein arrays and highly conductive nanowires.* Science, 2003. **301**(5641): p. 1882-84.

111. Pounds, S. and S.W. Morris, *Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values.* Bioinformatics, 2003. **19**(10): p. 1236-42.

112. Clayton, C., *The regulation of trypanosome gene expression by RNA-binding proteins.* PLoS Pathog, 2013. **9**(11): p. e1003680.

113. Michaeli, S., *Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome.* Future Microbiol, 2011. **6**(4): p. 459-74.

114. Clayton, C.E., *Life without transcriptional control? From fly to man and back again (vol 21, pg 1881, 2002).* EMBO J, 2002. **21**(14): p. 3917-17.

115. Berthier, D., *et al*., *Serial analysis of gene expression (SAGE) in bovine trypanotolerance: preliminary results.* Genet Sel Evol, 2003. **35 Suppl 1**: p. S35-47.

116. Wurst, M., *et al*., *An RNAi screen of the RRM-domain proteins of Trypanosoma brucei.* Mol Biochem Parasitol, 2009. **163**(1): p. 61-5.

117. Mani, J., *et al*., *Alba-domain proteins of Trypanosoma brucei are cytoplasmic RNA-binding proteins that interact with the translation machinery.* PLoS One, 2011. **6**(7): p. e22463.

118. Abbasi, N., Y.I. Park, and S.B. Choi, *Pumilio Puf domain RNA-binding proteins in Arabidopsis.* Plant Signal Behav, 2011. **6**(3): p. 364-8.

119. Archer, S.K., *et al*., *Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle.* PLoS Pathog, 2009. **5**(8): p. e1000565.

120. Stern, M.Z., *et al*., *Multiple roles for polypyrimidine tract binding (PTB) proteins in trypanosome RNA metabolism.* RNA, 2009. **15**(4): p. 648-65.

121. Manful, T., A. Fadda, and C. Clayton, *The role of the 5 '-3 ' exoribonuclease XRNA in transcriptome-wide mRNA degradation.* RNA, 2011. **17**(11): p. 2039-47.

122. Winkler, G.S., *et al*., *An altered-specificity ubiquitin-conjugating enzyme/ubiquitin-protein ligase pair.* J Mol Biol, 2004. **337**(1): p. 157-65.

123. Inada, T. and S. Makino, *Novel roles of the multi-functional CCR4-NOT complex in post-transcriptional regulation.* Front Genet, 2014. **5**: p. 135.

124. Halter, D., M.A. Collart, and O.O. Panasenko, *The Not4 E3 ligase and CCR4 deadenylase play distinct roles in protein quality control.* PLoS One, 2014. **9**(1): p. e86218.

125. Matsuda, R., *et al.*, *Protein quality control systems associated with no-go and nonstop mRNA surveillance in yeast.* Genes Cells, 2014. **19**(1): p. 1-12.

126. Gruber, A.R., *et al.*, *The Vienna RNA websuite.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W70-4.

127. Shen, S., *et al.*, *In vivo epitope tagging of Trypanosoma brucei genes using a one step PCR-based strategy.* Mol Biochem Parasitol, 2001. **113**(1): p. 171-3.

128. Vimberg, V., *et al.*, *Translation initiation region sequence preferences in Escherichia coli.* BMC Mol Biol, 2007. **8:** p. 100.

129. Nilsson, D., *et al.*, *Spliced Leader Trapping Reveals Widespread Alternative Splicing Patterns in the Highly Dynamic Transcriptome of Trypanosoma brucei.* PLoS Pathog, 2010. **6**(8): p. e 1001037.

130. Larsson, M., *et al.*, *High-throughput protein expression of cDNA products as a tool in functional genomics.* J Biotechnol, 2000. **80**(2): p. 143-57.

131. Angenendt, P., *et al.*, *Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products.* Mol Cell Proteomics, 2006. **5**(9): p. 1658-66.

132. Chakravarti, B., *et al.*, *Comparison of SYPRO Ruby and Deep Purple using commonly available UV transilluminator: Wide-scale application in proteomic research.* J Proteome Res, 2008. **7**(7): p. 2797-802.

133. Reimann, B., *et al.*, *Initial characterization of the nascent polypeptide-associated complex in yeast.* Yeast, 1999. **15**(5): p. 397-407.

134. del Alamo, M., *et al.*, *Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes.* PLoS Biol, 2011. **9**(7): p. e1001100.

135. Tarassov, K., *et al.*, *An in vivo map of the yeast protein interactome.* Science, 2008. **320**(5882): p. 1465-70.

136. Hata, H., *et al.*, *Dhh1p, a putative RNA helicase, associates with the general transcription factors Pop2p and Ccr4p from Saccharomyces cerevisiae.* Genetics, 1998. **148**(2): p. 571-9.

137. Farber, V., *et al.*, *Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs.* Nucleic Acids Res, 2013. **41**(2): p. 1211-22.

138. Kirstein-Miles, J., *et al.*, *The nascent polypeptide-associated complex is a key regulator of proteostasis.* EMBO J, 2013. **32**(10): p. 1451-68.

139. Coller, J.M., *et al.*, *The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes.* RNA, 2001. **7**(12): p. 1717-27.

140. Palfi, Z., *et al*., *U1 small nuclear RNP from Trypanosoma brucei: a minimal U1 snRNA with unusual protein components.* Nucleic Acids Res, 2005. **33**(8): p. 2493-503.

141. McConnell, T.S., R.P. Lokken, and J.A. Steitz, *Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA.* RNA, 2003. **9**(2): p. 193-201.

142. Droll, D., *et al*., *Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein.* PLoS Pathog, 2013. **9**(4): p. e1003286.

143. Kramer, S. and M. Carrington, *Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids.* Trends Parasitol, 2011. **27**(1): p. 23-30.

144. Estevez, A.M., *The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes.* Nucleic Acids Res, 2008. **36**(14): p. 4573-86.

145. Mayho, M., *et al*., *Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements.* Nucleic Acids Res, 2006. **34**(18): p. 5312-24.

146. Quijada, L., *et al*., *Expression of the human RNA-binding protein HuR in Trypanosoma brucei increases the abundance of mRNAs containing AU-rich regulatory elements.* Nucleic Acids Res, 2002. **30**(20): p. 4414-24.

147. Bringaud, F. and T. Baltz, *Differential regulation of two distinct families of glucose transporter genes in Trypanosoma brucei.* Mol Cell Biol, 1993. **13**(2): p. 1146-54.

148. Panasenko, O., *et al*., *The yeast Ccr4-Not complex controls ubiquitination of the nascent-associated polypeptide (NAC-EGD) complex.* J Biol Chem, 2006. **281**(42): p. 31389-98.

Part II:
Blood-Born miRNA/mRNA Signatures in Human African Trypanosomiasis

1 Introduction

African Trypanosomiases (AT) or sleeping sickness are diseases of humans and animals transmitted by *tse-tse* flies (**Fig. 1**). Human African Trypanosomiases (HAT) and Animal African Trypanosomiases (AAT) are caused by unicellular protozoa parasites called trypanosomes. *Trypanosoma brucei gambiense (Tbg)* and *Trypanosoma brucei rhodensiense (Tbr)* are the major disease-causing agents in humans and are also known to infect domestic and wild animals. This leads to a possible generation of an animal reservoir of HAT[1]. Recently, reports indicate that domestic animals such as pigs could be potential reservoirs for both human and animal-infective trypanosomes[2]. Such findings are very frustrating, as they actually render the eradication of the disease inconceivable, and therefore minimize all efforts toward maintaining a low transmission rate.

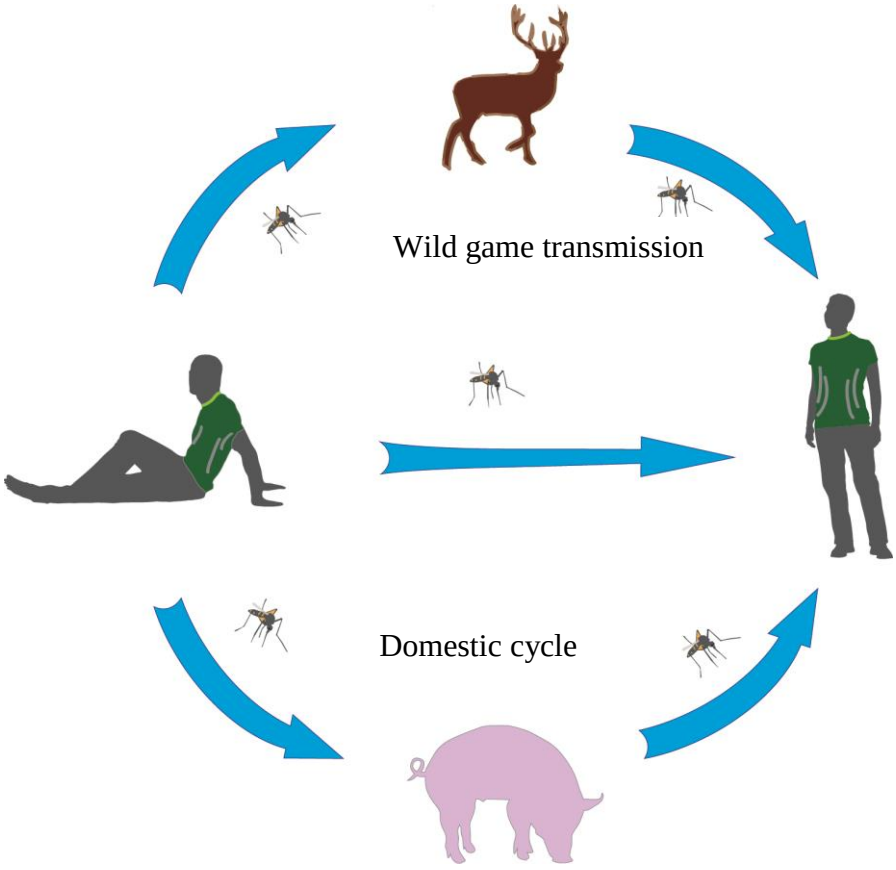


Figure 1: Transmission cycle of Human African Trypanosomiases (HAT) and Animal African Trypanosomiases. A possible scheme for the conservation of Animal reservoire

According to recent estimates, some 70 million people are at risk of contracting HAT and about one hundred and seventy-five thousand cases were reported between

2000 and 2009, with an annual infection rate of 10,000[3, 4]. Clinically, the disease is divided into two stages. An initial hemolymphatic stage; with no specific symptoms which progresses to a late encephalitic stage involving the central nervous system[5]. The progression profile from one stage to the other greatly varies from patient to patient, but also depends upon the infecting Trypanosome strain. Classically, *Tbg* is known to cause a chronic disease that is geolocalized within West and Central Africa; meanwhile *Tbr* causes an acute infection in East Africa. HAT caused by *Tbr* is less prevalent than *Tbg* disease and some exceptions have been observed in the reported disease profiles[6, 7].

Diagnosis of *Tbg* infection is done via the card agglutination test for trypanosomiasis (CATT). CATT positive subjects are then subjected to the trypanolysis test and subsequent microscopic examination for confirmation and disease staging[4]. Both the CATT and the trypanolysis test all depend on circulating immunoglobulin against variant surface antigens of the trypanosomes[8]. Presently, there is no homologue of the CATT for *Tbr* infection, and staging in both cases involves examination of cerebrospinal fluid (CSF). This is a very invasive procedure and discourages most of those living in endemic areas to participate in mass field screening. Although parasite DNA amplification is possible, the facilities are generally not available, and it is not suited for field diagnosis[4, 9]. Disease staging is of particular importance, as this provides the basis for the choice of therapeutic regiments, which is done in a case-by-case manner. Meanwhile, some molecular markers have been identified for staging; they all also rely on CSF. Moreover, patient classification differ from one country to the other and hence treatment outcomes[10-12].

HAT patients constitute a very heterogenous population. This heterogeneity makes the disease classification more complex. In most foci, several serologically positive, but parasitological negative subjects are constantly encountered. The possibility of false positives due to non-specific agglutination in the CATT assay cannot account for all cases[13, 14]. One possible explanation to such observations could be a low parasite load at the time of examination. To circumvent this problem, the mini anion exchange chromatography column has been introduced for parasite concentration[15]. Although this enhances the sensitivity of parasitological diagnosis, it does not make it 100% sensitive, and some patients are left out or wrongly assigned. Analysis of one of such patient population has led to the identification of three patient groups: the first group of these serologically positive, but parasitological negative

subjects, were in the very early stage of the infection and later developed the disease. A second group, maintain high serological titers, but never develop the disease (asymptomatic carriers), and the third group, who become CATT negative after some time, might have self-cured[7].

Many factors, may actually account for this heterogeneity, among which; the host and parasite genetic diversity, host immune responses, and multiple infections[16]. Regarding host genetic diversity, it is possible, that the phenomenon of trypanotolerance, which has long been reported in cattle, also occur in humans, rather than the self-cure hypothesis[17, 18]. The heterogeneity does not only threaten control strategies, but also complicate the decision of whether or not a patient should be treated, and even so, if stage I or stage II treatment should be administered. In the light of this, different recommendations have been adopted by control programs in different countries[19, 20]. An ideal situation however would be to develop reliable less invasive diagnostic standards that can guide therapeutic decision and encourage patient enrollment. As with most communicable and non-communicable diseases, molecular diagnosis is fast becoming a standard and many molecular biomarkers have been reported for several diseases[21, 22]. Several biomolecules have reportedly been used as biomarkers in some infectious and noninfectious diseases, amongst which miRNAs[23].

MicroRNAs (miRNAs) are small RNAs of about 21-nt which bind to the 3′ UTR of mRNAs and play key roles in the control of mRNA stability and gene expression control[24, 25]. miRNAs play diverse roles in mammals and are involved in several cellular processes ranging from immune responses, differentiation, proliferation and apoptosis, development and feedback loops[26]. Dysregulation of miRNA expression and functions have been associated with several diseases, and there have been several attempts to investigate the correlation of circulating miRNA profiles with disease progression, prognosis and staging[23, 27-29]. Besides, the potentials of miRNAs as biomarkers have extensively been studied in metabolic and genetic diseases such as cancer and diabetes. However, less attention has been attributed to the role of miRNA in infectious diseases. Most of the few cases addressed, have been focused on experimental infections and no natural infections[30-33]. There are reports of modulation of host gene expression post infection in some infectious diseases. In mice for example, macrophage activation and cytokine responses have been observed following infection by *Trypanosoma brucei brucei*[34, 35]. It is therefore reasonable, to hypothesize, that infection by disease-causing trypanosomes can lead to variations in

the expression profiles of miRNAs. This could be as a result of the activation of immune responses and other biochemical processes that are controlled by genes, whose expression are regulated by miRNA. It is with this in mind, that this study was designed to analyze the expression profiles of circulating miRNA with a view of biomarker identification and immune response analysis.

Objectives

In the struggle for survival during *T. brucei* infection, the mammalian host has developed strategies to eliminate the parasite. Innate and active immune responses as well as immunosuppression are all processes that are activated following infection. It is possible, that the activation of such defense mechanisms and related pathways involve the expression of genes whose expression is regulated at the posttranscriptional level by miRNA. This can result in differences in the expression profiles of the regulating miRNAs

Aim

This work was designed to analyze the expression profiles of circulating miRNA and mRNA in the peripheral blood cells of patients infected by *Trypanosoma brucei gambiense*, in an attempt to identify less invasive disease biomarkers and stage-specific markers and to understand the mechanisms that underlie the outcome of *Tbg* infection.

1.1 Human African Trypanosomiases

1.1.1 Historical Background

The awareness of the two variants of HAT in many endemic areas dates far back. However, knowledge about the precise causative agent was first reported in the late 19th century by David Bruce, while working on a wasting disease of cattle called nagana[36]. It was then established, that trypanosomes could infect wild game animals and serve as a reservoir. In 1899, the parasite was named *Trypanosoma brucei*, and in 1902, the first patient with *Tbg* infection was identified. In 1903, the first evidence of trypanosomes in CSF and Blood of sleeping sickness patients was established, and in 1910, *Tbr* was described for the first time. During the early 1950, HAT was almost brought under control owing to vast vector/parasite control campaigns. However, due to

political unrest, war, socio-economic factors and game reservoir, resurgence was observed in the late 1970[37, 38].

1.1.2 Epidemiology

About 70 million people living on an estimated area of about 1.55 million km^2 in 200 different foci in 36 African countries are thought to be at risk[4, 10]. Approximately 82% of people are at risk of *Tbg* infection in West and central Africa, while less than 18% are at risk of the more acute *Tbr* infection in east Africa[4]. An annual infection rate of about 10.000 cases has been reported, with the Democratic Republic of Congo, Congo, Uganda, Cote d'Ivoire, Guinea, Nigeria and south Sudan having the highest populations at risk of *Tbg* infection. Meanwhile, Uganda, Tanzani and Kenya are by far the countries with the largest populations at risk of *Tbr*[3, 4].

1.1.3 Life Cycle of *Trypanosome brucei* and Infection

Trypanosomes are spindle shaped unicellular protozoa parasites of the Trypanosomatidae family. They measure about 20–30 by 1.3–3.5 µM with a single flagellum that projects from the posterior end of the parasite and runs along the cell membrane. The flagellum is attached by an undulating membrane and is associated with the kinetoplast[10]. During its entire life cycle in both the human and vector host, the parasite multiplies by binary fission (**Fig. 2**). Human infection starts with the injection of metacyclic forms of the parasite into a healthy subject, by an infected *tse-tse* fly while taking a blood meal. The parasites then multiply locally for a few days before invading the lymphatic system, blood and other tissues and organs such as the central nervous system causing the sleep disorder that is characteristic of the disease.

The parasites then continue to multiply and differentiate into two distinct forms (i) the long slender dividing form, (ii) and the short stumpy quiescent form. The latter is then taken up by a fly while feeding and can be transmitted to other subjects. While in humans, the parasite is covered with a dense coat of surface glycoprotein, preventing it from lysis by complement factors[39]. Immunoglobulin, especially the IgM class against the surface glycoprotein can help destroy the parasite. However, a fraction of the parasites is capable of escaping the host immune response, and generate a new surface antigen via a mechanism of antigenic variation. The *Trypanosoma brucei* genome encodes about 1,000 different genes of the Variant Surface Glycoprotein (VSG) that are

expressed in a mutually exclusive fashion[40-43]. Switching from one VSG to another, may account for the variations observed in parasite load during infection.

1.1.4 Clinical Features

In its clinical presentation, two distinct stages of HAT are recognized, namely: the early hemolymphatic stages where parasites are found in blood and the lymphatic system and the late encephalitic stage with CNS involvement. The distinction is more observable in *Tbg* infection, than in *Tbr* infection[38]. Conventionally, it is accepted, that *Tbg* infection leads to a chronic disease that can last for years before the eventual death of the patient if left untreated, meanwhile, *Tbr* causes a chronic disease that is fatal within weeks to months if untreated. However, there are several exceptions reported with chronic illness described for *Tbr* infection and fulminant illness for *Tbg* infection[10].

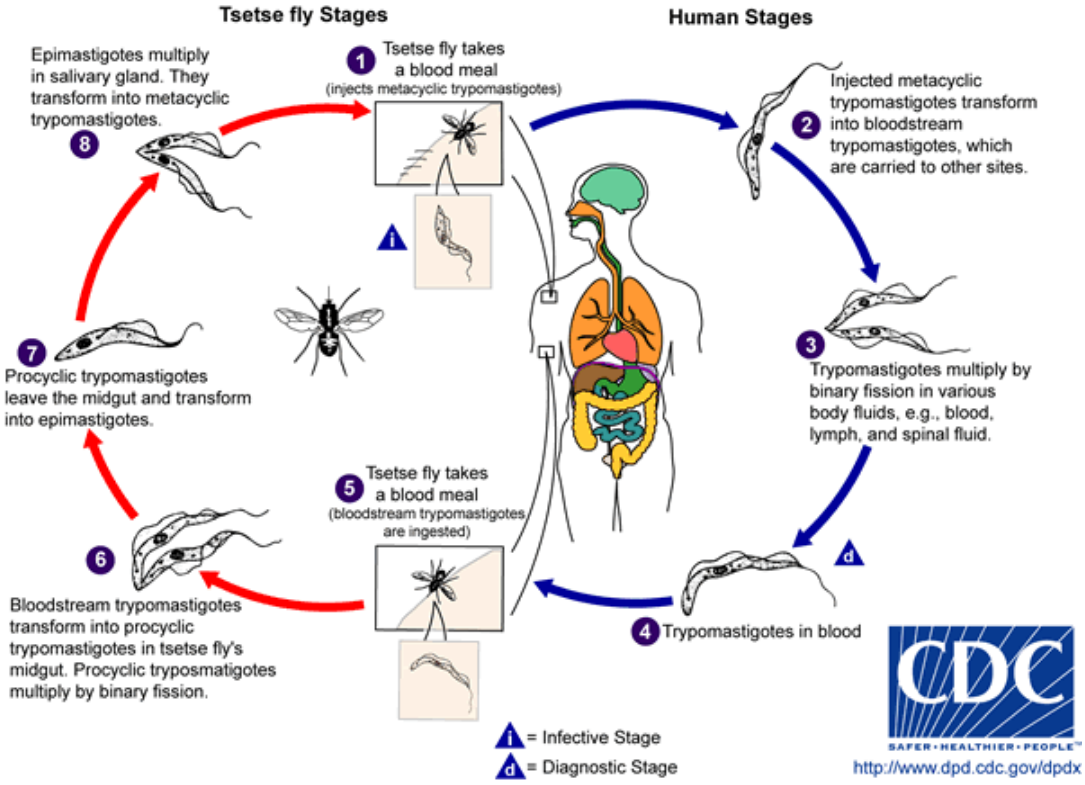


Figure 2: Representation of the life cycle of Trypanosoma brucei in the human and the tsetse fly. Alexander J. da Silva and Melanie Moser, Centers for Disease Control Public Health Image Library.

1.1.4. *T. brucei* gambiense Disease

After injection of the parasites, a chancre appears 5–15 days post infection. This will in most cases resolve without being noticed in most patients. Patients then starts

presenting intermittent nonspecific symptoms such as fever, fatigue, enlarge lymph nodes, headaches, arthralgia, and pruritus. Furthermore, in some cases a skin rash, splenomegaly and hepatomegaly are also some non-specific symptoms of the early stage of *Tbg* infection[44]. Following invasion of the central nervous system by trypanosomes and a consequential immune response, neuropschychiatric signs and symptoms starts appearing[45]. It may take months and in some case even years for the transition for first to second stage to be observed. The clinical features of the second stages are divided into groups including: motor, sensory, psychiatric, and sleep cycle disorder. Some observable mental disorders include: irritability, lassitude, headache, personality change, and psychosis. Kerandel's sign can also be noted and a reversal of the normal sleep-wake cycle, with daytime somnolence alternating with night-time insomnia. Weight loss, endocrine disorders (amenorrhea), and impotence are sometimes observed and if left untreated, patients ultimately die[46].

Following *Tbr* infection, an acute febrile illness begins 1–3 weeks post infection, with more pronounced, but non-specific symptoms including generalized lymphadenopathy keratitis and conjunctivitis. Practically, there is no clear boundary between first and second stage *Tbr* disease. Manifestations of CNS involvement may include drowsiness and tremor, congestive heart failure, arrhythmia, and pericardial effusion[47]. These conditions may kill the patients within the first six months post infection leaving no time for other *Tbr*-specific signs to be observed[48, 49].

1.1.5 Diagnosis of Sleeping Sickness

Diagnosis of sleeping sickness is a multi-step task that is performed routinely by control program during mass field screening and usually combine serological, parasitological and/molecular diagnosis. It generally involves screening, diagnosis confirmation, disease staging and enrollment for follow-up. During this process, active cases are detected and recruited into a surveillance scheme. As such, it is of prime importance, that field diagnosis, be simplified, sensitive and patient friendly. Field diagnosis of *Tbg* infection currently rely on the card agglutination test for trypanosomaisis (CATT/*Tbg*), which has been reported to be more sensitive than the cervical Lymph Node (CLN) palpitation and puncture[50]. A complementary serological test, the trypanolysis test has recently been evaluated, and is reported to complement the CATT and minimize false results[8]. Confirmation of serodiagnosis relies on parasitological detection of parasites in blood, lymph and cerebrospinal fluid.

Although successful in some cases, about 20% to 30% of infected subjects are missed out either as a result of low parasite load, and/or low sensitivity and experience[50].

Disease staging allow for the classification of patients in either of two major categories that are: first or hemolymphatic stage and the second or encephalitic stage, with the parasites crossing the blood-brain barrier. Without reliable tools for the detection of parasites in the brain, staging is often achieved by examination of CSF for the presence of trypanosomes, white cell count, and protein concentration. According to the world health organization recommendations, patients should be classified as being in second stage disease, if one of the following conditions are fulfilled: (i) raised white blood cell count (>5 cells/µl), (ii) presence of trypanosomes (iii), and increased protein content (>370 mg/l measured by dye-binding protein assay) [51, 52]. Despite these recommendations, control programs in different countries have adopted local measures depending on treatment outcomes[53-55].

1.1.5.1 Serological Diagnosis of Sleeping Sickness

Serological diagnosis is the first line diagnostic measure used during field screening. It generally aims at finding parasite-specific antigens or antibodies directed against parasite-specific antigens. In the latter case, one or more of three surface antigens (LiTat 1.3, 1.5, and 1.6) are used to search for circulating antibodies 3–6 weeks post infection[42]. As such, the CATT/Tbg, CATT/Latex and trypanolysis have been developed for the detection of circulating *Tbg*-specif antibodies. The specificity of the CATT/*Tbg*, and CATT/Latex *Tbg* are in the order of 87%–98% and 96%–99% respectively, with similar sensitivities (71%–100%)[56]. The reduced sensitivity of the CATT/*Tbg* can be explained in part, by the use of only one of the antigens (LiTat 1.3) as opposed to all three antigens in CATT/Latex.

The immune trypanolysis test which uses all three antigens was recently re-evaluated and reported to show a close to 100% specificity and sensitivity[8]. Both immunofluorescence assay (IFA) and enzyme-linked immunosorbent assays have successfully been used in some countries. However, the high cost and sophisticated equipment requirements coupled with the need for expertise limit their use in routine diagnosis[57, 58]. Antigen-specific tests have also been developed for diagnosis of trypanosome infection. In fact, although none of the test developed to date have actually been successful, this strategy is more attractive, as it will solve the dilemma of deciding whether or not to treat serologically positive but parasitological negative subjects. The

TryptectCIATT, one of such test was evaluated with promising results, but subsequent studies raised several doubts about its specificity[59, 60]. Other indirect methods have been evaluated using other body fluids such as saliva[61]. Recently, a new antibody-based test (SD BIOLINE HAT) was launched in the Democratic Republic of Congo which is expected to be cheap, easy and fast to use.

So far, there is no equivalent of the CATT/*Tbg* for field screening of *Tbr*. Several serological tests do exist however, but they are not used for routine field screening as is the case with CATT/*Tbg*. For example, the trypanosomiasis agglutination card test and other variants as well as other immunofluorescence and ELISA-based test are all used in reference centers[62, 63]. Confirmation of *Tbr* infection and disease staging is achieved using similar methods as with *Tbg*. But unlike in *Tbg* infection, *Tbr*-infected subjects have high parasite counts and therefore ease parasitological diagnosis.

1.1.5.2 Parasitological Diagnosis of Sleeping Sickness

Parasite detection is a major step toward confirmation of trypanosome infection, and in *Tbg* infection, this is often done on serologically positive patients (**Figure 3**). However, it is a more routine procedure in the diagnosis of *Tbr*, as very few serological tests exist. Parasitological examination is carried out on blood, chancre aspirates, CLN and CSF during mass population screening, or concomitantly with diagnosis of other infectious diseases[64]. Although trypanosome detection suffers from low sensitivity due to low parasite load (in *Tbg* especially) on the one hand and user experience on the other hand, it however allows for disease staging and diagnostic confirmation. It therefore helps in the decision of the treatment regimen. The sensitivity of parasite detection methods ranges between 40% to 80% and the sample must be processed as fast as possible, since trypanosomes are lysed by sunlight[65]. To increase sensitivity, several centrifugation techniques have been developed. The microhematocrite centrifugation technique or the Woo test, the quantitative buffy coat are all centrifugation techniques, that improve parasite detection by increasing parasite concentration[66, 67]. The quantitative buffy coat is more specific since parasite kinetoplast is stained with acridine orange, differentiating it from the other white cells. Mini-anion-exchange centrifugation technique, which is based on the separation of trypanosomes (that are less negatively charged than blood cells) from blood, and their

subsequent concentration by centrifugation increases sensitivity. The test sensitivity can be enhanced by using buffy coat[15].

1.1.5.3 Molecular Diagnosis of Sleeping Sickness

An arsenal of molecular tools has been developed for the molecular diagnosis of HAT, with PCR at their basis[68]. They range from the amplification species-specific DNA fragment, to sub-type identification[69-71]. Studies have been performed, to evaluate the specificity and sensitivity of PCR-based diagnosis of HAT and the results are promising although there are still some variations in accuracy[72, 73]. The major drawbacks observed so far, is the need of sophisticated equipment, trained personnel and constant electricity supply, which are conditions that are not always all met in endemic areas. With regards to the later, isothermal amplification techniques, such as the Loop-Mediated Isothermal Amplification (LAMP)[74] and Nucleic Acid Sequence-Based Amplification (NASBA)[75] have been developed as well as other combined techniques including the oligochromatography (OC), and the HAT-PCR-OC[76]. Current trends indicate that parasite detection methods might soon be replaced by PCR-based diagnosis, in areas where the facilities permit given the reported specificity and sensitivity[72].

1.1.5.4 Treatment of Sleeping Sickness

There are currently very few drug used for the treatment of HAT, and are classified into two groups. Their classification depends on whether or not they can cross the blood-brain barrier and to a lesser extent, their effectiveness on different parasite species. Meanwhile the precise mechanism of action in either case remains poorly understood, some are exclusively stage I drugs, and other recommended for both disease stages[77]. Generally, drugs used for the treatment of stage I disease are well tolerated and show mild side effects (though reversible). Unlike those used in stage I disease, stage II drugs are associated with post treatment related encephalopaties and death in some cases[78]. Pentamidine and Suramin are the drugs of choice for the treatment of stage I infection by *Tbg* and *Tbr* respectively.

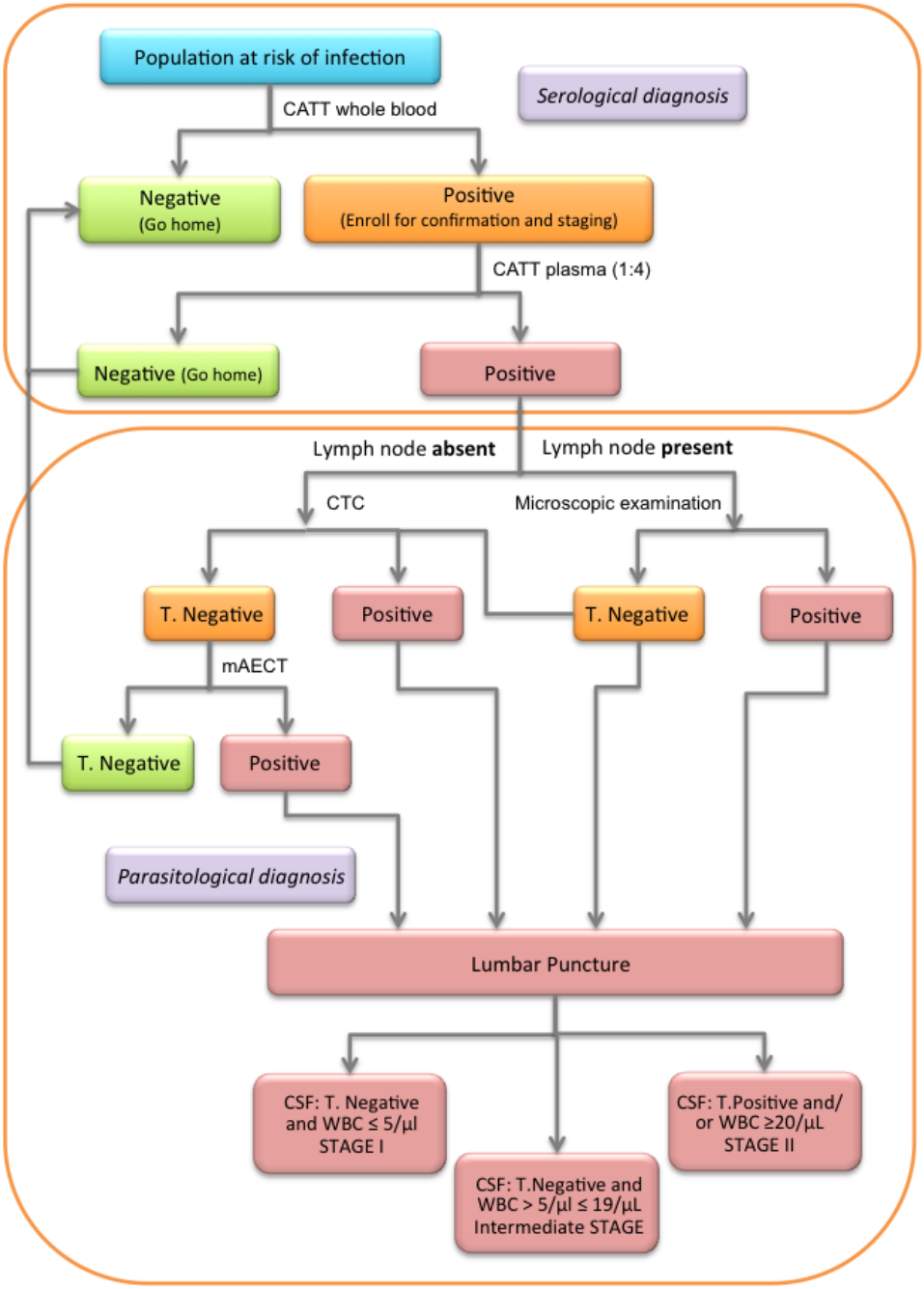


Figure 3: Example of a conventional field algorithm for the diagnosis and staging of *T.b. gambiense* HAT. (*Adapted from Bouteille and Buguet, 2012*). WBC: White blood cell count, CTC: Capillary centrifugation technique, CSF: Cerebrospinal fluid, CATT: Card Agglutination test for Trypanosomiases, mAECT: mini-anion exchange Chromatography technique, T. Neg: Trypanosome Negative

Treatment failure for these drugs is very rare and pentamidin is even more tolerated than Suramin, since it causes very few side effects[79]. Treatment of stage II *Tbg* and *Tbr* infection currently rely on melarsoprol, which due to severe adverse effects is not recommended for the treatment of stage I disease. Between 5–10% of patients treated with melarsoprol suffer from encephalopathy and die as a result of side effects

such as loss of consciousness, convulsions, edema, coma and subsequent death[80]. Alternative stage II treatment of *Tbg*, but not *Tbr* infection, are now available and have been reported to be more effective, with few relapse cases and side effects. Eflornithin and nifurtimox-eflornithine combination therapy (NECT) have recently been recommended by the WHO for the treatment of stage II *Tbg* infection especially in foci with high relapse of melarsoprol treatment[81]. Early diagnosis and treatment of infection remains the most efficient way to avoid the use of such toxic substances, which would not pass the strict safety regulations of today. The search for early, reliable and patient-friendly diagnostic biomarkers is thus the first indispensable step towards such goals.

1.2 miRNAs

miRNAs are a class of highly conserved gene regulatory RNA molecules in multicellular organisms playing key roles in the fate of many protein-coding genes. miRNAs are among the most abundant regulatory elements in the human genome, comprising up to about 5% of known human genes[82]. They are approximately 23 nucleotide long with anti-sense complementarity to the 3′ UTR of their target mRNA. Following maturation and integration into the RNA-induced silencing complex (RISC), they direct the complex to the target mRNA leading to cleavage or translational repression[26, 83]. They are known to control key cellular processes such as regulation of neurite outgrowth[84], cell differentiation[85], cell cycle control[86], cellular communication and signaling[87], cytokine interaction[88], apoptosis[89], Cholesterol biosynthesis[90] and there is mounting evidence for their involvement in some human diseases[91], protein homeostasis, glycosphingolipid biosynthesis, purine metabolism, axon guidance, and the regulation of actin cytoskeleton[92].

1.2.1 miRNA Biogenesis

Clusters of miRNA genes are commonly found in many regions of the genome. They frequently found within intergenic regions and introns of protein-coding genes[93, 94]. Regulation of miRNA transcription units greatly vary depending upon the genomic location of the transcription unit; intronic miRNA transcription units that are located within a host gene and in the same orientation are transcribed alongside the primary transcript of the host gene by the same promoter[94]. Intergenic miRNA transcription

units on the other hand are presumably transcribed from their own independent promoter[94]. The transcripts of these miRNAs are believed to be polycistronic with distinct 5′ and 3′ boundaries with Cap and polyA tails as in mRNAs[95].

In response to diverse endogenous as well as exogenous signals, miRNAs can either be transcribed by RNA polymerase II or RNA polymerase III in a process that is tightly controlled by transcription factors and other miRNAs or by DNA methylation[96-98]. Before miRNAs become biologically functional, they undergo a two-step cleavage process performed by two distinct ribonuclease II endonucleases (Drosher in the nucleus and Dicer in the cytoplasm).

The primary miRNA transcript or pri-miRNA is processed in the nucleus by the Drosher to yield a pre-miRNA or precursor miRNA which is exported into the cytoplasm via Exportin5 and Ran-GTP transporters. It further undergoes cleavage by the Dicer to yield a miRNA duplex of the mature miRNA or driver miRNA and passenger miRNA. The driver is intergrated into the RNA-induced silencing complex (RISC) and a passenger miRNA which can function coordinately with the driver strand to enhance repressibility[99, 100]. Once loaded into the RISC, the miRNA leads the complex toward the target mRNA, via complementary sequences at the 3′ UTR of the later leading either to translational repression or degradation.

1.2.2 Biological Functions of miRNAs

To date, studies have revealed, that in fine-tuning gene expression as well as the expression of other RNA species, miRNA activity can result in a variety of phenotypes. As such, miRNAs have been shown to be involved in diverse cellular processes such as growth, differentiation, immune responses and in disease states. Physiologically, miRNAs have been reported to play key roles in the maintenance of the survival of mature neurons and their function[101], in the regulation of apoptosis[102], in cell differentiation, and development[103], and in the modulation of immune responses in mammals[104]. miRNAs may also mediate angiogenesis *in vivo*[105] and when transported by exosomes, they can actively relay information between cells[106]. There is also evidence, that miRNAs may play a role in stabilizing molecular networks, by buffering against perturbations. This is as a result of the involvement of miR-7 in several interlocking feedback and feed-forward loops. In mammalian reproduction,

miRNA play major role in diverse physiological processes such as oocyte maturation, luteum development, early embryo development and testis differentiation[107, 108]

1.2.3 miRNA in Disease

The implication of miRNAs in both communicable and non-communicable diseases has been reported. In viral infections for example, it has been shown, that viruses can use host miRNA in their struggle to hijack the host cell metabolism. Meanwhile miRNAs are also involved in host innate immune responses and host cells do use these molecules in defense against both RNA and DNA viruses. There is therefore a mutually exclusive mechanism exerted by both host and parasite in their struggle for survival[109, 110]. Several anti-viral miRNAs have been reported to have negative effects on the replication and accumulation of some viruses involved in human diseases such as hepatitis C virus[111]. miRNAs participate in myocardial diseases through their effects on cardiac fibroblast[112] and their involvement in cancer is well documented. In gastro-intestinal caners for example, miRNA inhibit the expression of oncogenes and anti-oncogenes, thereby influencing tumorigenesis and cancer progression. Moreover, in some cancers miRNAs can act as tumor suppressors or oncogenes[113, 114].

1.2.4 Circulating miRNAs and Biomaker Potentials

The diagnostic potential of circulating miRNAs has become very eminent since they were first describes about half a decade ago[115]. Although more than 50% of all circulating miRNAs have been reported to be from blood cells[116], their relatively high plasma stability, association with disease conditions and ease of measurement makes them promising biomarkers for the diagnosis of several diseases. To date, there have been several reports on the biomarker potentials of miRNA in a various cancer types and other diseases. About 79 miRNA have been documented as circulating biomarkers of solid tumors[117].

In infectious diseases, mutually exclusive miRNA regulations have been reported in viral and bacterial infections. For example, during infection by *Mycobacterium tuberculosis*, the bacterium is said to control the expression of miR-99b in murine dendritic cells in its attempt to modulate host immunity[118]. The hepatitis C virus has also developed mechanisms to usurp host miRNA to support viral

proliferation[119]. During infection by protozoa parasites, miRNA deregulation has been reported to be at the basis of important physiological changes that are observed post infection. For example, following *Leishmania major* infection, major changes in macrophage function have been reported to be related or caused by differential miRNA expression[120]. Meanwhile, following *Cryptosporidium parvum* infection, there has been an NF-kB-dependent transactivation of several miRNA genes[121]. In billary epithelial cells (Cholangiocytes), *C. pavum* is capable of inducing the expression of the immune protein B7-H1 by down regulating miRNA-513[33]. There is therefore accumulating evidence in support of cross regulation of host miRNA in infectious diseases. This has raised the prospects that miRNAs, and especially circulating miRNAs could serve as diagnostic biomarkers in some of such diseases.

Despite this, very few studies have been devoted to the search for miRNA biomarkers in most infectious diseases. With no reliable staging biomarker, and the use of less reliable and highly invasive diagnostic procedure in HAT, the identification of possible circulating biomarkers could go a long way to enhance control measures, encourage participation and ease the work of control teams thereby reducing the use of highly toxic arsenics that are currently used for treatment.

2 Materials and Methods

2.1 Materials

Kits

Product	Catalogue N°	Manufacturer
miScript II RT Kit	218160	Qiagen
miScript SYBR® Green PCR Kit	218073	Qiagen

Chemicals

Product	Catalogue N°	Manufacturer
Biozyme LE Agarose	840004	Biozyme
Chloroform	288306	Sigma-Aldrich
Ethanol, absolute	24102	Sigma-Aldrich
Ethidium bromide	M3178.0010	Genaxxon
Ethylenediaminetetraacetic acid	E9884	Sigma-Aldrich
GeneRuler DNA Marker	SM0311	Thermo Scientific
Guanidine hydrochloride	6069.2	Carl Roth
Hydrochloric acid (HCl), 37%	85848.290	VWR International
Isopropanol	W292907	Sigma-Aldrich
Nuclease-free Water	AM9939	Life technologies
Parafilm® M	P8505	Labomoderne
Q5 High-Fidelity DNA polymerase	M0491	New England Biolabs
RNaseOUT™ Ribonuclease Inhibitor	10777-019	Invitrogen
Sodium chloride (NaCl)	S9888	Sigma-Aldrich
Sodium phosphate dibasic (Na_2HPO4)	S9763	Sigma-Aldrich
Sodium phosphate monobasic (NaH_2PO4)	S9638	Sigma-Aldrich
Trizma® Base	T1503	Sigma-Aldrich
Trizma® HCl	T3253	Sigma-Aldrich
TRIzol Reagent	15596-018	Invitrogen

Labware

Product	Catalogue N^{o}	Manufacturer
8 strip PCR tubes (0.2 ml)	AM12230	Life technologies
Adhesive PCR Seal	600208	Biozyme
Eppendorf safe lock micro centrifuge tubes (0.5 ml, 1.5 ml and 2 ml)	0030121694/ 0030121597/ 0030121570	Eppendorf
Gloves, Latex Medical Examination	BM11228-PF-AV	Blossom
Gloves, Nitril Freeform SE	FFS-700	Microflex
LightCycler® 480 Multiwell Plate 384, white	04729749001	Roche
MultiScreen PCR µ96 Plate	LSKM PCR 50	EMD-Millipore
Thin wall standard 96-well PCR plate	82-0600-A	Peqlab

Equipment

Name	Manufacturer
Biofuge, pico	Hereaus Instruments
Electronic Balances, Kern 434,440-45	Kern & Sohn GmbH
Heating Block	Grant Instrument
Ice maker	Scotsman
Microcomputer electrophoresis power supply	Renner GmbH
Microwave oven	Bosch
Nanodrop Spectrophotometer N1000	Peqlab
Refrigerated centrifuge 2k15	Sigma
Swing wing Centrifuge R5810	Eppendorf
Thermocycler	MJ Research, LifeEco
TKA MilliQ watter supply	Millipore
Ultraspec 2000 Spectrophotometer	Pharmacia Biotech
Vortex Mixer	Neolab
Lightcycler®480 System	Roche

2.2 Methods

2.2.1 Ethical Issues

Written informed consent forms were obtained from patients and healthy individuals whose blood samples were collected and included in the present study. At the local level, ethical approval was obtained from the World Health Organization (WHO) control program for Trypanosomiases in West Africa (RPC 222/14.06.2007) and from the Heidelberg Ethical Commission (S-171/2012). All individuals who participated in the present study received an explanation of the scope of the study before they signed the consent forms.

2.2.2 Blood Samples

During routine field screening by teams of the WHO control program for Trypanosomiases in West Africa (Jammoneau and Coworkers), indigenes of in the Bofa sleeping sickness focus (Guinea) were screened with the Card Agglutination Test for Trypanosomiasis (CATT) for whole blood. Samples with a positive CATT result were screened using the CATT plasma dilution test; all individuals who were positive at a dilution of ≤¼ were further examined for parasites using the buffy coat concentration technique[15], and by examination of lymph node aspirates if available, Stage determination was done by white cell count for all newly infected individuals. Five hundred microliters of plasma or whole blood was stored for a trypanolysis test[122]. All individuals with a positive CATT test, with or without confirmed presence of the parasite, and who accepted follow up and treatment were invited to the Bofa local Health center for enrollment into the surveillance program. 2.5 ml of blood was collected by venopuncture directly into the PAXgene Blood RNA tubes (PreAnalytics-BD Company) from all consented participants. The tubes were kept in a refrigerator at −20°C for 2 days, then at 80°C.

2.2.3 Total RNA Extraction

Total RNA was extracted from blood samples using the peqGold RNA extraction reagent (PeqLab) following an optimized procedure. Blood samples in Paxgene tubes were centrifuged at 5000×g for 10 minutes at 4°C. The supernatant was discarded and the pellet completely re-suspended in 10ml of nuclease-free water by vortexing.

Samples were again centrifuged at 5000×g for 10 minutes at 4°C and the resulting pellet re-suspended in 2 ml of TriFast PeqGold. 400 µl of chloroform was added and the samples homogenized for 30 seconds at room temperature before allowed for 3 minutes at room temperature. The aqueous phase was separated by centrifuging at 12000×g for 15 minutes at 4°C and transferred into an RNase-free eppendorf tube containing 500µl of Isopropanol. The tubes were kept at –20°C for one hour and centrifuged at 12000×g for 10 minutes at 4°C. The resulting RNA pellet was washed two times in 75% ethanol, precipitated with 3 M sodium acetate and re-suspended in 50 µl of water. The quality of total RNA was checked by gel analysis using the total RNA Nano chip assay on an Agilent 2100 Bioanalyzer (Agilent Technologies GmbH, Berlin, Germany). RNA concentrations were determined using the NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

2.2.4 Molecular Diagnosis

PCR-based diagnosis was performed on all patient samples using species-specific primers. DNA was extracted from samples by ethanol precipitation of the aqueous phase obtained after RNA extraction using peqGold Trifast following manufacturer's recommendations (Peqlab). The PCR reaction was carried out in a 25 µl reaction using the Q5 high-fidelity DNA polymerase (New England Biolabs). The primers used allow for specific detection of *T. b. gambiense*[69]. PCR Product DNA was visualized by ethidium bromide staining on a 1.5% agarose gel.

2.2.5 miRNA Expression Profiling

Analysis of the differential expression of circulating miRNAs was done using the miRNA Microarray System with miRNA Complete Labeling and Hyb Kit (which represents 1,205 human and 144 human viral miRNAs) following the manufacturer's instructions (by the DKFZ core facility). Briefly, after total RNA extraction and quality control using the Agilent Bioanalyzer, 100 ng of total RNA was dephosphorylated using Calf Intestinal Alkaline Phosphatase at 37°C for 30 minutes. The samples were then denatured in 100% DMSO at 100°C for 5 minutes and ligated to Cyanine3-pCp at 16°C in a circulating water bath for 2 hours and purified on a Micro bio spin column. The eluate was vacuum dried in a vacuum concentrator at 55°C. Samples were resuspended in 18 µl of Nuclease-free water. 4.5 µl of the 10× GE Blocking Agent and 22.5 µl of 2×

Hi-RPM Hybridization Buffer was added to each sample and mixed by vortexing. Samples were then heated at 100°C for 5 minutes and kept on ice. Hybridization was done in a SureHyb chamber at 55°C for 20 hours in a hybridization oven. Slides were washed two times at room temperature and once at 37°C for 5 minutes and scanned using an Agilent scanner (SureScan). Data was extracted using Agilent feature extraction software and analyzed with Chipster microarray data analysis software. Gene functions were annotated using the GeneCard database (http://www.genecards.org/)[123].

2.2.6 qRT-PCR

qRT-PCR was performed to confirm the profiles observed from miRNA expression profiling. To this end, 0.75 µg of total RNA was reverse transcribed into cDNA in a total volume of 20 µl using the miScript Reverse transcription kit (Qiagen) according to the manufacturers recommendations. Following cDNA synthesis, the resulting cDNA was diluted 10 fold before being used for real time PCR. The miScript primer assay for Syber green-based real time PCR (Qiagen) was used for qRT-PCR in a total volume of 12 µl, containing 1 µl of diluted cDNA in a lightCycler 480 system (Roche). The entire reaction was composed of 40 cycles, consisting of an initial activation step at 95°C for 15 minutes followed by 40 consecutive cycles of 94°C for 15 seconds, 55°C for 30 seconds and 70°C for 30 seconds for transcript quantification.

2.2.7 Gene Expression Profiling

Gene expression profiling was performed (by the DKFZ Genomic and proteomic core facility) using the illumina Human Sentrix-12 BeadChip arrays, which contain more than 47,000 probes. Biotin-labeled cDNA samples for hybridization on Illumina Human Sentrix-12 BeadChip arrays (Illumina, Inc.) were prepared according to Illumina's recommended sample labeling procedure[124]. In brief, 200 ng total RNA was used for complementary DNA (cDNA) synthesis, followed by an amplification/labeling step (*in vitro* transcription) to synthesize biotin-labeled cRNA according to the Illumina® Total Prep™ RNA Amplification Kit (Life Technologies). Biotin-16-UTP was purchased from Roche Applied Science, Penzberg, Germany. The cRNA was column purified according to TotalPrep RNA Amplification Kit, and eluted in 60 µl of water. Quality of cRNA was controlled using the RNA Nano Chip Assay on

an Agilent 2100 Bioanalyzer and spectrophotometrically quantified (NanoDrop). Hybridization was performed at 58°C, in GEX-HCB buffer (Illumina Inc.) at a concentration of 100 ng cRNA/µl, unsealed in a wet chamber for 20 hours. Spike-in controls for low, medium and highly abundant RNAs were added, as well as mismatch control and biotinylation control oligonucleotides. Microarrays were washed once in High Temp Wash buffer (Illumina Inc.) at 55°C and then twice in E1BC buffer (Illumina Inc.) at room temperature for 5 minutes (in between washed with ethanol at room temperature). After blocking for 5 min in 4 ml of 1% (wt/vol) Blocker Casein in phosphate buffered saline Hammarsten grade (Pierce Biotechnology, Inc., Rockford, IL), array signals are developed by a 10-min incubation in 2 ml of 1 µg/ml Cy3-streptavidin (Amersham Biosciences, Buckinghamshire, UK) solution and 1% blocking solution. After a final wash in E1BC, the arrays are dried and scanned. Microarray scanning was done using an iScan array scanner. Data extraction was done for all beads individually, and outliers were removed when >2.5 MAD (median absolute deviation). All remaining data points were used for the calculation of the mean average signal for a given probe, and standard deviation for each probe was calculated.

2.2.8 Target Prediction and Core Analysis

miRNA target prediction was done using the target prediction software incorporated into the Ingenuity Pathway Analysis (IPA) software. To this end, both highly predicted and experimentally identified miRNA targets with relevance to pathogen induction as well as immune responses were queried. All resulting miRNA targets were scored against all regulates genes that were differentially regulated from the gene expression profiling experiments. miRNAs and corresponding targets that went through this filter were subjected to a core analysis in IPA to find out cross relationships and potential downstream effects involving other molecules that could be major players in infection.

2.2.9 Statistical Analysis

Data analysis was done using Chipster microarray data analysis software. All samples were quintile normalized across chips and filtered according to standard deviation (0.95) and interquatile range. The empirical Baye's two group t-tests ($p < 0.05$) was used to test for differential miRNA expression between different sample groups

and the Benjamini-Hochberg correction was applied over all p-values of the differential expression analysis. For linkage clustering, the pearsons correlation coefficient was calculated. Quantitative Real Time (qRT-) PCR was carried out in triplicates for a confirmation of microarray data. Resulting data were expressed as mean ± SE and all miRNA with a mean difference having a p-value of <0.05 for a two-sided unpaired student *t-test* were considered significantly regulated.

3 Results and Discussion

3.1 Patient Screening

During a routine screening campaign, a total of 14,445 individuals were screened (by the sleeping sickness control program in West Africa) with the CATT test. Of this number, 324 tested positive for the CATT on whole blood while 114 had a positive test for the CATT plasma at a fourfold dilution. Trypanosomes could be found in 45 out of 114 subjects, and the remaining 69 subjects were classified as seropositive. Forty samples were chosen for our study (**Table 1**). They included eight control samples from sero-negative, parasite-negative people; five CATT positive, trypanolysis positive but parasitologically and PCR-negative individuals, and seven CATT positive, trypanolysis, parasitologically and PCR-negative individuals. The remaining 20 subjects were parasitologically positive patients: nine in stage I and eleven in stage II. Sixteen of these were positive after a single round of PCR, and the remaining two were positive after a second PCR round. RNA was prepared from the 40 samples and used for miRNA and gene expression analysis.

Interestingly, a treated patient who was among the seropositives could be classified as infected both by miRNA and the trypanolysis test for which it tested positive for all three antigens. Besides, some seropositive patients who tested trypanolysis negative were classified as being positive. This could be due to co-infection with other trypanosome strains, or unrelated parasites. Recently, the trapanolysis test was miniaturized, and used for the screening of about 480 treated and untreated patients, with a close to 100% sensitivity and specificity[125]. It is therefore possible to speculate, that seropositive subjects classified by both by miRNA and the trypanolysis test are patients either at an early stage of infection, or with controlled infection. This further explains why they could not be confirmed parasitologically as patients.

The differentially regulated miRNAs were able to distinguish between infected and uninfected subjects. Within the group of seropositives, six of them were classified as infected by the miRNAs and other as uninfected. Among those classified as infected, three of them were positive for at least one of the antigens for the trypanolysis test.

Table 1: Sample classification based on multiple diagnostic tests

Patient	CATT	Mn-BC Parasite count	White Cell count	Trypa-nolysis	miR-Class.	Status	Stage
Bo.488/6	+	>100	51	+	A	HAT	II
Bo 482/6	+	>50	32	+	A	HAT	II
Bo.483/6	+	>100	80	+	A	HAT	II
Bo.477/6	+	>20	15	+	A	HAT	II
Bo.479/6	+	2	13	+	A	HAT	II
Bo.484/6	+	>100	1	+	A	HAT	I
Bo.470/6	+	>100	0	+	A	HAT	I
Bo.478/6	+	>10	228	+	A	HAT	II
Bo.480/6	+	>100	5	+	A	HAT	I
Bo.473/6	+	6	6	+	A	HAT	II
Bo.485/6	+	>50	6	+	A	HAT	II
Bo.502/6	+/-		1	+	A	HAT	I
Bo.486/6	+	+	6	+	A	HAT	II
Bo.487/6	+	+	2	+	A	HAT	I
Bo.481/6	+	6	1	+	A	HAT	I
Bo.474/6	+	10	212	+	A	HAT	II
Bo.475/6	+	10	5	+	A	HAT	I
Bo.476/6	+/-	>50	541	+	A	HAT	II
Bo.472/6	+	+	0	+	A	HAT	I
Bo.471/6	+	>50	0	+	A	HAT	I
Bo.492/6	+	Neg		+**	A	Seropo/AT	
Bo.498/6	+	Neg		+**	B	Seropo	
Bo.499/6	+	Neg		-	A	Seropo	
Bo.491/6	+	Neg		-	A	Seropo	
Bo.490/6	+	Neg		+**	A	Seropo	
Bo500/6	+/-	Neg		+	B	Seropo	
Bo489/6	+	Neg		+	A	Seropo	

Bo495/6	+	Neg	-	B	Seropo
Bo.527/6	+	Neg	+*	B	Seropo
Bo.493/6	+	Neg	-	B	Seropo
Bo.494/6	+	Neg	-	A	Seropo/AT
Bo.520/6	+	Neg	-	B	Seropo
Bo.537/6	-	Neg	-	B	Control
Bo.538/6	-	Neg	-	B	Control
Bo.509/6	-	Neg	-	B	Control
Bo.511/6	-	Neg	-	B	Control
Bo.514/6	-	Neg	-	B	Control
Bo.518/6	-	Neg	-	B	Control
Bo.521/6	-	Neg	-	B	Control
Bo.529/6	-	Neg	-	B	Control

for Typanoöysis, +, +* and +** = positive for all, one and two antigens respectively
A= Classified as Infected, B= Classified as non-infected, AT=treated patient

3.2 Molecular Diagnosis

PCR-based diagnosis was carried out on all patients as well as seropositive subjects using primers that specifically amplify the *Trypanosome brucei gambiense* glycoprotein (TgsGP). The presence of a 308 bp fragment implies the presence of *Tbg* DNA as previously reported[69]. Of the 20 patients tested, 16 were positive during a first PCR (**Fig. 4**), and the other four were re-amplified in a second PCR. Two seropositive subjects tested positive by PCR, meanwhile the others as well as other patients gave some non-specific bands slightly above 500 bp. In some cases, multiple bands were observed, both for patient and seropositves. However, the said bands could not be correlated or interpreted as a positive test, since the size does not correspond to that of the expected product. Sixteen patient samples tested positive for *Tbg* infection at first PCR run, and the others could only be amplified from the products of the first PCR. Some confirmed patients gave unspecific amplification of fragments either below or above the TgsGP fragment.

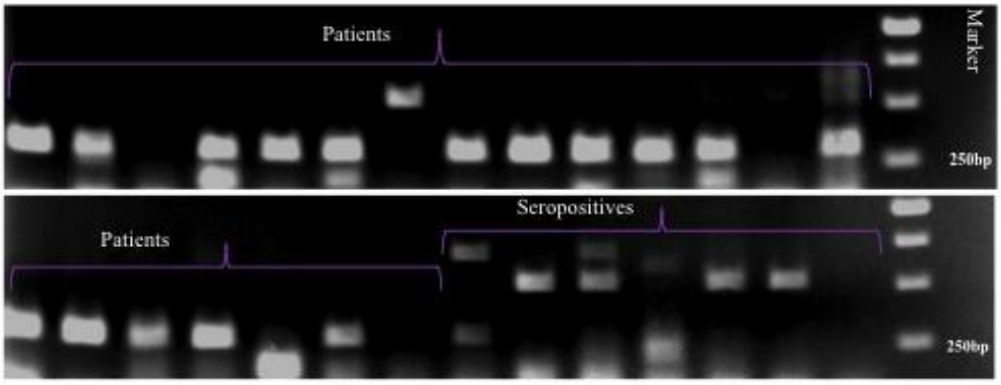


Figure 4: Gel electroforegram showing molecular diagnosis of T. brucei gambiense by TgsGP PCR. (Specific bands at 308bp, nonspecific amplification at 500bp and above)

This was the case with most of the seropositives as well. This could be as a result of a co-infection by other trypanosome species, or parasites that conserved the fragment but are not trypanosomes. Such results have recently been observed during the molecular diagnosis of *Fusobacterium nucleatum*, where Gram staining, culture and molecular diagnosis were discrepant[126]. Low parasite load can explain in part the fact that some patient samples did not amplifiy during the first run. It is also possible, that since DNA was extracted from the left over from RNA extraction, the yield and quality of DNA might have equally played a role.

3.3 miRNA Expression Analysis

We analyzed the expression levels of 1205 miRNAs using circulating miRNA extracted from peripheral blood of *T. brucei gambiense* infected subjects and uninfected controls. A total of fourteen miRNAs were found to be differentially expressed between all patients, irrespective of disease stage and the control group (**Table 2**). Of this number, 13 were significantly differentially regulated between patients with stage-II disease and controls while ten miRNAs were differentially expressed between stage-I patients and controls**.** Combined, ten microRNAs were differentially regulated between all patients with respect to the control group. Two miRNA (miR-338-5p miR-146b-5p) was up regulated only in patients with stage I disease; meanwhile, four miRNAs (miR-22*, miR-195, miR-144* and miR-374c) were differentially down regulated in patients with stage II disease. Only one miRNA was upregulated in all patients irrespective of group. Despite this, none of the stage-regulated miRNA could reliably be considered as a stage marker, due to low specificity.

Table 2: Summary of miRNA differentially regulated between patient vs control

Probe ID	Log_2FC	A. p-Value
*hsa-miR-199a-3p	-6.7	2E - 06
hsa-miR-126	-6.4	2E - 06
*hsa-miR-27b	-6.6	2E - 06
*hsa-miR-98	-6.0	0.0001
*hsa-miR-4291	-5.6	0.0002
*hsa-miR-409-3p	-5.1	0.001
hsa-miR-144*	-4.2	0.009
hsa-miR-195	-4.4	0.01
*hsa-miR-454	-4.0	0.02
*hsa-miR-193b	4.1	0.02

*= Differentially regulated in both Stages FC= Fold Change, Ad = adjusted

Within the seropositive group, a multiple group analysis comprising patients controls and seropositive was performed. Six of the seropositives were classified as infected, and the others were classified as uninfected. Four of these samples tested trypanolysis positive, meanwhile one trypanolysis positive subject was classified as uninfected (**Fig. 4**). In fact, the seropositive group could be one such heterogenous group. Owing to the fact that unspecific reaction of the CATT could classify a subject in this group, or just as seen from that molecular diagnosis, co-infected subjects are in this group as well. Besides, some of the subjects in this group could be patients or with asymptomatic carriers and subjects with controlled infection with low parasite load. This could explain the observed classification of these samples as seen in (**Figure 5)**; however, it does not explain why a trypanolysis-positive subject was classified as un infected.

Perhaps the general trend for most the differentially regulated miRNAs was, for this sample, more toward the uninfected than the infected. This could mean that the subject was either having a controlled infection leading to restricted differential regulation of the measured miRNA, or the pathways involved in infection control did not involve these miRNAs.

Generally, most of the differentially regulated miRNAs were down regulated in patients except for miR-193b (**Table 2)**. In diseases such as cancer, this miRNA has been reported to be down regulated and is involved in the regulation of cellular proliferation, invasion and migration as well as regulation of cyclin D1 in melanoma[127, 128]. In Hepatocellular Cancer (HCC), whose main cause is chronic

hepatitis C infection, mir-193b has been shown to be 5 fold up regulated[129]. This could be an indication that this microRNA is upregulated post-infection. miR-199a-3p has been shown to play diverse roles in cancer and other diseases; it is down regulated in human Osteosarcoma, where it regulates cell proliferation and migration[130].

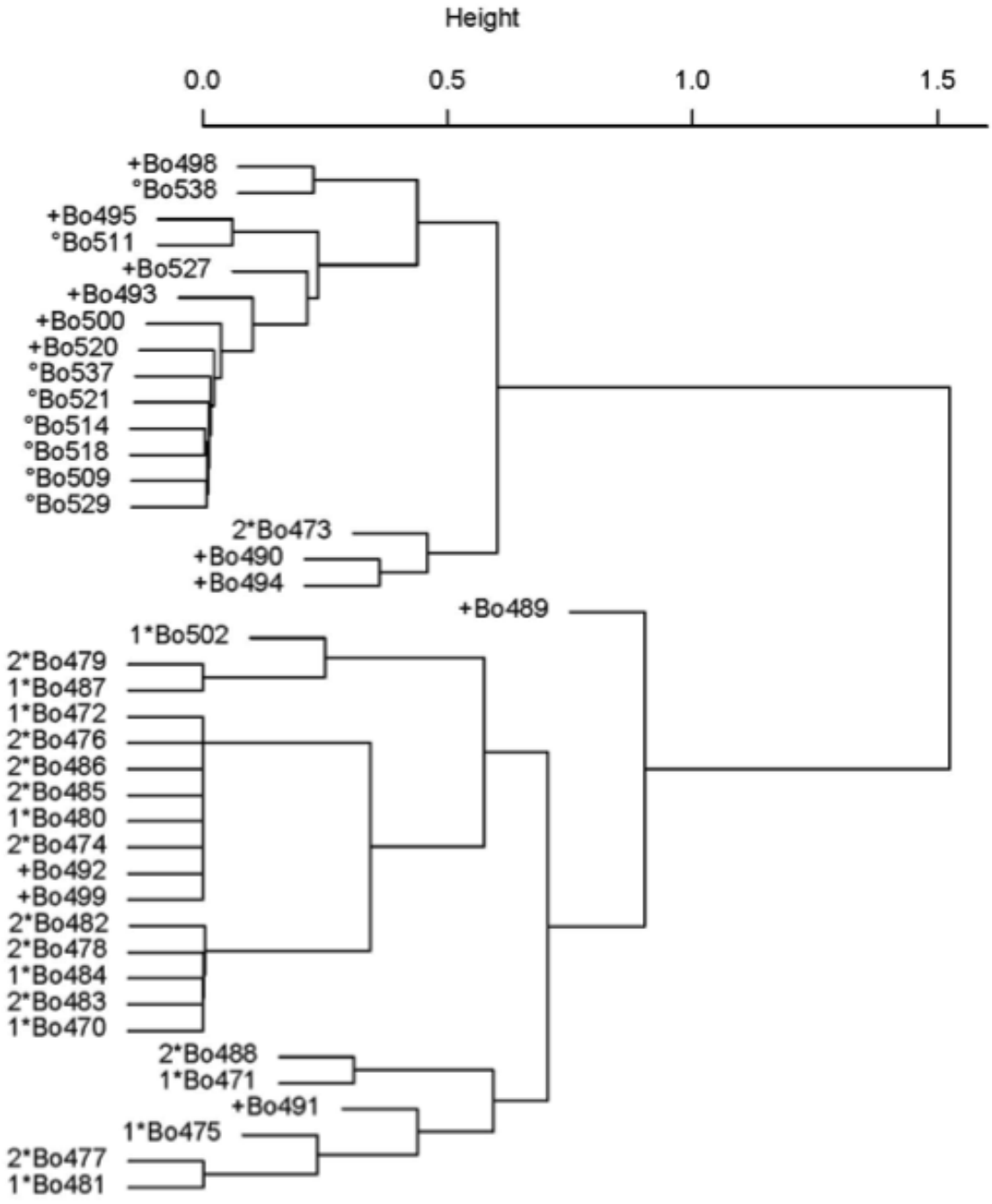


Figure 5: Cluster Dendogram for all Samples showing associations between samples.
Symbol: 1 = Stage I, 2 = stage II, * = Infected, + =Seropositive, °= Control

It is equally down regulated in endometrial cancer and regulates cellular proliferation by targeting the mammalian targets of ripamycin[131] and in breast cancer

it regulates proliferation and survival by targeting caveolin-2[132]. In a recent report, this microRNA was found to be significantly upregulated in sera from gastric cancer patients and was significantly associated with tumor invasion and with lymph node metastasis. It was later proposed was to be a reliable early marker for gastric cancer[133]. It was one of the microRNAs with 100% selectivity as reported recently in gastric cancer and is also involved in the suppression of hepatitis B viral replication without affecting cell proliferation[134]. Although we found this microRNA to be down regulated in patients it may however not be a good diagnostic biomarker for HAT, as it is also reported to have similar selectivity in other diseases as well[133].

miRNA-27b just like miR-199a-3p, has been shown to be down regulated post infection during cytomegalovirus infection and exert an antiviral activity[135]. During *Cryptosporadum parvum* infection, this microRNA has been reported to be up regulated and shown to target KH-type Splicing Regulatory Protein (KSRP) in order to coordinate Toll-like Receptor (TLR)- and NF- κ B-mediated defense by gastrointestinal epithelial cells.

miR-126* was also down regulated in all patients with respect to controls, and it has also been reported to be down regulated in sera from patients with acute myocardial infarction[136]. It plays vital roles in lipid metabolism, where for example it has an atherosclerosis limiting effect[137]. In cancer, its roles are contrasting where on the one hand, it is said to exert anti-cancer effects by impairing tumor proliferation, invasion, migration ad survival. On the other hand, it also shows a cancer supportive effect by enhancing blood vessel growth and inflammation[138].

miR-98 just like most of the other miRNA this miRNA was equally down regulated, and like miRNA-27b it is known to be involved in the activation of NF-kB-mediated defense mechanisms involving the toll-like receptor. In fact, following lipopolysaccharide stimulation or during *C. parvum* infection of gastrointestinal epithelial cells (Cholangiocyte), both miRNAs have been shown to enhance the expression of Cytokine-Inducible Src Homology 2-Containing protein[139]. This miRNA is also thought to inhibit the migration and invasion of human esophageal squamous cell carcinoma[140].

miRNA 409-3p is known to have several counter cancer activities. In bladder cancer, this miRNA is also down regulated and by targeting c-Met, this microRNA can impair cell migration and invasion[141]. In gastric cancer, it is also down regulated, where it controls cell proliferation and apoptosis by targeting the transcriptional

regulator PHF10[142]. Again in gastric cancer, this microRNA suppresses tumor cell invasion and metastasis by direct targeting of radixin[143].

miRNA-195 as observed has also been reported to be down regulated in Human Tongue Squamous Cell Carcinoma, where it inhibits cell cycle progression, promotes apoptosis, and reduces Cyclin D1 and Bcl-2 expression[144]. In endometriosis, this miRNA promotes apoptosis, meanwhile enhancing proliferation and growth by targeting Fractalkine[145]. In normal cells, miR-195 competes with HuR for the control of Stim 1 mRNA stability.

Some miRNA such as miR-4291, until now have not been reported to be involved in any disease or healthy condition. miR-144-3p and miRNA-454 both have barely been studied, and no previsions can be made as to the potential role they may play in infectious diseases. The former impairs reverse cholesterol transport and enhances pro inflammatory cytokine production in atherosclerosis[146]; meanwhile the latter is thought to play a role in colon tumorigenesis by regulating TGF-β/Smad signaling[147].

3.4 Stage-Specific Regulated miRNA

Apart from the ten miRNAs that were differentially regulated in all patients irrespective of disease stage, two and four microRNAs were differentially regulated only in a stage-specific manner for stage I and stage II disease respectively. Patients with stage I disease showed a deregulation of miR-146b-5p and miR-338-5p which were down and up regulated respectively (**Table 3**).

The level of mi-146b-5p has been reported to be down regulated in monocytes during obesity, and has been shown to inhibit NF-kB-mediated inflammatory responses by targeting TNF receptor-associated factor-6 (TRAF6)[148]. It has also been shown, that its expression is regulated by interferon gamma via the JAK/STAT pathway and has a negative effect on the activity of NF-kB by targeting IRAK1[149]. In cancer, this microRNA has been reported to impair glioma migration and invasion by targeting MMP16[150]. Besides, this microRNA showed a very low selectivity (<60%) when used to classify patients and controls, and alone, did not allow for the classification of stage I patients (**Fig. 6**).

Table 3: miRNAs differentially regulated between Stage I disease and controls

Probe ID	Log_2FC	Ad. p-Value
hsa-miR-199a-3p	-6.8	0.0002
hsa-miR-27b	-6.7	0.000
hsa-miR-126*	-6.7	0.0003
hsa-miR-98	-5.9	0.003
hsa-miR-4291	-5.2	0.007
hsa-miR-409-3p	-5.5	0.007
*hsa-miR-338 -5p	4.7	0.01
*hsa-miR-146b-5p	-4.4	0.02
hsa-miR-193b	4.4	0.04
hsa-miR-454	-4.1	0.04

*miRNAs differentially regulated only in stage I disease, FC= Fold Change, Ad = adjusted

miR-338-5p on the other hand has been shown to be involved in the regulation of its host gene, the Apoptosis-Associated Tyrosine Kinase (AATK) in rat hippocampal neurons during maturation[151]. It has a counter effect on autophagy in colorectal cancer, where it is equally up regulated and enhances and promotes the migration of colorectal cancer cells by targeting phosphatidylinositol 3-kinase catalytic subunit type 3 (PIK3C3).

The down regulation of miR-146b-5p could possibly be related to excessive production of interferon gamma, which is produced by natural killer cells, together with CD4 and CD4 cells during the neuropathogenesis if sleeping sickness. These cells are activated by the production of the trypanosome lymphocytic triggering factor by trypanosomes. They in turn liberate interferon gamma, which activate macrophages to produce tumor necrosis factor, interleukin-1 and reactive nitrogen species that are thought to be in part responsible for the toxic effects on the brain[36].

Some microRNAs were differentially regulated only when patients with stage II disease and controls were compared (**Table 4**). Three such microRNA were found (miR-22*, miR-374c and miR-144*). When both patient groups were analyzed against the control, miR-146b-5p could not be identified among those miRNA that were differentially regulated in patients with respect to control (**Fig. 7**)

Nonetheless, when individual patient subgroups of patients were considered, it was found both in stage I and stage II subjects. All three miRNAs were down regulated in patients, and when taken individually, miR-22 family has been reported to play

several roles in cancer and other diseases. For example, miR-22 has is known to be involved in the selective induction of p53-dependent apoptosis by repressing p21 and inhibit the progression of cancer by inducing cellular senescence[152, 153]. In some instances, it may function as a proto-oncogene whose aberrations can lead to hematopoietic malignancies promoting stemness and metastasis[154].

miR-144* is involved in normal physiological processes such as hematopoiesis, but its expression post infection leads to a negative regulation of host immune response during influenza infection and tends to promote viral replication by suppressing the expression of TRAF6.

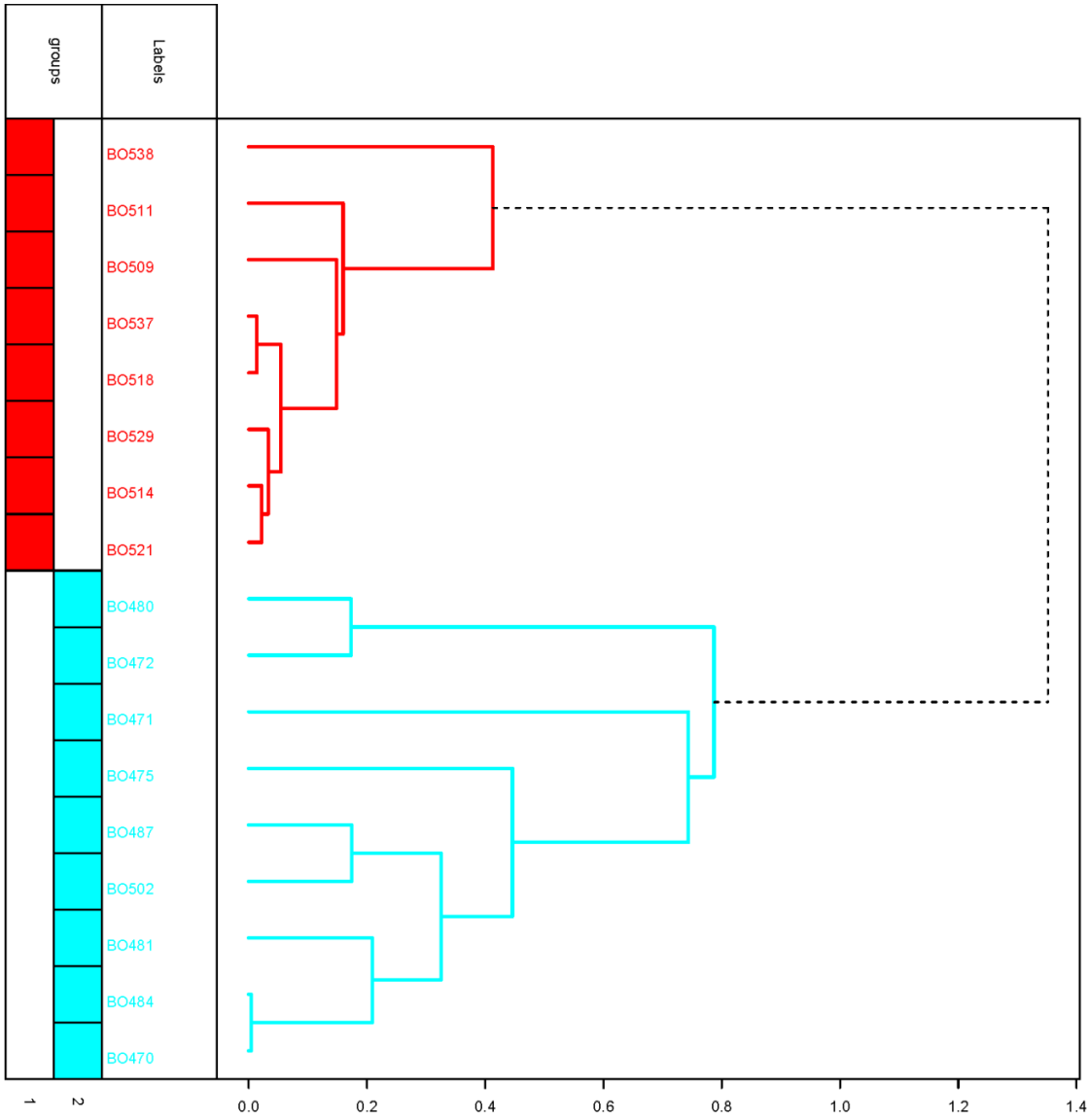


Figure 6: Cluster dendogram for stage I patients *vs* controls for ten miRNAs. Seperation of patients from controls (1) Group I Patients, (2), Uninfected controls

Table 4: miRNAs differentially regulated in Stage II disease and controls

Probe ID	Log_2F C	A. p-Value
hsa-miR-199a-3p	-6.9	1.40E-05
hsa-miR-27b	-6.7	6.00E-05
hsa-miR-126*	-6.3	0.0002
hsa-miR-4291	-6.1	0.0006
hsa-miR-98	-6.3	0.0006
hsa-miR-22	-5.1	0.003
hsa-miR-409-3p	-5.1	0.01
hsa-miR-195	-5.1	0.01
hsa-miR-144	-4.7	0.01
*hsa-miR-146b-5p	-4.3	0.02
*hsa-miR-374c	-4.6	0.02
hsa-miR-454	-4.1	0.04
hsa-miR-193b	4.1	0.04

*differentially regulated in stage II subjects with respect to controls FC= Fold Change, Ad = adjusted

Analysis of single blood cell populations during tuberculosis infection reveals that this miRNA is mainly expressed in T-cells, and is up regulated in patients. It inhibits TNF-α and IFN-γ production and T cell proliferation and therefore could be involved in the modulation of host immune responses by modifying cytokine production[155].

Taken together, the data obtained confirms the already existing theory that most circulating miRNAs originate from blood cells. Since most of them are equally expressed in other diseases and healthy conditions, they may not be suitable for any diagnostic purpose, even if three microRNAs could nicely select all patients from controls. However, there is strong evidence, that *Trypanosoma brucei gambiense* infection leads to deregulation of several microRNAs, most of which have been reported to be involved in host immune responses against communicable diseases.

Given that not all microRNAs differentially regulated during *Tbg* infection are known to pay any roles host immune response, it is possible, that other infection coexisting with trypanosomiasis might have led to the differential expression of such microRNAs since patients were not tested for other infectious diseases or even

metabolic diseases. Modulation of cytokine production and function seem to be the junction at which the global effects of the action of the individual miRNAs meet.

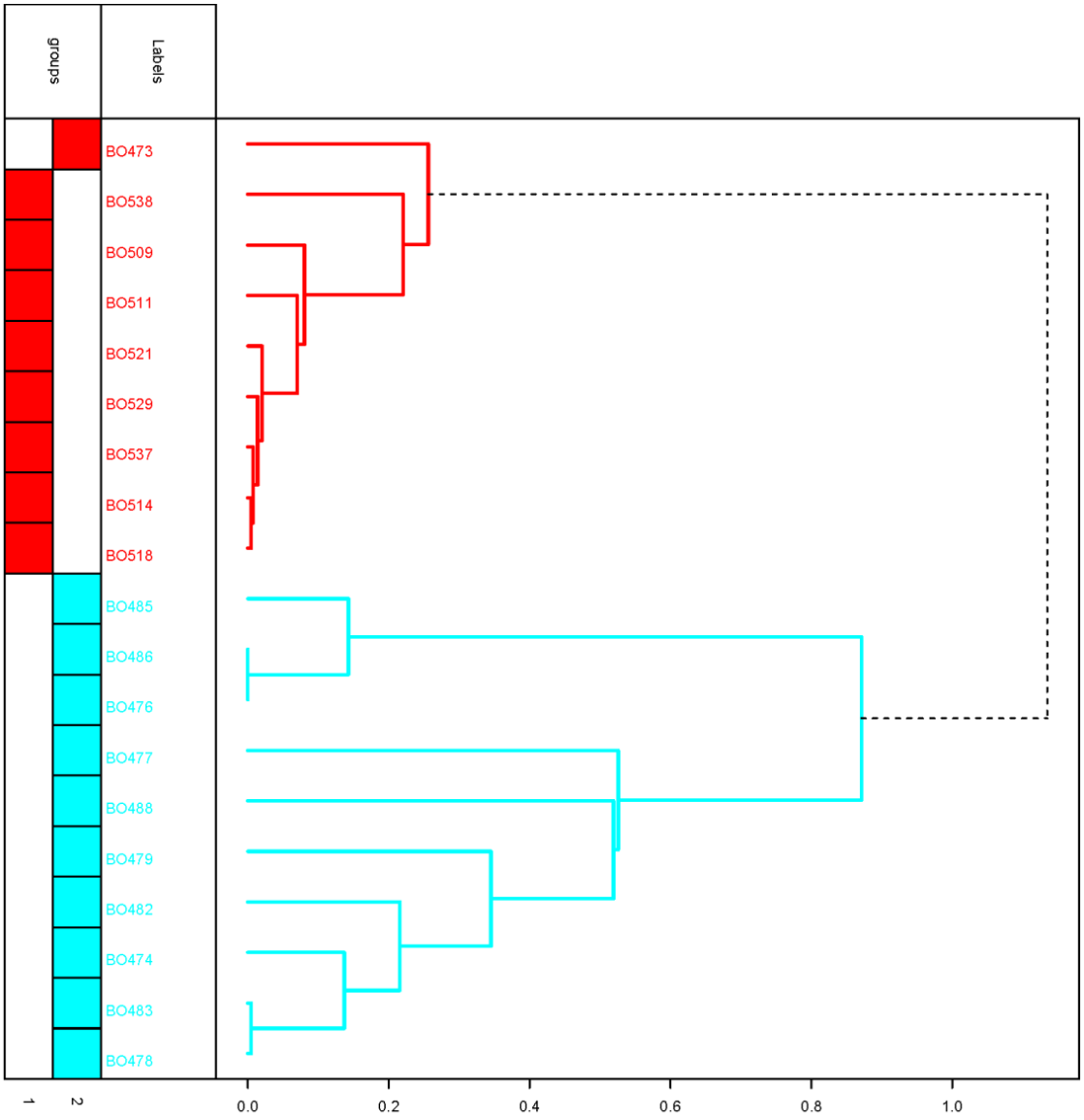


Figure 20: miRNA classification of Stage II patients and controls. *1= controls, 2= patients*

It is, however, not clear if the observed differential regulation of some of microRNA could be as a result of an exclusive mechanism from the infecting parasites in a quest for survival. Even if trypanosomes are not intracellular parasites, this cannot be entirely ruled out. The inability of any one of the identified microRNAs to be suitable as a diagnostic biomarker or stage maker could be as a result of the heterogeneity within the sample groups, and the classification criteria used to attribute the disease stage. It could be that the selectivity of some of these microRNAs becomes

significant with increased sample size since the actual sample size was very limited (20 patients; 11 stage II and 9 stage I, 12 seropositives and 8 uninfected controls) and therefore potentially underpowered.

The prospects of identifying a reliable circulating biomarker for HAT may be possible if single cell populations are separated and analyzed. This might be helpful, not only in finding diagnostic biomarkers, but also correlating miRNA expression profiles with gene expression profiles.

3.5 Biomarker Application of Differentially regulated miRNAs

Ingenuity pathway analysis was used to investigate any existing biomarker application for any of the differentially regulated miRNA. It turned out that miR-199a-3p and miR-195 (**Table 5)** were already reported to be potential diagnostic biomarkers in some diseases. MiRNA-199a-3p was one of the three miRNAs that showed 100% selectivity in HAT. Its involvement in other diseases however precludes it as a reliable diagnostic biomarker for *Tbg* HAT

Of the 13 miRNAs, two miRNAs (miR-193b and miR-338-5p) were up regulated while others were down regulated (**Fig. 8**). Three individual miRNAs (miR-199a-3p, miR-27b and miR-126*) were able to isolate all patients from controls ($p<0.05$) (**Fig. 9**)

Table 5: miRNA and biomarker applications from IPA

Probe ID	Biomarker application	A. p-value
hsa-miR-199a-3p	Diagnosis	4.00E-06
hsa-miR-27b		1.00E-05
hsa-miR-126*		1.70E-05
hsa-miR-98	Unspecified Application	7.20E-05
hsa-miR-4291		0.0003
hsa-miR-409-3p		0.001
hsa-miR-193b		0.006
hsa-miR-195	Diagnosis,Unspecified Application	0.01
hsa-miR-144*		0.01
hsa-miR-454	Unspecified Application	0.04

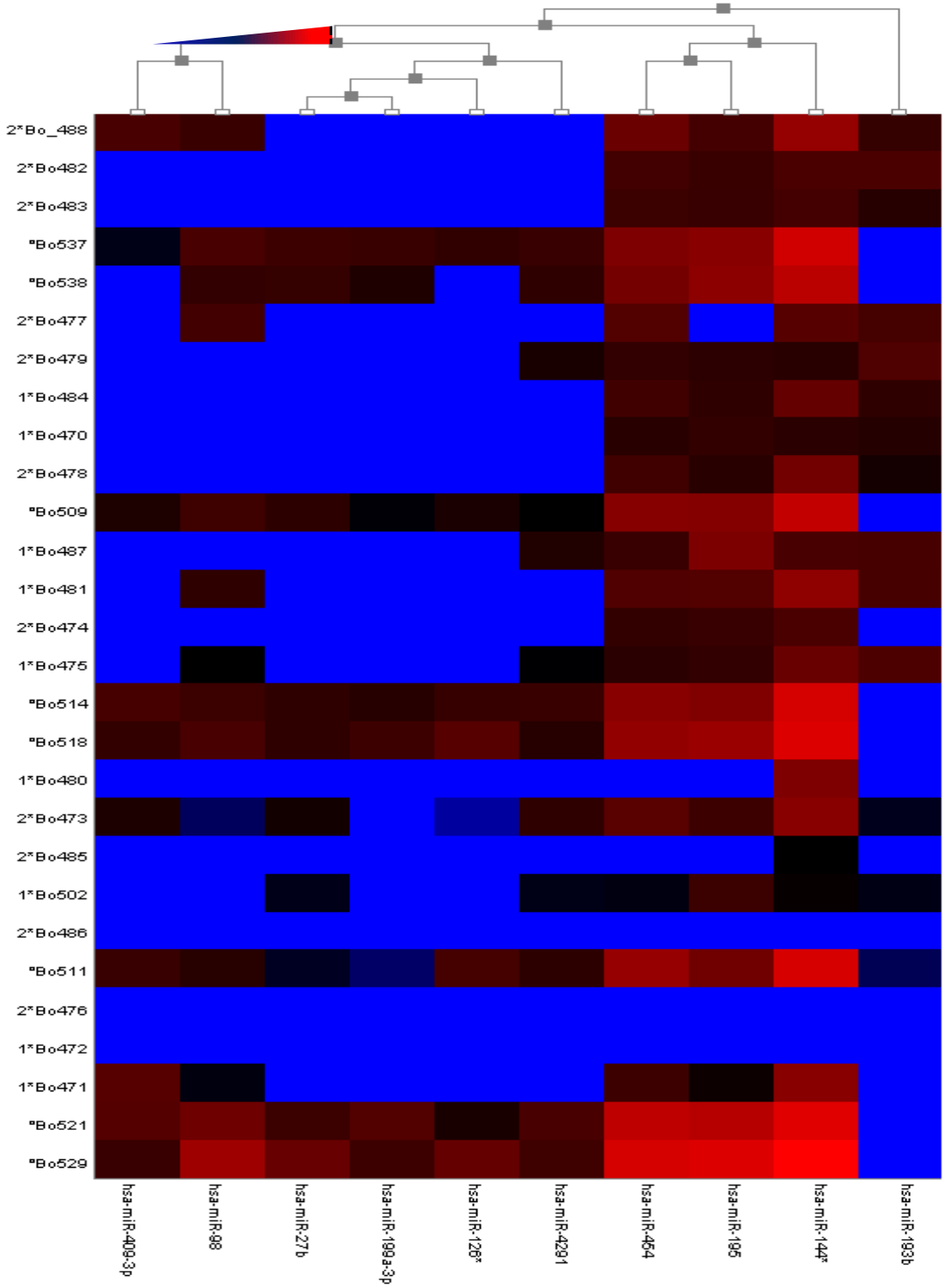


Figure 8: Heirachical clustering of Patients and Controls for 10 regulated miRNAs.
Symbol 1 = Stage II, 2 = stage II, * = Infected, + =Seropositive, °= Control

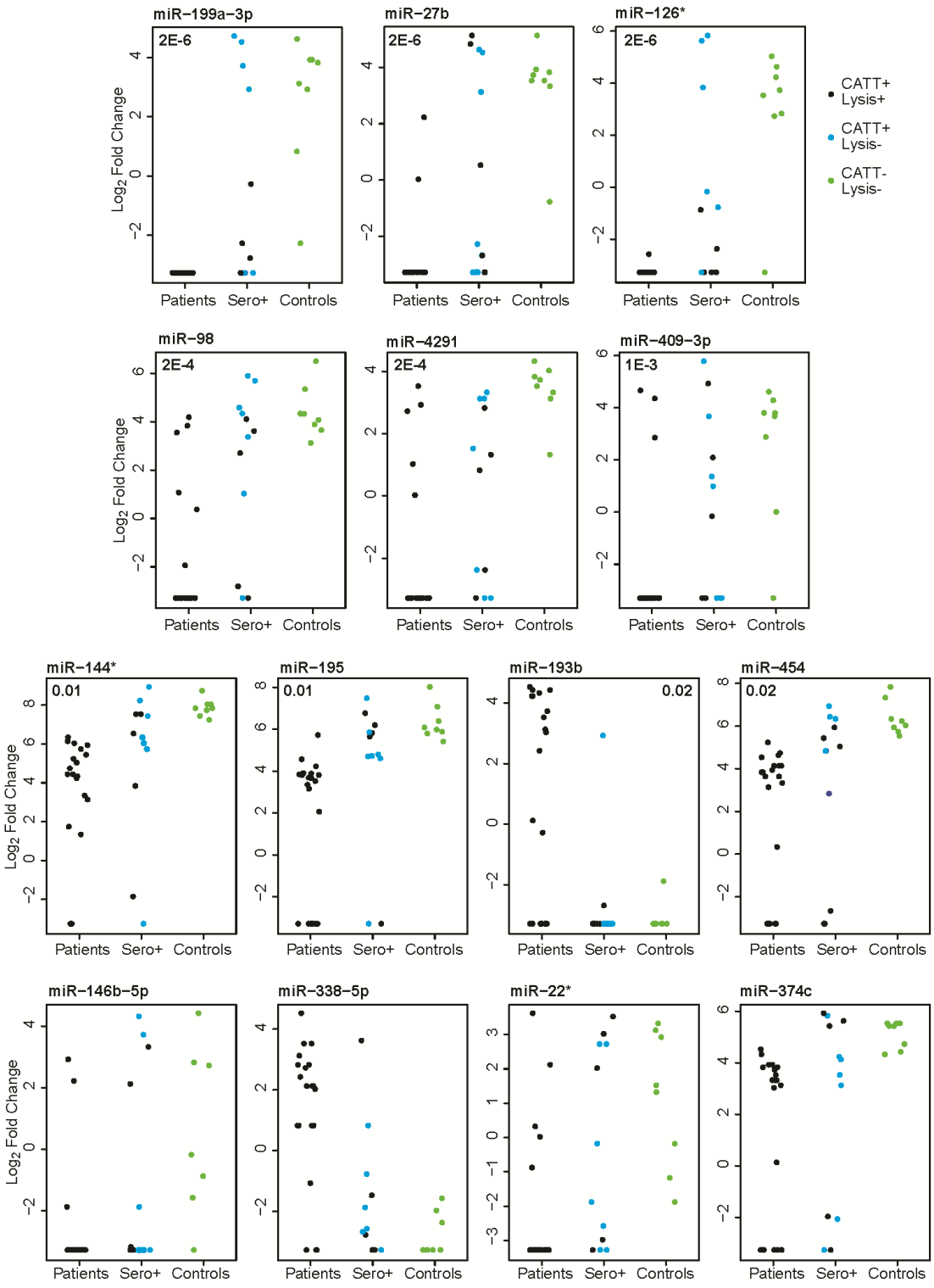


Figure 9: Fold change of individual differentially regulated miRNAs for all sample groups. *CATT and lysis Negative (Green). CATT positive, lysis Negative (light blue) CATT and Lysis Positive (Black)*

3.6 Validation of expression profiles by qRT-PCR

To confirm the observed expression profiles, miRNAs expression patterns was further analysed by qPCR (**Table 9**). Three miRNAs (miR-199a-3p, miR-27b, and miR126*) were analyzed for their expression patterns in patients and controls. The expression profiles for all three miRNAs could be confirmed ($p<0.05$), with miR-199a-3p and miR-27b showing and average 2fold change and miR126* showing an 8 fold change. Since there was no distinct miRNA(s) that were exclusively differentially regulated for the seropositive group, these samples were not considered for further validation by qRT-PCR.

Table 6**: qRT-PCR data for selected differentially regulated miRNAs**

qRT-PCR	**ΔCt**	**ΔCt**	
Probe	**Patient**	**Control**	**p-Value**
miR-199a-3p	11,6 ± 0,6	10,69 ± 0,4	0,03
miR-27b	10,3 ± 1,2	9,1 ±0,8	0,01
miR-126*	9,4 ± 0,7	6,44 ± 0,04	5,40E-10

3.7 miRNA target prediction and core analysis

The targets of the ten miRNAs that were differentially regulated between patients and controls were predicted using algorithms incorporated into the Ingenuity Pathway Analysis (IPA). More than 3,000 highly predicted or experimentally investigated putative targets were found. Since gene expression analysis was performed in parallel, the targets were scored against the list of differentially regulated mRNA transcripts to see if any of the genes showing differential mRNA levels could also be a possible target of the ten miRNAs. This lead to a list of 21 mRNA (**Table 7**) targeted by the ten miRNA, and that are involved either directly or indirectly in infectious diseases.

TIMP2, a moderately elevated mRNA encoding a metalloprotease inhibitor, is a possible target of two of the down-regulated miRNAs (miR-4291 and miR-454). Among the down regulated genes, four (GPR146, EIF2S1, PLA2G4D and MAPK10) were possible targets of one up regulated miRNA (miRNA-193b). Of the predicted targets, pathway analysis indicated that some of the regulated targets were involved in NF-kB signaling, some in apoptosis signaling, others in immune deficiency, leukocyte extravasations, pathogen recognition and response, oxidative stress response, G-protein signal transduction and cell cycle regulation (Index II).

It is not possible to draw any conclusions whatsoever, regarding the relationship between the expression patterns of the miRNA and their potential mRNA targets. the origin of both the miRNAs and the mRNAs are diverse. Even if majority of the miRNA might have come from blood cells, this does not directly imply a similar origin for the mRNA. There may of course be some cross correlation between the expression patterns of some miRNA/´mRNA pairs, but at this point it is practically impossible to draw any clear boundaries.

3.8 Gene Expression Profiling

Gene expression profiling was carried out in parallel to the miRNA expression profiling. A total of about 656 genes were differentially regulated between patients and controls. Of this number, 38 of them were differentially regulated between patients and controls, meanwhile 21 were found to be targeted by at least one of the differentially expressed miRNA and to be involved with pathogen infection. Of the targeted genes, only three were up regulated, of which one (TIMP2) was already reported to play a role in infection. This gene was found to be targeted by two of the down regulated miRNAs (miR-4291 and miR-454). Among the down-regulated genes, four (GPR146, EIF2S1, PLA2G4D, and MAPK10) were targeted by one up regulated miRNA (miRNA-193b).

A core analysis performed on these genes, revealed other related genes whose expression are related to parasitic infections. A total of 25 targets and related genes were found to be involved in parasitic infections (**Table 8**), of which 9 (36%) were already shown to have either biomarker, diagnostic, drug efficacy, prognostic or disease progression evaluation potentials.

Considering that the samples were from different cell populations mixed together, it is difficult to speculate on the basis of the changes in gene expression observed. It is possible that some of the observed changes are due to changes in lymphocyte behavior, or changes in the proportions of different cell types in the blood of patients. Furthermore, the fact that some down regulated gene were identified as target of some up regulated miRNAs is not a direct implication of any cause-effect relationship, as it is possible, that the miRNA and the gene are from different cell populations. This is also true for genes whose targeting miRNAs were not seen to be differentially regulated or even reverse trends in miRNA-mRNA expression patterns.

As mentioned earlier, it is difficult to hypothesize on any cause-effect relationship whatsoever, since cell populations were used. However, it can clearly be seen that key players of innate and adaptive immunity were differentially modulate.

Table 7: Differentially expressed mRNAs targeted by dysregulated miRNAs

MiRNA	Predicting Algorithm	Target Symbol
hsa-miR-98	TargetScan Human	ACSL6
hsa-miR-98	Ingenuity Expert Findings,TargetScan Human	BCL2L1
hsa-miR-195	TargetScan Human	PIK3R1
hsa-miR-195	TargetScan Human	*UBE2V1
hsa-miR-193b	TargetScan Human	EIF2S1
hsa-miR-193b	TargetScan Human	*MAPK10
hsa-miR-193b	TargetScan Human	PLA2G4D
hsa-miR-199a-3p	TargetScan Human	CXCL11
hsa-miR-199a-3p	TarBase	*MET
hsa-miR-27b	TargetScan Human	CXCL11
hsa-miR-27b	TargetScan Human	FZD7
hsa-miR-27b	TargetScan Human	*MAPK10
hsa-miR-27b	TargetScan Human	*MET
hsa-miR-27b	TargetScan Human	*RFXAP
hsa-miR-27b	TargetScan Human	*UBE2V1
hsa-miR-454	TargetScan Human	*MAPK10
hsa-miR-454	TargetScan Human	*MET
hsa-miR-454	TargetScan Human	*RFXAP
hsa-miR-454	TargetScan Human	*TIMP2
hsa-miR-4291	TargetScan Human	*TIMP2
hsa-miR-4291	TargetScan Human	*UBE2V1

*= Targeted by more than one miRNA

Table 8: Dysregulated targets and related genes involved in infectious diseases

Symbol	Entrez Gene Name	Application(s)
ACSL6	acyl-CoA synthetase long-chain family member	
BAD-Bcl-XL		
BCL2L1	BCL2-like 1	Efficacy,Prognosis
CXCL11	chemokine (C-X-C motif) ligand 11	
EIF2S1	eukaryotic translation initiation factor 2, subunit	Efficacy
FZD7	frizzled family receptor 7	Disease Progression
MAPK10	mitogen-activated protein kinase 10	
Met2	methyltransferase 2	
MET	met proto-oncogene (hepatocyte growth factor receptor)	Diagnosis, Disease Progression,
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1	
PKR dimer	--	
PLA2G4D	phospholipase A2, group IVD (cytosolic)	
RFXAP	regulatory factor X-associated protein	
TIMP2	TIMP metallopeptidase inhibitor 2	Efficacy,Prognosis Unspecified Application
UBE2V1	ubiquitin-conjugating enzyme E2 variant 1	
ACSL6	acyl-CoA synthetase long-chain family member 6	
CCDC80	coiled-coil domain containing 80	
Frizzled	--	Diagnosis
FZD7	frizzled family receptor 7	Disease Progression
MYOC	myocilin, trabecular meshwork inducible glucocorticoid response	
PLA2	--	Efficacy
PLA2G4D	phospholipase A2, group IVD (cytosolic)	
ROR2	receptor tyrosine kinase-like orphan receptor 2	
WNT3	wingless-type MMTV integration site family	
WNT7A	wingless-type MMTV integration site family	

4 Conclusion

Analysis of the expression patterns of circulating and blood cell-born miRNA and mRNA was performed using Agilent and illumine platforms respectively. A total of 13 miRNAs were found to be differentially expressed between patients and controls. None of the analyzed miRNA was able to show a clear expression profile with seropositive subjects. Meanwhile, three miRNAs were able to distinguish between patients and controls with a close to 100% selectivity. However, these miRNA were also already reported to have similar expression patterns in other infectious and non-infectious diseases. They are therefore not suited to be designated as potential disease-specific biomarkers at this level.

Most of the differentially regulated miRNAs have been reported to be involved in host defense mechanisms, and pathway analysis of differentially expressed mRNA targeted by the differentially regulated miRNA also points to this direction. Gene expression analyses led to the identification of 21 transcripts that are potential targets of the differentially regulated miRNAs, and are involved in infection. Like some of the miRNAs, some of the genes were previously reported to have biomarker applications. Considering that a mixture of cell populations was used, and compounded by the fact that some of the transcripts might have been from other sources, it is not possible to correlate miRNA/mRNA expression patterns.

The differential expression observed here cannot entirely be attributed to HAT, even if it is a major player, as it cannot be excluded that most of the patients might be suffering from other diseases and infections. The seropositive group remains a very heterogenous group, and differentially expressed miRNA alone cannot be considered reliable enough for the classification of individuals in this group. Besides, more scrutiny is needed before the initial classification of subjects in this group using established serological tests. It is possible that combining the immune trypanolysis test and miRNA expression profile could help improve the classification. The root cause of the observed differential expression can be assumed to be resulting from host immune responses against microbial infection. Limited sample size might have played a considerable role in the outcome of this study.Therefore, on the basis of the effect sizes observed in this study,future efforts to identify less invasive biomarkers for HAT should make an effort to consider sample size in the case where this is possible and, in addition, implement cell sorting.

References

1. Funk, S., et al., *Identifying Transmission Cycles at the Human-Animal Interface: The Role of Animal Reservoirs in Maintaining Gambiense Human African Trypanosomiasis.* Plos Computational Biology, 2013. 9(1).
2. Hamill, L.C., *et al*., *Domestic pigs as potential reservoirs of human and animal trypanosomiasis in Northern Tanzania.* Parasites & Vectors, 2013. 6.
3. Simarro, P.P., *et al*., *The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward.* PLoS Negl Trop Dis, **2011**. 5(2): p. e1007.
4. Simarro, P.P., *et al*., *Estimating and Mapping the Population at Risk of Sleeping Sickness.* Plos Neglected Tropical Diseases, 2012. 6(10).
5. Dumas, M. and B. Bouteille, *[Human African trypanosomiasis]*. C R Seances Soc Biol Fil, **1996**. 190(4): p. 395-408.
6. Jamonneau, V., *et al*., *Characterization of Trypanosoma brucei s.l. infecting asymptomatic sleeping-sickness patients in Cote d'Ivoire: a new genetic group?* Annals of Tropical Medicine and Parasitology, 2004. 98(4): p. 329-337.
7. Jamonneau, V., *et al*., *Untreated human infections by Trypanosoma brucei gambiense are not 100% fatal.* PLoS Negl Trop Dis, 2012. 6(6): p. e1691.
8. Jamonneau, V., *et al*., *Revisiting the immune trypanolysis test to optimise epidemiological surveillance and control of sleeping sickness in West Africa.* PLoS Negl Trop Dis, 2010. 4(12): p. e917.
9. Njiru, Z.K., *et al*., *African trypanosomiasis: sensitive and rapid detection of the sub-genus Trypanozoon by loop-mediated isothermal amplification (LAMP) of parasite DNA.* Int J Parasitol, **2008**. 38(5): p. 589-99.
10. Chappuis, F., *et al*., *Options for field diagnosis of human african trypanosomiasis.* Clin Microbiol Rev, 2005. 18(1): p. 133-46.
11. Burchmore, R., *Parasites in the brain? The search for sleeping sickness biomarkers.* Expert Rev Anti Infect Ther, 2012. 10(11): p. 1283-6.
12. Tiberti, N., *et al*., *Cerebrospinal Fluid Neopterin as Marker of the Meningo-Encephalitic Stage of Trypanosoma brucei gambiense Sleeping Sickness.* Plos One, 2012. 7(7).
13. Ilboudo, H., *et al*., *Diversity of response to Trypanosoma brucei gambiense infections in the Forecariah mangrove focus (Guinea): perspectives for a better control of sleeping sickness.* Microbes and Infection, 2011. 13(11): p. 943-952.
14. Wastling, S.L., *et al*., *Latent Trypanosoma brucei gambiense foci in Uganda: a silent epidemic in children and adults?* Parasitology, 2011. 138(12): p. 1480-1487.
15. Camara, M., *et al*., *Sleeping sickness diagnosis: use of buffy coats improves the sensitivity of the mini anion exchange centrifugation test.* Tropical Medicine & International Health, 2010. 15(7): p. 796-799.
16. Sternberg, J.M. and L. MacLean, *A spectrum of disease in Human African trypanosomiasis: the host and parasite genetics of virulence.* Parasitology, 2010. 137(14): p. 2007-2015.

17. Sternberg, J.M. and L. Maclean, *A spectrum of disease in human African trypanosomiasis: the host and parasite genetics of virulence.* Parasitology, 2010. 137(14): p. 2007-15.

18. Naessens, J., *Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?* Int J Parasitol, 2006. 36(5): p. 521-8.

19. Simarro, P.P., *et al*., *Attitude towards CATT-positive individuals without parasitological confirmation in the African Trypanosomiasis (T.b. gambiense) focus of Quicama (Angola).* Trop Med Int Health, 1999. 4(12): p. 858-61.

20. Chappuis, F., et al., *Card agglutination test for trypanosomiasis (CATT) end-dilution titer and cerebrospinal fluid cell count as predictors of human African Trypanosomiasis (Trypanosoma brucei gambiense) among serologically suspected individuals in southern Sudan.* Am J Trop Med Hyg, 2004. 71(3): p. 313-7.

21. Krishna, N.K. and K.M. Cunnion, *Role of Molecular Diagnostics in the Management of Infectious Disease Emergencies.* Medical Clinics of North America, 2012. 96(6): p. 1067-+.

22. Levy, P.Y. and F. Fenollar, *The role of molecular diagnostics in implant-associated bone and joint infection.* Clin Microbiol Infect, 2012. 18(12): p. 1168-75.

23. Gandhi, R., et al., *Circulating MicroRNAs as biomarkers for disease staging in multiple sclerosis.* Annals of Neurology, 2013. 73(6): p. 729-740.

24. Huntzinger, E. and E. Izaurralde, *Gene silencing by microRNAs: contributions of translational repression and mRNA decay.* Nat Rev Genet, 2011. 12(2): p. 99-110.

25. Djuranovic, S., A. Nahvi, and R. Green, *A Parsimonious Model for Gene Regulation by miRNAs.* Science, 2011. 331(6017): p. 550-553.

26. Bushati, N. and S.M. Cohen, *MicroRNA functions.* Annual Review of Cell and Developmental Biology, 2007. 23: p. 175-205.

27. Amiel, J., L. de Pontual, and A. Henrion-Caude, *miRNA, Development and Disease.* Advances in Genetics, Vol 80, 2012. 80: p. 1-36.

28. Feng, Y.L. and X.Y. Yu, *Cardinal roles of miRNA in cardiac development and disease.* Science China-Life Sciences, 2011. 54(12): p. 1113-1120.

29. Zhang, W.C., et al., *The role of microRNAs in lung cancer progression.* Medical Oncology, 2013. 30(3).

30. Keller, A., et al., *Toward the blood-borne miRNome of human diseases.* Nature Methods, 2011. 8(10): p. 841-U104.

31. Delic, D., et al., *Hepatic miRNA expression reprogrammed by Plasmodium chabaudi malaria.* Parasitology Research, 2011. 108(5): p. 1111-1121.

32. El-Assaad, F., et al., *Differential MicroRNA Expression in Experimental Cerebral and Noncerebral Malaria.* Infection and Immunity, 2011. 79(6): p. 2379-2384.

33. Gong, A.Y., et al., *Cryptosporidium parvum Induces B7-H1 Expression in Cholangiocytes by Down-Regulating MicroRNA-513.* Journal of Infectious Diseases, 2010. 201(1): p. 160-169.

34. Morrison, L.J., et al., *Role for Parasite Genetic Diversity in Differential Host Responses to Trypanosoma brucei Infection.* Infection and Immunity, 2010. 78(3): p. 1096-1108.

35. Zeiner, G.M., et al., *Toxoplasma gondii Infection Specifically Increases the Levels of Key Host MicroRNAs.* Plos One, 2010. 5(1).

36. Kennedy, P.G.E., *Human African trypanosomiasis of the CNS: current issues and challenges.* Journal of Clinical Investigation, 2004. 113(4): p. 496-504.

37. Kuzoe, F.A., *Current situation of African trypanosomiasis.* Acta Trop, 1993. 54(3-4): p. 153-62.

38. Kennedy, P.G., *Human African trypanosomiasis-neurological aspects.* J Neurol, 2006. 253(4): p. 411-6.

39. Borst, P. and A.H. Fairlamb, *Surface receptors and transporters of Trypanosoma brucei.* Annu Rev Microbiol, 1998. 52: p. 745-78.

40. Hutchinson, O.C., et al., *Variant Surface Glycoprotein gene repertoires in Trypanosoma brucei have diverged to become strain-specific.* Bmc Genomics, 2007. 8.

41. Borst, P., *Antigenic variation and allelic exclusion.* Cell, 2002. 109(1): p. 5-8.

42. Vanhamme, L., et al., *An update on antigenic variation in African trypanosomes.* Trends Parasitol, 2001. 17(7): p. 338-43.

43. Morrison, L.J., L. Marcello, and R. McCulloch, *Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity.* Cell Microbiol, 2009. 11(12): p. 1724-34.

44. Boa, Y.F., et al., *[The different present-day clinical picture of human African trypanosomiasis caused by T. b. gambiense. Analysis of 300 cases from a focus in Daloa, Ivory Coast]*. Bull Soc Pathol Exot Filiales, 1988. 81(3 Pt 2): p. 427-44.

45. Greenwood, B.M. and H.C. Whittle, *The pathogenesis of sleeping sickness.* Trans R Soc Trop Med Hyg, 1980. 74(6): p. 716-25.

46. MacLean, L., et al., *Stage Progression and Neurological Symptoms in Trypanosoma brucei rhodesiense Sleeping Sickness: Role of the CNS Inflammatory Response.* Plos Neglected Tropical Diseases, 2012. 6(10).

47. Greenwood, B.M. and H.C. Whittle, *The Pathogenesis of Sleeping Sickness.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1980. 74(6): p. 716-725.

48. Odiit, M., F. Kansiime, and J.C.K. Enyaru, *Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda.* East African Medical Journal, 1997. 74(12): p. 792-795.

49. Stich, A., P.M. Abel, and S. Krishna, *Human African trypanosomiasis.* British Medical Journal, 2002. 325(7357): p. 203-206.

50. Robays, J., et al., *The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo.* Trop Med Int Health, 2004. 9(5): p. 542-50.

51. *Control and surveillance of African trypanosomiasis. Report of a WHO Expert Committee.* World Health Organ Tech Rep Ser, 1998. 881: p. I-VI, 1-114.

52. World Health, O., *Control and surveillance of human African trypanosomiasis.* World Health Organ Tech Rep Ser, 2013(984): p. 1-237.

53. Doua, F., et al., *The efficacy of pentamidine in the treatment of early-late stage Trypanosoma brucei gambiense trypanosomiasis.* Am J Trop Med Hyg, 1996. 55(6): p. 586-8.

54. Lejon, V., et al., *Intrathecal immune response pattern for improved diagnosis of central nervous system involvement in trypanosomiasis.* J Infect Dis, 2003. 187(9): p. 1475-83.

55. Stanghellini, A. and T. Josenando, *The situation of sleeping sickness in Angola: a calamity.* Trop Med Int Health, 2001. 6(5): p. 330-4.

56. Jamonneau, V., et al., *Preliminary evaluation of LATEX/T. b. gambiense and alternative versions of CATT/T. b. gambiense for the serodiagnosis of human african trypanosomiasis of a population at risk in Cote d'Ivoire: considerations for mass-screening.* Acta Trop, 2000. 76(2): p. 175-83.

57. Lejon, V., et al., *A semi-quantitative ELISA for detection of Trypanosoma brucei gambiense specific antibodies in serum and cerebrospinal fluid of sleeping sickness patients.* Acta Trop, 1998. 69(2): p. 151-64.

58. Noireau, F., et al., *Serodiagnosis of sleeping sickness in the Republic of the Congo: comparison of indirect immunofluorescent antibody test and card agglutination test.* Trans R Soc Trop Med Hyg, 1988. 82(2): p. 237-40.

59. Asonganyi, T., et al., *A multi-centre evaluation of the card indirect agglutination test for trypanosomiasis (TrypTect CIATT).* Ann Trop Med Parasitol, 1998. 92(8): p. 837-44.

60. Nantulya, V.M., *TrypTect CIATT--a card indirect agglutination trypanosomiasis test for diagnosis of Trypanosoma brucei gambiense and T. b. rhodesiense infections.* Trans R Soc Trop Med Hyg, 1997. 91(5): p. 551-3.

61. Lejon, V., J. Kwete, and P. Buscher, *Towards saliva-based screening for sleeping sickness?* Trop Med Int Health, 2003. 8(7): p. 585-8.

62. Komba, E., et al., *Multicentre evaluation of an antigen-detection ELISA for the diagnosis of Trypanosoma brucei rhodesiense sleeping sickness.* Bull World Health Organ, 1992. 70(1): p. 57-61.

63. Akol, M.N., et al., *Trypanosomosis agglutination card test for Trypanosoma brucei rhodesiense sleeping sickness.* East Afr Med J, 1999. 76(1): p. 38-41.

64. Sinha, A., et al., *African trypanosomiasis in two travelers from the United States.* Clin Infect Dis, 1999. 29(4): p. 840-4.

65. Simarro, P.P., F.J. Louis, and J. Jannin, *[Sleeping sickness, forgotten illness: what are the consequences in the field?]*. Med Trop (Mars), 2003. 63(3): p. 231-5.

66. Bailey, J.W. and D.H. Smith, *The Use of the Acridine-Orange Qbc(R) Technique in the Diagnosis of African Trypanosomiasis.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1992. 86(6): p. 630-630.

67. Woo, P.T., *Evaluation of the haematocrit centrifuge and other techniques for the field diagnosis of human trypanosomiasis and filariasis.* Acta Trop, 1971. 28(3): p. 298-303.

68. Matovu, E., et al., *Phase II Evaluation of Sensitivity and Specificity of PCR and NASBA Followed by Oligochromatography for Diagnosis of Human African*

Trypanosomiasis in Clinical Samples from D. R. Congo and Uganda. Plos Neglected Tropical Diseases, 2010. 4(7).

69. Radwanska, M., et al., *Novel primer sequences for polymerase chain reaction-based detection of Trypanosoma brucei gambiense.* American Journal of Tropical Medicine and Hygiene, 2002. 67(3): p. 289-295.
70. Becker, S., et al., *Real-time PCR for detection of Trypanosoma brucei in human blood samples.* Diagnostic Microbiology and Infectious Disease, 2004. 50(3): p. 193-199.
71. Kyambadde, J.W., et al., *Detection of trypanosomes in suspected sleeping sickness patients in Uganda using the polymerase chain reaction.* Bulletin of the World Health Organization, 2000. 78(1): p. 119-124.
72. Mugasa, C.M., et al., *Diagnostic Accuracy of Molecular Amplification Tests for Human African Trypanosomiasis-Systematic Review.* Plos Neglected Tropical Diseases, 2012. 6(1).
73. Schijman, A.G., et al., *International Study to Evaluate PCR Methods for Detection of Trypanosoma cruzi DNA in Blood Samples from Chagas Disease Patients.* Plos Neglected Tropical Diseases, 2011. 5(1).
74. Njiru, Z.K., et al., *African trypanosomiasis: Sensitive and rapid detection of the sub-genus Trypanozoon by loop-mediated isothermal amplification (LAMP) of parasite DNA.* International Journal for Parasitology, 2008. 38(5): p. 589-599.
75. Mugasa, C.M., G.J. Schoone, and H. Schallig, *Detection of Trypanosoma brucei parasites in blood samples using real-time nucleic acid sequence-based amplification (NASBA).* Tropical Medicine & International Health, 2009. 14: p. 175-175.
76. Deborggraeve, S., et al., *Molecular dipstick test for diagnosis of sleeping sickness.* Journal of Clinical Microbiology, 2006. 44(8): p. 2884-2889.
77. Alsford, S., et al., *High-throughput decoding of antitrypanosomal drug efficacy and resistance.* Nature, 2012. 482(7384): p. 232-6.
78. Wery, M., *Drug used in the treatment of sleeping sickness (human African trypanosomiasis: HAT).* Int J Antimicrob Agents, 1994. 4(3): p. 227-38.
79. Paine, M.F., et al., *Diamidines for human African trypanosomiasis.* Curr Opin Investig Drugs, 2010. 11(8): p. 876-83.
80. Blum, J., S. Nkunku, and C. Burri, *Clinical description of encephalopathic syndromes and risk factors for their occurrence and outcome during melarsoprol treatment of human African trypanosomiasis.* Tropical Medicine & International Health, 2001. 6(5): p. 390-400.
81. Priotto, G., et al., *Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial.* Lancet, 2009. 374(9683): p. 56-64.
82. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic Acids Research, 2008. 36: p. D154-D158.
83. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* Cell, 1993. 75(5): p. 855-62.

84. Hong, J., et al., *MicroRNA function is required for neurite outgrowth of mature neurons in the mouse postnatal cerebral cortex.* Frontiers in Cellular Neuroscience, 2013. 7.

85. Polesskaya, A., et al., *Genome-Wide Exploration of miRNA Function in Mammalian Muscle Cell Differentiation.* Plos One, 2013. 8(8).

86. Pickering, M.T., B.M. Stadler, and T.F. Kowalik, *miR-17 and miR-20a temper an E2F1-induced G1 checkpoint to regulate cell cycle progression.* Oncogene, 2009. 28(1): p. 140-5.

87. Ji, Q., et al., *MicroRNA miR-34 inhibits human pancreatic cancer tumor-initiating cells.* PLoS One, 2009. 4(8): p. e6816.

88. Lu, T.X., A. Munitz, and M.E. Rothenberg, *MicroRNA-21 Is Up-Regulated in Allergic Airway Inflammation and Regulates IL-12p35 Expression.* Journal of Immunology, 2009. 182(8): p. 4994-5002.

89. Chang, T.C., et al., *Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis.* Molecular Cell, 2007. 26(5): p. 745-752.

90. Krutzfeldt, J., et al., *Silencing of microRNAs in vivo with 'antagomirs'.* Nature, 2005. 438(7068): p. 685-9.

91. Majumder, S., et al., *MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27(kip1).* Cancer Research, 2009. 69(2): p. 211s-211s.

92. Tsang, J.S., M.S. Ebert, and A. van Oudenaarden, *Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures.* Mol Cell, 2010. 38(1): p. 140-53.

93. Macfarlane, L.A. and P.R. Murphy, *MicroRNA: Biogenesis, Function and Role in Cancer.* Curr Genomics, 2010. 11(7): p. 537-61.

94. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units.* Genome Res, 2004. 14(10A): p. 1902-10.

95. Saini, H.K., S. Griffiths-Jones, and A.J. Enright, *Genomic analysis of human microRNA transcripts.* Proc Natl Acad Sci U S A, 2007. 104(45): p. 17719-24.

96. Cramer, P., *Structure and function of RNA polymerase II.* Adv Protein Chem, 2004. 67: p. 1-42.

97. Gu, T.J., et al., *Alu-directed transcriptional regulation of some novel miRNAs.* Bmc Genomics, 2009. 10.

98. Han, L., et al., *DNA methylation regulates microRNA expression.* Cancer Biology & Therapy, 2007. 6(8): p. 1284-1288.

99. Yang, X.L., et al., *Both mature miR-17-5p and passenger strand miR-17-3p target TIMP3 and induce prostate tumor growth and invasion.* Nucleic Acids Research, 2013. 41(21): p. 9688-9704.

100. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* Embo Journal, 2002. 21(17): p. 4663-4670.

101. Sempere, L.F., et al., *Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation.* Genome Biology, 2004. 5(3).

102. He, C.L., et al., *Functional elucidation of MiR-34 in osteosarcoma cells and primary tumor samples.* Biochemical and Biophysical Research Communications, 2009. 388(1): p. 35-40.

103. Yang, Y.Y., et al., *The Bantam microRNA Is Associated with Drosophila Fragile X Mental Retardation Protein and Regulates the Fate of Germline Stem Cells.* Plos Genetics, 2009. 5(4).

104. Xiao, C.C. and K. Rajewsky, *MicroRNA Control in the Immune System: Basic Principles (vol 136, pg 26, 2009).* Cell, 2009. 137(2): p. 380-380.

105. Wang, S.S., et al., *The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis.* Developmental Cell, 2008. 15(2): p. 261-271.

106. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nature Cell Biology, 2007. 9(6): p. 654-U72.

107. Yu, Z.R., T. Raabe, and N.B. Hecht, *MicroRNA Mirn122a reduces expression of the posttranscriptionally regulated germ cell transition protein 2 (Tnp2) messenger RNA (mRNA) by mRNA cleavage.* Biology of Reproduction, 2005. 73(3): p. 427-433.

108. Mishima, Y., et al., *Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430.* Current Biology, 2006. 16(21): p. 2135-2142.

109. O'Connell, R.M., et al., *MicroRNA-155 is induced during the macrophage inflammatory response.* Proceedings of the National Academy of Sciences of the United States of America, 2007. 104(5): p. 1604-1609.

110. Williams, A.E., *Functional aspects of animal microRNAs.* Cell Mol Life Sci, 2008. 65(4): p. 545-62.

111. Pedersen, I.M., et al., *Interferon modulation of cellular microRNAs as an antiviral mechanism.* Nature, 2007. 449(7164): p. 919-U13.

112. Thum, T., et al., *MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts.* Nature, 2008. 456(7224): p. 980-4.

113. Shenouda, S.K. and S.K. Alahari, *MicroRNA function in cancer: oncogene or a tumor suppressor?* Cancer Metastasis Rev, 2009. 28(3-4): p. 369-78.

114. Saito, Y., H. Suzuki, and T. Hibi, *The role of microRNAs in gastrointestinal cancers.* J Gastroenterol, 2009. 44 Suppl 19: p. 18-22.

115. Mitchell, P.S., et al., *Circulating microRNAs as stable blood-based markers for cancer detection.* Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(30): p. 10513-10518.

116. Pritchard, C.C., et al., *Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies.* Cancer Prevention Research, 2012. 5(3): p. 492-497.

117. Wittmann, J. and H.M. Jack, *Serum microRNAs as powerful cancer biomarkers.* Biochimica Et Biophysica Acta-Reviews on Cancer, 2010. 1806(2): p. 200-207.

118. Singh, Y., et al., *Mycobacterium tuberculosis Controls MicroRNA-99b (miR-99b) Expression in Infected Murine Dendritic Cells to Modulate Host Immunity.* Journal of Biological Chemistry, 2013. 288(7): p. 5056-5061.

119. Gupta, A., et al., *MicroRNAs, Hepatitis C Virus, and HCV/HIV-1 Co-Infection: New Insights in Pathogenesis and Therapy.* Viruses-Basel, 2012. 4(11): p. 2485-2513.

120. Lemaire, J., et al., *MicroRNA Expression Profile in Human Macrophages in Response to Leishmania major Infection.* Plos Neglected Tropical Diseases, 2013. 7(10).

121. Zhou, R., et al., *NF-kappaB p65-Dependent Transactivation of miRNA Genes following Cryptosporidium parvum Infection Stimulates Epithelial Cell Immune Responses.* Plos Pathogens, 2009. 5(12).

122. Jamonneau, V., et al., *Revisiting the Immune Trypanolysis Test to Optimise Epidemiological Surveillance and Control of Sleeping Sickness in West Africa.* Plos Neglected Tropical Diseases, 2010. 4(12).

123. Stelzer, G., et al., *In-silico human genomics with GeneCards.* Hum Genomics, 2011. 5(6): p. 709-17.

124. Eberwine, J., et al., *Analysis of Gene-Expression in Single Live Neurons.* Proceedings of the National Academy of Sciences of the United States of America, 1992. 89(7): p. 3010-3014.

125. Camara, O., et al., *Immune trypanolysis test with blood spotted on filter paper for epidemiological surveillance of sleeping sickness.* Trop Med Int Health, 2014.

126. Salipante, S.J., et al., *Coinfection of Fusobacterium nucleatum and Actinomyces israelii in mastoiditis diagnosed by next-generation DNA sequencing.* J Clin Microbiol, 2014. 52(5): p. 1789-92.

127. Chen, J., et al., *MicroRNA-193b represses cell proliferation and regulates cyclin D1 in melanoma.* Am J Pathol, 2010. 176(5): p. 2520-9.

128. Xu, C., et al., *MicroRNA-193b regulates proliferation, migration and invasion in human hepatocellular carcinoma cells.* Eur J Cancer, 2010. 46(15): p. 2828-36.

129. Braconi, C., et al., *Hepatitis C virus proteins modulate microRNA expression and chemosensitivity in malignant hepatocytes.* Clin Cancer Res, 2010. 16(3): p. 957-66.

130. Duan, Z.F., et al., *MicroRNA-199a-3p Is Downregulated in Human Osteosarcoma and Regulates Cell Proliferation and Migration.* Molecular Cancer Therapeutics, 2011. 10(8): p. 1337-1345.

131. Wu, D., et al., *MicroRNA-199a-3p regulates endometrial cancer cell proliferation by targeting mammalian target of rapamycin (mTOR).* Int J Gynecol Cancer, 2013. 23(7): p. 1191-7.

132. Shatseva, T., et al., *MicroRNA miR-199a-3p regulates cell proliferation and survival by targeting caveolin-2.* J Cell Sci, 2011. 124(Pt 16): p. 2826-36.

133. Li, C., et al., *miRNA-199a-3p in plasma as a potential diagnostic biomarker for gastric cancer.* Ann Surg Oncol, 2013. 20 Suppl 3: p. S397-405.

134. Zhang, G.L., et al., *Suppression of hepatitis B virus replication by microRNA-199a-3p and microRNA-210.* Antiviral Res, 2010. 88(2): p. 169-75.

135. Buck, A.H., et al., *Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection.* RNA, 2010. 16(2): p. 307-15.

136. Long, G.W., et al., *Human Circulating MicroRNA-1 and MicroRNA-126 as Potential Novel Indicators for Acute Myocardial Infarction.* International Journal of Biological Sciences, 2012. 8(6): p. 811-818.

137. Schober, A., et al., *MicroRNA-126-5p promotes endothelial proliferation and limits atherosclerosis by suppressing Dlk1.* Nature Medicine, 2014. 20(4): p. 368-+.

138. Meister, J. and M.H.H. Schmidt, *miR-126 and miR-126*: New Players in Cancer.* Thescientificworldjournal, 2010. 10: p. 2090-2100.

139. Hu, G.K., et al., *MicroRNA-98 and let-7 Confer Cholangiocyte Expression of Cytokine-Inducible Src Homology 2-Containing Protein in Response to Microbial Challenge.* Journal of Immunology, 2009. 183(3): p. 1617-1624.

140. Huang, S.D., et al., *MicroRNA-98 and microRNA-214 post-transcriptionally regulate enhancer of zeste homolog 2 and inhibit migration and invasion in human esophageal squamous cell carcinoma.* Molecular Cancer, 2012. 11.

141. Xu, X., et al., *MicroRNA-409-3p inhibits migration and invasion of bladder cancer cells via targeting c-Met.* Molecules and Cells, 2013. 36(1): p. 62-68.

142. Li, C.L., et al., *MicroRNA-409-3p regulates cell proliferation and apoptosis by targeting PHF10 in gastric cancer.* Cancer Letters, 2012. 320(2): p. 189-197.

143. Zheng, B., et al., *MicroRNA-409 suppresses tumour cell invasion and metastasis by directly targeting radixin in gastric cancers.* Oncogene, 2012. 31(42): p. 4509-4516.

144. Jia, L.F., et al., *Prognostic Implications of MicoRNA miR-195 Expression in Human Tongue Squamous Cell Carcinoma.* Plos One, 2013. 8(2).

145. Wang, Y., et al., *MiR-195 inhibits proliferation and growth and induces apoptosis of endometrial stromal cells by targeting FKN.* International Journal of Clinical and Experimental Pathology, 2013. 6(12): p. 2824-2834.

146. Hu, Y.W., et al., *An agomir of miR-144-3p accelerates plaque formation through impairing reverse cholesterol transport and promoting pro-inflammatory cytokine production.* PLoS One, 2014. 9(4): p. e94997.

147. Liu, L., et al., *The Oncogenic Role of microRNA-130a/301a/454 in Human Colorectal Cancer via Targeting Smad4 Expression.* Plos One, 2013. 8(2).

148. Hulsmans, M., et al., *Decrease of miR-146b-5p in Monocytes during Obesity Is Associated with Loss of the Anti-Inflammatory but Not Insulin Signaling Action of Adiponectin.* Plos One, 2012. 7(2).

149. Kutty, R.K., et al., *Differential regulation of microRNA-146a and microRNA-146b-5p in human retinal pigment epithelial cells by interleukin-1 beta, tumor necrosis factor-alpha, and interferon-gamma.* Molecular Vision, 2013. 19: p. 737-750.

150. Li, Y.Y., et al., *miR-146b-5p inhibits glioma migration and invasion by targeting MMP16.* Cancer Letters, 2013. 339(2): p. 260-269.

151. Kos, A., et al., *A Potential Regulatory Role for Intronic microRNA-338-3p for Its Host Gene Encoding Apoptosis-Associated Tyrosine Kinase.* Plos One, 2012. 7(2).

152. Xu, D., et al., *miR-22 represses cancer progression by inducing cellular senescence.* Journal of Cell Biology, 2011. 193(2): p. 409-424.

153. Tsuchiya, N., et al., *Tumor Suppressor miR-22 Determines p53-Dependent Cellular Fate through Post-transcriptional Regulation of p21.* Cancer Research, 2011. 71(13): p. 4628-4639.

154. Song, S.J., et al., *The Oncogenic MicroRNA miR-22 Targets the TET2 Tumor Suppressor to Promote Hematopoietic Stem Cell Self-Renewal and Transformation.* Cell Stem Cell, 2013. 13(1): p. 87-101.

155. Liu, Y., et al., *Modulation of T cell cytokine production by miR-144* with elevated expression in patients with pulmonary tuberculosis.* Mol Immunol, 2011. 48(9-10): p. 1084-90.

Index I: Primers and Oligonucleotides

Gene ID	Sense EcoR1	Antisense BamH1
tb927.10.1510	gccgaattcatgatgttgtatgagca	acggatccccgctagctgggcacgttaac
tb927.10.1510	gccgaattcatgatggttccgtcggg	acggatccccgctacttcttgccttgagt
tb927.6.850	gccgaattcatgatgaacaatagcaa	acggatccccgtcagtcccctcccaccac
tb927.3.1920	gccgaattcatgatgacaaataataa	acggatccccgtcaacgcagctcattctc
tb927.10.8720	gccgaa ttcatgatgacggatgatgt	acggatccccgtcatgcgcaccatgaaag
tb927.8.1960	gccgaattcatgatggtggtggagta	acggatccccgttactttagttcatccca
tb927.6.600	gccgaattcatgatgatgcagtatgg	acggatccccgtcagctatgacccttttac
tb927.4.410	gccgaattcatgatgcaccaaaccca	acggatccccgtcacgcactggcatcacc
tb927.10.2100	gccgaattcatgatgggaaaggaaaaggtg	acggatccccgtttcttcgaagccttcac
tb09.211.0120	gccgaattcatgatgagcgctaacgatg	acggatccccggttgggatccatcgtc
tb927.11.9700	gccgaattcatgatgccctctattactcaggag	acggatccccgcacttcgctcgctga
tb927.3.1760	gccgaattcatgatggtggtggtgggg	acggatccccgcctttgggaccgctt
tb927.8.5380	gccgaattcatgatgtctcaaaatgaagaagc	acggatccccgccacacggtcgcg

Primer List for Gene Amplification

GeneID	**forward primer**	reverse primer
tb927.10.4430	aaaaagcaggctcagctatgtcgtcggatga	agaaagctgggtctatgttcctttcttgct
tb427_puf11	aaaaagcaggctccaaaatggtatcccctga	agaaagctgggtttatttatgcgatgcagg
tb927.10.12660	aaaaagcaggctcaaacatgtctggttggga	agaaagctgggtctacagcgttggcatgca
tb927.6.820	aaaaagcaggctctacaatggaggccagtgc	agaaagctgggttcatccctttcctgccgcg
tb927.7.4730	aaaaagcaggctcggtaatgcttcgtagggg	agaaagctgggttcactcaccgactgcccc
tb927.10.11760	aaaaagcaggctctaaaatgagttcaaccaa	agaaagctgggttcactcggcatcgaagtg
tb927.1.2600	aaaaagcaggctccattatggaagtacgcga	agaaagctgggtctaacattctccgtcatc
tb927.11.500	aaaaagcaggctcacaaatgtctcaggttcc	agaaagctgggttcatttacgggcaggccg
tb927.11.530	aaaaagcaggctcacaaatggaaccggagtg	agaaagctgggtttaattctgctgctgctg
tb927.11.12100	aaaaagcaggctccagcatgttcaaccagtt	agaaagctgggtttaagcagtaatacacgt
tb927.3.2930	aaaaagcaggctcctgtatgttctaccccaa	agaaagctgggttcaaccagcggcaccgcg
tb927.10.12090	aaaaagcaggctccagtatgccaccgcgggc	agaaagctgggttcaccgttgaacacgttg
tb927.7.320	aaaaagcaggctccaacatgtataccagtgg	agaaagctgggtctaaatcacgcgactctc

tb927.11.12120	aaaaagcaggctccacaatggcagctgtcac	agaaagctgggtttacgacagcgtaggagg
tb927.8.2780	aaaaagcaggctctaagatgggagactcgat	agaaagctgggttcactccattcgaaccgg
tb927.8.4450	aaaaagcaggctcaatcatggggatcaaggc	agaaagctgggtttacgacgggtcgcgctt
tb927.10.13540	aaaaagcaggctcaggaatgaatttttcacc	agaaagctgggtctactcgcccgatccgct
tb927.6.4530	aaaaagcaggctcgcatatgtcctcttctgc	agaaagctgggtctaatacgaatgcaaccg
tb927.7.1180	aaaaagcaggctcccgaatgcaacgatacaa	agaaagctgggttcacctcgcttccaacac
tb927.8.6440	aaaaagcaggctcaacaatggaagacaatct	agaaagctgggttcaacatacaattaaagt
tb927.3.2000	aaaaagcaggctcccatatggatgagaagcc	agaaagctgggttcattggcgagcagcaga
tb927.5.1080	aaaaagcaggctccaccatgctcgacacgag	agaaagctgggtttaatttttttttacttt
tb927.10.11270	aaaaagcaggctcatatatggtggatccttc	agaaagctgggttcagttgtggacttgcac
tb927.7.880	aaaaagcaggctcaacaatgaacaaacttac	agaaagctgggtttacgactcttcatttcc
tb927.7.3730	aaaaagcaggctcgacaatggaaacgcacca	agaaagctgggtttagtttgggctcatgaa
tb927.10.13570	aaaaagcaggctcagcaatggccgatcgagg	agaaagctgggtctaggccacactccggca
tb927.3.1030	aaaaagcaggctcaaatatgccggcaaaaaa	agaaagctgggttcatctccgcagttcctc
tb927.10.13720	aaaaagcaggctcacggatggcgcatgttgg	agaaagctgggttcaacatttatttttcaac
tb927.5.1750	aaaaagcaggctcccttatgtacgcaaaggg	agaaagctgggtttagctgcgagacaccgc
tb927.4.4230	aaaaagcaggctcaaagatgggaaggcggat	agaaagctgggtttagcgatccacgccaaa
tb927.9.4560	aaaaagcaggctccaatatgtatttcccgcg	agaaagctgggtttagaacaaccgaatacc
tb927.8.990	aaaaagcaggctcgaggatgagcacgggtga	agaaagctgggtctatgaattgtgacgcat
tb927.11.3340	aaaaagcaggctcacacatgtccaagttatt	agaaagctgggtttactccagcgcagaaat
tb927.9.12360	aaaaagcaggctcaattatggagggcactgg	agaaagctgggtttatgctgactgatgcgg
tb927.11.16590	aaaaagcaggctcagggatgagcacatcagc	agaaagctgggtctacacccacctcctccc
tb927.11.5850	aaaaagcaggctcagaaatgcctgtgcttcc	agaaagctgggttcacaaggggcgaccacg
tb927.11.15350	aaaaagcaggctcgagcatgtcctccacaca	agaaagctgggtctactcaaggtctgccgc
tb927.6.4440	aaaaagcaggctccgctatgccaacgccaac	agaaagctgggttcagctgcgctgcttctg
tb927.9.13990	aaaaagcaggctccctgatgcaaggaggaaa	agaaagctgggttcatgaagtggatttatt
tb927.9.8740	aaaaagcaggctctcaaatgtacagccaacc	agaaagctgggtttaactccactgctgtgt
tb927.11.14100	aaaaagcaggctccaaaatgcaaacctgtct	agaaagctgggttcactcattctgttcgtc
tb927.6.3480	aaaaagcaggctcgttaatgtggtacagtaa	agaaagctgggtttaccagggtccatatac
tb927.3.3960	aaaaagcaggctcgcacatgatgccgtgcaa	agaaagctgggtttaatccactggtgaagg
tb927.4.400	aaaaagcaggctcgggtatggaaagtatgcc	agaaagctgggtttaggcataaagccttgc
tb927.9.13280	aaaaagcaggctcacgcatgtcgctcaccgc	agaaagctgggtctagtgtcccttctccgg
tb927.11.16020	aaaaagcaggctccgagatggcagcagcgtt	agaaagctgggtttagtttgcctcattgtt
tb927.3.3940	aaaaagcaggctcagtcatggaagttccttt	agaaagctgggtttacaaaaacggcacaaa
tb927.8.6650	aaaaagcaggctcggaaatgacagaccacca	agaaagctgggttcaacctacacgcatggt

tb927.11.950	aaaaagcaggctcaataatgcaggaaggcca	agaaagctgggtttagtacccttgtcgct
tb927.8.710	aaaaagcaggctcggaaatgataactccaga	agaaagctgggtttaacgaccatcgtccat
tb927.3.3670	aaaaagcaggctcaagaatgcgaggaagagg	agaaagctgggtctaagaagatttcttccc
tb927.3.720	aaaaagcaggctcagacatgcagggctattt	agaaagctgggtttatgccatgggcggaac
tb927.1.2200	aaaaagcaggctcacctatgcgtttgtatct	agaaagctgggtctatttctccgcctcgag
tb927.3.740	aaaaagcaggctccgcaatgatgcaacggga	agaaagctgggttcaatgtaacccagtttg
tb927.3.790	aaaaagcaggctccaggatggccgagcggcg	agaaagctgggtctagcgcttccataaccg
tb927.3.5250	aaaaagcaggctccgcaatggggcgtggaaa	agaaagctgggtctattctgtggttagcag
tb927.4.1680	aaaaagcaggctctctaatgcctcccaagcg	agaaagctgggttcagtaaactggttgcca
tb927.5.810	aaaaagcaggctcgaaaatgagcactgcaac	agaaagctgggttcacaaggaaagaaacat
tb927.5.1570	aaaaagcaggctctcgaatggcaaagaagaa	agaaagctgggtctagaaaacataaaggcc
tb927.5.1580	aaaaagcaggctctaatatggcgagaaaccg	agaaagctgggttcacagagggtacacggg
tb927.6.4050	aaaaagcaggctcttttatggggttcccttc	agaaagctgggtttagtgagaagcaccggg
tb927.6.4720	aaaaagcaggctcgaccatgcgaccggagca	agaaagctgggtctaattcgcacgggttat
tb927.7.250	aaaaagcaggctcgaagatgcagaaacagcg	agaaagctgggtttacacccttgcttccat
tb927.7.2140	aaaaagcaggctcaaacatgtccttctcaat	agaaagctgggttcactgccctaaagaagt
tb927.7.2580	aaaaagcaggctcaaatatgcgtcacggagc	agaaagctgggtctatgcatctatcggcga
tb927.7.2660	aaaaagcaggctcaattatgcagctcgtagc	agaaagctgggtttaattcctgcccctccc
tb927.7.2670	aaaaagcaggctccattatgagccatccaca	agaaagctgggttcaatcgctgattcctcc
tb927.7.2680	aaaaagcaggctcggatatgttccttcccga	agaaagctgggttcaaaagggtacgtgtag
tb927.7.4980	aaaaagcaggctcattcatgcctggttgcct	agaaagctgggttcagtaaaccacgatcag
tb927.8.4020	aaaaagcaggctctagaatgaccatctcgtc	agaaagctgggtctaaggcaacgcagaccc
tb927.9.3460	aaaaagcaggctccgggatgggaaggaagga	agaaagctgggtttacttacgccgcggtgc
tb927.9.9450	aaaaagcaggctccccaatgtattctagcga	agaaagctgggttcacgaacgacccggatg
tb927.9.9520	aaaaagcaggctctaacatgacacaggtcat	agaaagctgggttcaccacaacggcacgga
tb927.10.1540	aaaaagcaggctctaacatgcctcccgctga	agaaagctgggtctacgcttttgccgacg
tb927.10.5150	aaaaagcaggctcaaaaatgtatcggccaac	agaaagctgggtttagatggtttcacgagg
tb927.10.5250	aaaaagcaggctccatcatgtctggcactaa	agaaagctgggttcatctctgtttctgcgt
tb927.10.11730	aaaaagcaggctcacaaatggatagctcgct	agaaagctgggttcacggttccccactcgt
tb927.10.12330	aaaaagcaggctcgggaatgcaaccatttgg	agaaagctgggtttagggggattcggcagc
tb927.10.12740	aaaaagcaggctcctggatgacgatggatat	agaaagctgggtttacctcgcgtatggcca
tb927.10.12760	aaaaagcaggctcaaatatgtcatcgcagga	agaaagctgggtttatgacctgcagagaac
tb927.10.12780	aaaaagcaggctccataatgccctcaaagaa	agaaagctgggtttataacccgtaaggatt
tb927.10.12800	aaaaagcaggctccaagatgaccacagcaac	agaaagctgggtttacagctgctggtaagg
tb927.10.14930	aaaaagcaggctctaccatgaacgatgccgg	agaaagctgggtttaaccaccacctcgagt

tb927.10.14950	aaaaagcaggctcgatcatgtacggtaacca	agaaagctgggtttaagtacttccctcaga
tb927.11.1980	aaaaagcaggctccacaatgagcagcgcggt	agaaagctgggttcacagggtcttgggctt
tb927.11.7450	aaaaagcaggctctacgatgagtagggatcg	agaaagctgggtctaaaaacgccgcctgcc
tb927.11.7890	aaaaagcaggctcggatatgtcaatgcacag	agaaagctgggttcagctggatctcccgtc
tb927.11.8470	aaaaagcaggctcaattatgagcgtgggcaa	agaaagctgggtttacctttgctgcgcagt
tb927.11.16550	aaaaagcaggctcgagaatgagtgaagcgaa	agaaagctgggtttacacatatgtaagcaa
tb927.9.10280	aaaaagcaggctctaacatgatcttccccag	agaaagctgggttcacttggctacccactc
tb927.6.3490	aaaaagcaggctcggcaatgcacatttcaac	agaaagctgggttcaatcctccaaccttac
tb927.11.14950	aaaaagcaggctcaggaatggccttcaacca	agaaagctgggtctactgctgcagatggtt
tb927.11.4460	aaaaagcaggctcgcatatgacgacaggaaa	agaaagctgggtttacgctgtagccttctc
tb927.11.4450	aaaaagcaggctcaaagatgtcagaaaaagg	agaaagctgggtctactccttttatcagc
tb927.4.2040	aaaaagcaggctcgacgatgccttcatatcc	agaaagctgggtttactcctcattgccacc
tb927.4.2030	aaaaagcaggctcgacgatgccttcatatcc	agaaagctgggttcagttgccttcacccaa
tb927.2.3880	aaaaagcaggctccgtcatggaacagtctcc	agaaagctgggtttaacagaacatctgtgg
tb927.5.1960	aaaaagcaggctctaaaatgatcaatgaagc	agaaagctgggttcatagttttaatgttct
tb927.3.1910	aaaaagcaggctctagaatgtccatgactga	agaaagctgggtctactgacgctgctggcg
tb927.10.11990	aaaaagcaggctcgtatatgcgtccgttaac	agaaagctgggtctacgtcaagcgtgcctc
tb927.3.3970	aaaaagcaggctcttcaatggtttccttcac	agaaagctgggtttatcccgcctcttcatc
tb927.10.15870	aaaaagcaggctccctgatgaactacgccga	agaaagctgggttcaagatctccttgggac
tb927.7.7280	aaaaagcaggctcgcccatgccaaacagtga	agaaagctgggtctacgttgtttcaatcgt
tb927.9.12900	aaaaagcaggctcgaaaatgtatcgacacgg	agaaagctgggttcactttcgctcatgtag
tb927.7.5380	aaaaagcaggctcaagaatgagcttattttc	agaaagctgggtttacgtaagtgcagcaaa
tb927.7.6410	aaaaagcaggctctcgcatgtcttcgggggc	agaaagctgggttcaagcagtggaagccgc
tb927.9.6870	aaaaagcaggctcagtgatgtctcgagtgta	agaaagctgggttcacggggaacaaggccc
tb927.7.1390	aaaaagcaggctccgtgatgactgtgaacat	agaaagctgggtttagctccaacttgggta
tb927.3.5460	aaaaagcaggctcctccatggagggccgcaa	agaaagctgggttcaggaggaatgcggtag
tb927.10.3500	aaaaagcaggctctcgcatggggcgtgatag	agaaagctgggtttaaccgtcaatacctgc
tb927.5.760	aaaaagcaggctcggagatgagtggaagcgc	agaaagctgggtctaactgctgctgcgctt
tb927.11.7140	aaaaagcaggctcttttatggccgcaacaat	agaaagctgggtctacgaaggtgtttctgg
tb927.3.1340	aaaaagcaggctcaaaaatgcctcccagacg	agaaagctgggtttactccttttcttccac
tb927.10.15170	aaaaagcaggctcaaacatgtcctccctgac	agaaagctgggtctactggttttcctccct
tb927.8.4540	aaaaagcaggctcgaaaatggaggggtctca	agaaagctgggtctatttcccaactcgttt
tb927.9.9060	aaaaagcaggctctttcatgcctgtctgcaa	agaaagctgggttcacttcttggaaagagt
tb927.6.4770	aaaaagcaggctcaatcatgtacccccgaca	agaaagctgggttcaataatatgtttctcg
tb927.10.7440	aaaaagcaggctcgaatatgcaggatgacat	agaaagctgggtctacaggtacagcgacga

tb927.10.3990	aaaaagcaggctctacgatggtaaccgatga	agaaagctgggtttacgcggtgtacagctc
tb927.6.2330	aaaaagcaggctcgacgatgcccgcacgtgg	agaaagctgggtttagcgtagtgcgtacat
tb927.11.550	aaaaagcaggctccgtgatggctgcgacacg	agaaagctgggtttagttgtagcggccgtt
tb927.5.3750	aaaaagcaggctcatcaatgccaagttcgtc	agaaagctgggtttattggcgcaccctctt
tb927.7.4900	aaaaagcaggctcagacatgggtgttccaaa	agaaagctgggttcatcgtttcctgtctcc
tb927.8.2810	aaaaagcaggctcagaaatgggtgtcccatt	agaaagctgggtttaggtccccagggtggc
tb927.5.2450	aaaaagcaggctctagaatgggcgtgccaaa	agaaagctgggtcaaatttgtaacttcacc
tb927.10.6220	aaaaagcaggctcagaaatggggattgccgg	agaaagctgggttcagcgcccaatgactct
tb927.11.1550	aaaaagcaggctctgttatgggcaccaaggg	agaaagctgggttcattttacacctccatc
tb927.10.2220	aaaaagcaggctccgacatgatgcggaggac	agaaagctgggtttaaaagaaccacagtct
tb927.8.2370	aaaaagcaggctcaataatggatgaagttgg	agaaagctgggttcagcggtttgtgcgaaa
tb927.10.10850	aaaaagcaggctctaaaatgtctgactggga	agaaagctgggtttatagataatgcattgt
tb927.11.8290	aaaaagcaggctccacgatggatgaaacacc	agaaagctgggttcaaacgagacttgtggg
tb927.8.3220	aaaaagcaggctccactatggggattaaagg	agaaagctgggtctaccgtcgcgtataaga
tb927.5.1200	aaaaagcaggctcgggaatggcagtcatagt	agaaagctgggtttagactaaccgaattag
tb927.9.7070	aaaaagcaggctcacatatgagtgaaacaaa	agaaagctgggtttactccttgggaaaata
tb927.1.2580	aaaaagcaggctccagaatgcagtttggccc	agaaagctgggtttacccactactccccgc
tb927.10.7450	aaaaagcaggctcgtgtatgcctgtgattaa	agaaagctgggtttactcgatgttacctaa
tb927.2.2180	aaaaagcaggctctctcatggcaacaatatc	agaaagctgggttcattctagggccatgct
tb927.6.670	aaaaagcaggctccgccatgcagtaccaacg	agaaagctgggttcacgtttctgtttttat
tb927.11.16600	aaaaagcaggctcttacatgtcacttccacc	agaaagctgggtttatccttcctttagggg
tb927.7.5460	aaaaagcaggctcaaaaatgcaattcgatga	agaaagctgggttcaggagttgttagttcc
tb427tmp.02.5380	aaaaagcaggctcgaagatgtttacgaaaaa	agaaagctgggtttactgttgcccctcctc
tb927.8.5180	aaaaagcaggctcaccgatggatgcgttgac	agaaagctgggttcaatcctgaggtacgga
tb927.9.3530	aaaaagcaggctcagtaatggcttcggtgaa	agaaagctgggtctaagaagtggtggaagc
tb927.5.4030	aaaaagcaggctctgcaatgtcggcacagca	agaaagctgggtttactggaatgggttctg
tb927.3.1780	aaaaagcaggctctcgaatgctttcccaata	agaaagctgggtctatccaccttttaactga
tb927.10.1510	aaaaagcaggctcgtccatgttgtatgagca	agaaagctgggtctagctgggcacgttaac
tb927.10.1510	aaaaagcaggctcatctatggttccgtcggg	agaaagctgggtctacttcttgccttgagt
tb927.6.850	aaaaagcaggctctcacatgaacaatagcaa	agaaagctgggttcagtcccctcccaccac
tb927.3.1920	aaaaagcaggctcgataatgacaaataataa	agaaagctgggttcaacgcagctcattctc
tb927.10.8720	aaaaagcaggctcagaaatgacggatgatgt	agaaagctgggttcatgcgcaccatgaaag
tb927.8.1960	aaaaagcaggctcgaatatggtggtggagta	agaaagctgggtttactttagttcatccca
tb927.6.600	aaaaagcaggctcgaaaatgatgcagtatgg	agaaagctgggttcagctatgacccttac
tb927.4.410	aaaaagcaggctcgaccatgcaccaaaccca	agaaagctgggttcacgcactggcatcacc

tb927.6.1670	aaaaagcaggctcagaaatgagcgttccgtg	agaaagctgggtctttagcaacctcgcgcc
tb927.11.13970	aaaaagcaggctctgatatggattcaaaggg	agaaagctgggtttacaacgttgccgatgt
tb927.8.2850	aaaaagcaggctcaactatgcaggtgacacg	agaaagctgggttcaccttcttagtgcttt
tb927.10.8360	aaaaagcaggctcggaaatggacgtcgacaa	agaaagctgggtctacccactaagacggta
tb927.9.13510	aaaaagcaggctcaaatatgaagctcattac	agaaagctgggtctacttctcagtcaaatg
tb927.5.2140	aaaaagcaggctcgtttatgttcagtgagca	agaaagctgggttcaagactcaccaggcaa
tb927.10.7080	aaaaagcaggctccggcatgcccgcggagct	agaaagctgggttcaaagaggcctaaagtt
tb927.7.1170	aaaaagcaggctctgttatgattacagaggg	agaaagctgggtctactcacgctcccgttt
tb927.5.1490	aaaaagcaggctcaaacatgaactgctccgc	agaaagctgggtttaatttcccccaaggga
tb927.9.5460	aaaaagcaggctcgaaaatggactcacctga	agaaagctgggtctaaacccgcctcacagc
tb927.8.4500	aaaaagcaggctctacgatggagcaccgcgc	agaaagctgggtttatttagatttctcaaa
tb927.8.4820	aaaaagcaggctcatcaatgcacgtctacac	agaaagctgggttcaagtcaaggtcccctt
tb927.11.10560	aaaaagcaggctcaataatgctgttcaaacc	agaaagctgggttcacgcatatataactga
tb927.9.4680	aaaaagcaggctcaatcatggcccaacaagg	agaaagctgggtttactcaccaaggtaggc
tb927.11.2260	aaaaagcaggctcgtttatgatggctgaatc	agaaagctgggtttaggccttgctagcgcc
tb927.10.16070	aaaaagcaggctccaagatgcagaccttact	agaaagctgggttcacaattgatacaaggg
tb927.11.11770	aaaaagcaggctcttatatgaatccagaagc	agaaagctgggtctaaagcgtgaagtcggg
tb927.6.1870	aaaaagcaggctcgttgatgcaaaatttacg	agaaagctgggttcaccaaagccgacgatt
tb927.11.6160	aaaaagcaggctcggagatgtcaggttgggc	agaaagctgggtttaagctttacctgcgtt
tb927.9.11050	aaaaagcaggctcgacaatgagaactacaat	agaaagctgggtctagcggcgctggttggg
tb927.6.3870	aaaaagcaggctcaagtatggcatcagaccc	agaaagctgggttcaatatggcccaccgcg
tb927.10.1630	aaaaagcaggctctaacatgccacctaaaaa	agaaagctgggtctactcgttgctttcatc
tb927.1.4540	aaaaagcaggctccacgatgtttttgaagga	agaaagctgggtctaagatgccgctagtac
tb927.1.4650	aaaaagcaggctccacgatgtccttggagga	agaaagctgggtctacatctcagcgtaatc
tb927.7.4670	aaaaagcaggctcaatcatgcatcattgtcg	agaaagctgggttcaataaattgtacgctt
tb927.7.930	aaaaagcaggctcagcgatgtctcgtcggcg	agaaagctgggttcaccctatttcagcatt
tb927.6.4960	aaaaagcaggctcagcgatgtcatttacggc	agaaagctgggttcacatgtcttccgtgct
tb927.11.7310	aaaaagcaggctcgccgatgcctagtagggc	agaaagctgggtctaatgctcgctgccaga
tb927.4.1630	aaaaagcaggctcgtatatgtctggctctct	agaaagctgggttcacgatttgtttttctt
tb927.7.7380	aaaaagcaggctcagccatggatcccttcgt	agaaagctgggttcagttaacctgatacga
tb927.3.4210	aaaaagcaggctcaaaaatgatacggccgtg	agaaagctgggttcactcaagaagcccaga
tb927.10.310	aaaaagcaggctctaggatgtggactgtgca	agaaagctgggttcagccggagagcggttg
tb927.11.13960	aaaaagcaggctcaaagatgagtgttaggaa	agaaagctgggtttacttcccgcggttacg
tb927.6.2550	aaaaagcaggctcgaaaatgagcgcaccacc	agaaagctgggtttaccgacgcctcttctt
tb927.4.1310	aaaaagcaggctctgagatgttcggtggtgg	agaaagctgggttcaaaactcttttgtcat

tb927.3.5150	aaaaagcaggctcacttatggtttcttgcaag	agaaagctgggtttactcgcgaaggtgtcg
tb927.11.11030	aaaaagcaggctctaaaatgggggtgccaac	agaaagctgggtctaaaatgcaaactgttc
tb927.11.2370	aaaaagcaggctctaaaatgcccaacccata	agaaagctgggttcactgcgggactgccgc
tb927.6.4950	aaaaagcaggctcggtgatgcaagcggaaga	agaaagctgggtctaccacgccttgatttt
tb927.1.1470	aaaaagcaggctcaatgggtggtattccttc	agaaagctgggttcactgaccccagaagcc
tb927.1.3070	aaaaagcaggctcgatgtctgcgtgggcctc	agaaagctgggttcatggtgcggaagatgt
tb927.1.4020	aaaaagcaggctcgatgcatttcattatttt	agaaagctgggtttaaacttctaaacctga
tb927.1.4280	aaaaagcaggctctatggagaaggtagcaac	agaaagctgggttcacatcactggttcaac
tb927.1.5000	aaaaagcaggctccatgacaactgaagaggg	agaaagctgggtttaccgcagccgaggaac
tb927.1.670	aaaaagcaggctcgatgatgcaagtcaaact	agaaagctgggttcatgcggcagctcgaca
tb927.10.10020	aaaaagcaggctccatgaattctgctgcgag	agaaagctgggtttacgaccgggcgttccg
tb927.10.11600	aaaaagcaggctcgatggagtattcgtcaga	agaaagctgggtctattgataacggtgatg
tb927.10.11600	aaaaagcaggctcgatggagtattcgtcaga	agaaagctgggtctattgataacggtgatg
tb927.10.11630	aaaaagcaggctcaatgagtgccaagttaca	agaaagctgggtttaaactacgatcaaatc
tb927.10.12020	aaaaagcaggctcaatgtctgtgtgtcctcc	agaaagctgggttcagggtaaaaagatgtt
tb927.10.12210	aaaaagcaggctctatggttcaccgtcaaga	agaaagctgggtctaactcgttgtggatcc
tb927.10.1250	aaaaagcaggctcaatgacaagtgtggaaac	agaaagctgggttcagtccaccaactgccc
tb927.10.12720	aaaaagcaggctccatggagacgcacctcgg	agaaagctgggtttagcgccatggtcgagg
tb927.10.13610	aaaaagcaggctcgatgctgtcgtcccgtag	agaaagctgggtctaaaaccgccttcgcgg
tb927.10.13800	aaaaagcaggctcaatgtctgaggctagcga	agaaagctgggtctactgagaacaatcggc
tb927.10.13980	aaaaagcaggctcaatgtcgatgatcgccat	agaaagctgggtctacgatggtccgagact
tb927.10.14150	aaaaagcaggctcaatgagcaagacagagac	agaaagctgggtttacgcaaaatcctcccc
tb927.10.14660	aaaaagcaggctccatggagggcctgtcgcc	agaaagctgggttcatctgtcttcagttct
tb927.10.14720	aaaaagcaggctcaatgtacggtggttatgg	agaaagctgggtctagagttttgtctccct
tb927.10.14810	aaaaagcaggctcgatggctcgcggtcaaat	agaaagctgggttcacttcttcggctgggg
tb927.10.14900	aaaaagcaggctctatggagtacccagcgga	agaaagctgggttcacccccttgcggcgaat
tb927.10.15310	aaaaagcaggctcaatgtatccaaaaaataa	agaaagctgggtttacaaaggggttgtagt
tb927.10.15760	aaaaagcaggctctatgaaatcagaggggac	agaaagctgggtttacgattgctcagcagc
tb927.10.1860	aaaaagcaggctctatgacagataatcaagg	agaaagctgggttcagagatcgatcagctt
tb927.10.2190	aaaaagcaggctctatgttccacaacagtta	agaaagctgggtctaagcgccctctactgg
tb927.10.2240	aaaaagcaggctcgatgtccattacacaaag	agaaagctgggtctagggatagaaccgtgc
tb927.10.240	aaaaagcaggctctatgtctttgctgctgtc	agaaagctgggttcaagctgcctcgccgcc
tb927.10.3140	aaaaagcaggctccatgtacaccgtggctcc	agaaagctgggttcattcaagacacgcgct
tb927.10.370	aaaaagcaggctcgatggaagttgctgcggt	agaaagctgggtttaaaactgcagaacatg
tb927.10.3950	aaaaagcaggctccatggtacagcagggaaa	agaaagctgggtttaaaaagccatgaaaaa

tb927.10.5130	aaaaagcaggctcaatgatccttaaggaaga	agaaagctgggttcatgttcggtcccgcac
tb927.10.550	aaaaagcaggctccatgaacgtttacggcaa	agaaagctgggtctatctatgcgtcgagtc
tb927.10.5830	aaaaagcaggctcgatgtatttacattccat	agaaagctgggttcaaattttgacgggctt
tb927.10.5990	aaaaagcaggctccatgtcatcttcccgtgt	agaaagctgggtctaccgatgctgggaata
tb927.10.6450	aaaaagcaggctcaatggatcaccctcccgc	agaaagctgggtctaaatgaatgtcgtagc
tb927.10.7790	aaaaagcaggctccatgtatcctggaaagga	agaaagctgggttcagtttttgctgcgtt
tb927.10.9020	aaaaagcaggctcaatgatgggtgatgctgg	agaaagctgggtctacgagtctgaaagatt
tb927.10.9150	aaaaagcaggctccatggggaaagcaccagt	agaaagctgggtctacaaatcagggttcac
tb927.10.9240	aaaaagcaggctctatggatgacaaacaggc	agaaagctgggtctagttaagcagttcatt
tb927.10.9330	aaaaagcaggctcgatggaacacgaaattca	agaaagctgggtttacattcttcccggtac
tb927.11.10020	aaaaagcaggctccatgcaactcttatcgtg	agaaagctgggttcatttctggagagagcg
tb927.11.11860	aaaaagcaggctcaatgagagaaaaggaagc	agaaagctgggttcaaagttccgggctgat
tb927.11.12070	aaaaagcaggctccatgaacgctaggaagcc	agaaagctgggtctacaactgaagactttc
tb927.11.12640	aaaaagcaggctctatgagctctgcggtttc	agaaagctgggttcattcattcaactcctt
tb927.11.12670	aaaaagcaggctccatgttcgacttgatggc	agaaagctgggtctacgcatcggggtgttt
tb927.11.12730	aaaaagcaggctcaatgagctgccttcatgg	agaaagctgggtttaacccggtttgtggca
tb927.11.1340	aaaaagcaggctcaatgagttctgtgaggct	agaaagctgggtttattcatcagtgaagta
tb927.11.14220	aaaaagcaggctcaatggacagcatcactgc	agaaagctgggtctattgttgcggctgctg
tb927.11.14520	aaaaagcaggctctatgttacggcatttgtt	agaaagctgggtttagtggctcaggcgagc
tb927.11.14750	aaaaagcaggctcaatgttccacgatctgac	agaaagctgggtttatttgtacgttggagg
tb927.11.14800	aaaaagcaggctcaatgccctcagaatcttt	agaaagctgggtctatcgctgacgcagatt
tb927.11.15000	aaaaagcaggctctatggtccggcggaataa	agaaagctgggtttactctccacgagcacg
tb927.11.16130	aaaaagcaggctccatgcccagcgagcggac	agaaagctgggtttatgccctttcataaat
tb927.11.16350	aaaaagcaggctcgatgggtatacttcagca	agaaagctgggtttacgagtacttatccac
tb927.11.16470	aaaaagcaggctcaatgaccgaaactgatgg	agaaagctgggttcaagccaattcacgttt
tb927.11.16610	aaaaagcaggctcaatgagctctcagcaccg	agaaagctgggtttacgcatccactacgag
tb927.11.2030	aaaaagcaggctcgatgtcttcatctttctt	agaaagctgggttcaacggttcaagtaata
tb927.11.220	aaaaagcaggctctatgattgaagcggacgt	agaaagctgggttcacatcaacgacctggt
tb927.11.2250	aaaaagcaggctccatgtacaataacgactc	agaaagctgggtttaggaaggcccatctgc
tb927.11.2300	aaaaagcaggctccatggccgaccacgagtt	agaaagctgggtttacataaagtcgtcgtc
tb927.11.2610	aaaaagcaggctctatggcagccgctgttac	agaaagctgggtctacgacgtgcgttcgac
tb927.11.2900	aaaaagcaggctcaatgacagggatgctccg	agaaagctgggttcaggagttaaagtttac
tb927.11.3100	aaaaagcaggctcgatgaccgcactgtcaaa	agaaagctgggtctagcgggaggaaaaccg
tb927.11.3440	aaaaagcaggctctatggtgtatgagtttgt	agaaagctgggtctaagacggatgcatgtc
tb927.11.360	aaaaagcaggctcgatgaagactattgccct	agaaagctgggttcaagtaccttcccgctc

tb927.11.3650	aaaaagcaggctccatggccgcggctccaag	agaaagctgggtttattgcttgaggtatat
tb927.11.4320	aaaaagcaggctcgatggagaaaattgttcg	agaaagctgggtctaagcatctggtggtgc
tb927.11.4400	aaaaagcaggctccatgcttggttcatggct	agaaagctgggttcaactctgggatttcgt
tb927.11.4440	aaaaagcaggctcgatgcctcaaagcctgtt	agaaagctgggtctatttcccccggcgatg
tb927.11.4550	aaaaagcaggctcaatgtcgaatgcacttga	agaaagctgggttcacgcgtcatcatctct
tb927.11.5840	aaaaagcaggctcaatgaatgcaaacgacga	agaaagctgggttcagttcagcgagtggat
tb927.11.5960	aaaaagcaggctctatgttctcgaaatccgt	agaaagctgggttcaaacggttgcctgctc
tb927.11.6010	aaaaagcaggctccatgcgcggtgccgatca	agaaagctgggtttatacgagctccctttt
tb927.11.6130	aaaaagcaggctccatggcggaagaggtgtc	agaaagctgggtttacgcctcttcgcacca
tb927.11.6240	aaaaagcaggctccatgcgtccgaggttgta	agaaagctgggttcaacagcaccttacttc
tb927.11.6430	aaaaagcaggctctatggagtccgtgccgcg	agaaagctgggttcaaaaagaccactgcgc
tb927.11.6440	aaaaagcaggctctatgcctccaaaaaacaa	agaaagctgggtctaccacctattagggag
tb927.11.6600	aaaaagcaggctctatggatcccagaaaccc	agaaagctgggttcaagcacttatgccaca
tb927.11.6870	aaaaagcaggctcaatggcgggctttcaaat	agaaagctgggttcactccaattcctccat
tb927.11.7100	aaaaagcaggctctatgcaacaggagaaccg	agaaagctgggttcaccgagcgccgatgtg
tb927.11.7590	aaaaagcaggctccatgtggcactcttttct	agaaagctgggtttacacttccacttgggc
tb927.11.8020	aaaaagcaggctcgatggcacttgaagggtc	agaaagctgggtctacacgtgcctcagttc
tb927.11.830	aaaaagcaggctctatgggccacgcttttg	agaaagctgggttcacacactcctgccttc
tb927.11.860	aaaaagcaggctcaatgcacgctggcactcc	agaaagctgggtctaggcccttgttgcacc
tb927.11.9100	aaaaagcaggctcaatggtcttttcaattcc	agaaagctgggttcacaggcgcctaaacac
tb927.11.9400	aaaaagcaggctcaatgcaacagcaaaaccg	agaaagctgggttcaatccctgcttccttc
tb927.11.9410	aaaaagcaggctctatgcatttacaaataaa	agaaagctgggttcattcttctttctgttt
tb927.11.9440	aaaaagcaggctccatgaacgaattaaagaa	agaaagctgggtttacatatttaatctaat
tb927.11.9530	aaaaagcaggctcaatgacggactgcatcaa	agaaagctgggttcagttcacgggttgctc
tb927.2.3160	aaaaagcaggctcaatgtccagtcggtgctc	agaaagctgggttcacctcgcaatccaaga
tb927.2.3920	aaaaagcaggctctatgttttcttgttgcct	agaaagctgggtctacttgcgcctagaata
tb927.2.4440	aaaaagcaggctccatgtacgatgagaaagt	agaaagctgggttcagacttttttttttct
tb927.2.4710	aaaaagcaggctccatgcaacaatataccct	agaaagctgggttcagtcccttacgcggtc
tb927.2.4930	aaaaagcaggctctatgtctcacattctttt	agaaagctgggttcatttcatcaacgtcat
tb927.2.5000	aaaaagcaggctccatggcaacaagaggtgt	agaaagctgggttcacttgtgttgaagcct
tb927.2.5150	aaaaagcaggctcaatggattacgccgccgc	agaaagctgggtctatataaccccaagttg
tb927.2.5200	aaaaagcaggctcaatggatgagcaaaacag	agaaagctgggtctaattgtcggaatcgct
tb927.3.1300	aaaaagcaggctccatgtggttcacaaacat	agaaagctgggtctactttacatttgtcca
tb927.3.1810	aaaaagcaggctctatgcgaacaaccacgca	agaaagctgggttcactgggtgtgctgtgg
tb927.3.2900	aaaaagcaggctcgatggcagcttacggtat	agaaagctgggttcagtccgtgtcttcgcc

tb927.3.3060	aaaaagcaggctcgatggattttgattttgt	agaaagctgggtctaaaacggatccgctga
tb927.3.3440	aaaaagcaggctccatgaacaattcacttaa	agaaagctgggtctattccactttgcatct
tb927.3.3490	aaaaagcaggctcgatggcaacagaattgaa	agaaagctgggtctacttgggtagtgctgc
tb927.3.5010	aaaaagcaggctcaatggcggataatgaaga	agaaagctgggtttatgacccctcctcga
tb927.3.5090	aaaaagcaggctctatgccctccgctgaaac	agaaagctgggttcaactccacgactgcca
tb927.3.5320	aaaaagcaggctctatggcaatgtgtgattg	agaaagctgggttcaatggatgtgaagaac
tb927.3.580	aaaaagcaggctcgatggatttggaaaattg	agaaagctgggtctacatatattcatcatc
tb927.4.1540	aaaaagcaggctccatgcttcgtgtattctc	agaaagctgggttcaattctttcccctcct
tb927.4.1910	aaaaagcaggctccatgtcttcctttaacat	agaaagctgggtttacgaagtctgtccctc
tb927.4.2520	aaaaagcaggctcaatgaggcggcccaatgg	agaaagctgggttcatgccgcgggaccctt
tb927.4.3330	aaaaagcaggctcgatgcctgcttcagcctc	agaaagctgggtctaacgatgatttgagag
tb927.4.3470	aaaaagcaggctcgatgtctcttcgtgacga	agaaagctgggtctatgcactcccatgcat
tb927.4.3970	aaaaagcaggctcaatgtcggcaggttcgga	agaaagctgggtttaattgctaaagaactc
tb927.4.4570	aaaaagcaggctcaatgggtgttgcacccag	agaaagctgggtttaggacttcctccagtc
tb927.4.550	aaaaagcaggctcaatgtacggcaacaaagg	agaaagctgggttcactcctcagttttaaa
tb927.4.640	aaaaagcaggctcgatggacgctgttgtaaa	agaaagctgggttcaaaaccttctcttgac
tb927.4.990	aaaaagcaggctcgatggacaccgctgctga	agaaagctgggttcaaacaaacgattttcg
tb927.5.1600	aaaaagcaggctcaatgagtgaaccattgat	agaaagctgggtctattcttcattttcctc
tb927.5.1990	aaaaagcaggctctatgacgcaatcgcaatc	agaaagctgggtttagcgactgaagcgggc
tb927.5.2120	aaaaagcaggctcgatggagaatcaggcacc	agaaagctgggtctacttcatcatggtatt
tb927.5.2620	aaaaagcaggctcaatgcacgtacttggcag	agaaagctgggtttacagctgttcaagaga
tb927.5.3120	aaaaagcaggctcgatggctgaacaaccgga	agaaagctgggtctaaagtgtcatggcttg
tb927.5.3450	aaaaagcaggctcgatggagtcgaatgtgtc	agaaagctgggtctatctccgtttcttctt
tb927.5.4020	aaaaagcaggctcgatgatgatttcccgcgc	agaaagctgggtttagaaagagagggccgt
tb927.5.4100	aaaaagcaggctctatgacttttctacccgt	agaaagctgggtttaaaggctcagtttgaa
tb927.5.4320	aaaaagcaggctcgatggaagagggcggtga	agaaagctgggtttacgtgtcctcagagtg
tb927.5.4330	aaaaagcaggctcgatgagatcgtttatgta	agaaagctgggttcacgccacctcgcccgc
tb927.5.4450	aaaaagcaggctccatgtcaatgagcaacgg	agaaagctgggttcactttacagcagtttt
tb927.5.4620	aaaaagcaggctcaatgcaccggcttgcaac	agaaagctgggtttaagagtacatgaggat
tb927.5.750	aaaaagcaggctcgatgtctgtagttgttag	agaaagctgggtctactctaacatgtactg
tb927.6.2720	aaaaagcaggctctatgccgcactccatgga	agaaagctgggtttacaaaagtttccattg
tb927.6.3220	aaaaagcaggctcaatgccagttcgacagaa	agaaagctgggttcaatcactaagtaactt
tb927.6.3420	aaaaagcaggctcgatgcggcgaatttttag	agaaagctgggttcaaacttgcgtcagcag
tb927.6.3950	aaaaagcaggctctatggaggatccacggga	agaaagctgggtctaacggcgatattcctc
tb927.6.4270	aaaaagcaggctcaatgaagcggtcttttt	agaaagctgggttcaatgaactgcagcact

tb927.6.5010	aaaaagcaggctcgatggaacattcgcgtcg	agaaagctgggtttaaacactcttccctcc
tb927.7.1710	aaaaagcaggctcgatgcgcttgttgaaaaa	agaaagctgggtttacacaactacacttgc
tb927.7.1890	aaaaagcaggctcaatgaatgtttctgagtt	agaaagctgggtctatgacgtaaacctctt
tb927.7.2160	aaaaagcaggctcaatgcaagcatctgaatc	agaaagctgggttcacgctgcaaccttctg
tb927.7.2440	aaaaagcaggctcaatgaagatatcatttat	agaaagctgggtttaattcaatcccttctc
tb927.7.2610	aaaaagcaggctctatggatggcgagggatg	agaaagctgggtttatctcaaatccatctc
tb927.7.2780	aaaaagcaggctcaatgtctaaagctccttc	agaaagctgggtctatgccgccgctgccgg
tb927.7.2980	aaaaagcaggctccatgcgccggtctgtata	agaaagctgggtttatttaaacaccttaaa
tb927.7.3220	aaaaagcaggctcaatgcgagctgaaaaggt	agaaagctgggttcactcgggaacattttg
tb927.7.3350	aaaaagcaggctcaatgtgcagctgttgctg	agaaagctgggttcaagcaataggttcgtt
tb927.7.3550	aaaaagcaggctcaatgtctatattcggtga	agaaagctgggtttagaaacctcgcccaac
tb927.7.4120	aaaaagcaggctcgatggatctggattcttt	agaaagctgggtttagcttggtggtctaaa
tb927.7.4430	aaaaagcaggctcaatgcgttccgttattcg	agaaagctgggttcaatcattccatggact
tb927.7.4520	aaaaagcaggctccatgaccgatacaattga	agaaagctgggtttaaaatagcgacatttc
tb927.7.4660	aaaaagcaggctcaatgcgaacacgatttag	agaaagctgggtttacttccatggaacacg
tb927.7.5480	aaaaagcaggctccatgctcagtcttacgcg	agaaagctgggtctacaccgccatctccat
tb927.7.5680	aaaaagcaggctctatgaccgaccttcacat	agaaagctgggtttagtatttactgcggga
tb927.7.7170	aaaaagcaggctctatgaatccacacgcctg	agaaagctgggtttatcggtagtccctgta
tb927.7.7190	aaaaagcaggctcaatgggaaaaaagaaggc	agaaagctgggtttacactccgacactgcc
tb927.7.7220	aaaaagcaggctcaatgagcaccagttcgtc	agaaagctgggtctattgttcgtgagaggc
tb927.8.1460	aaaaagcaggctcgatggcatcccaaaatgc	agaaagctgggtctaactggagcgcgaacc
tb927.8.1680	aaaaagcaggctccatgatggcgagtgcccc	agaaagctgggttcaactggggtcaccaat
tb927.8.1770	aaaaagcaggctcgatggggttgttcatcca	agaaagctgggttcaggactgctgtgttgc
tb927.8.3090	aaaaagcaggctccatgcaagtctctccacg	agaaagctgggtctaacttgaaaacggctt
tb927.8.3780	aaaaagcaggctcaatgccacccaaaaagaa	agaaagctgggtctaaccctcttttcatc
tb927.8.3850	aaaaagcaggctcaatgatagtcagaacgcc	agaaagctgggttcaagcggctaccgcgcc
tb927.8.4190	aaaaagcaggctcaatgagcggccataggca	agaaagctgggtttactccccagtaaaaat
tb927.8.4200	aaaaagcaggctccatgctagatctatttgc	agaaagctgggttcacccccttatgttgcgt
tb927.8.4220	aaaaagcaggctccatgggcaagaaggctga	agaaagctgggttcaagaaagatcatacgc
tb927.8.4370	aaaaagcaggctcaatggctgacagtaacaa	agaaagctgggttcactttggtgcactctg
tb927.8.4390	aaaaagcaggctcgatggggaaacgatcgat	agaaagctgggttcagtcctcaggactgta
tb927.8.4650	aaaaagcaggctcgatgcaaaggtgtgtcca	agaaagctgggtttacacagcgggcttgaa
tb927.8.5370	aaaaagcaggctccatgtccctctcaccaca	agaaagctgggttcaggcactcggcgccaa
tb927.8.6750	aaaaagcaggctcaatgaagatcttcaggga	agaaagctgggttcacacgcgttcgcccct
tb927.8.7760	aaaaagcaggctcaatgggtgctgctccaag	agaaagctgggtttatctatctgcgcaata

tb927.8.7820	aaaaagcaggctctatgctttttcacacgac	agaaagctgggttcatttcccatcaaatcc
tb927.8.910	aaaaagcaggctcaatgactgaattcaatta	agaaagctgggtttaattgaagcgacatcc
tb927.9.10770	aaaaagcaggctccatggctgcatttgctgc	agaaagctgggtctacatgccaatgtgacg
tb927.9.1120	aaaaagcaggctctatggggattagtttttg	agaaagctgggtttatatatcctcttctcc
tb927.9.12320	aaaaagcaggctcaatgtctaagcagcagca	agaaagctgggtctaatggcacatggcatc
tb927.9.13150	aaaaagcaggctcgatgggatacaggttgtt	agaaagctgggtttatgacgtgcaggtggc
tb927.9.13970	aaaaagcaggctctatgcactgccgcctgct	agaaagctgggttcattgtctaggaataag
tb927.9.14120	aaaaagcaggctcaatggagcattttgcccc	agaaagctgggtctaccggttcctatgctt
tb927.9.15040	aaaaagcaggctcgatgaagcgcgcgcgcaa	agaaagctgggttcacgaagttctgtggtt
tb927.9.17940	aaaaagcaggctccatgacatgtaggagttc	agaaagctgggttcattcagatccaaagcc
tb927.9.3000	aaaaagcaggctcaatgcaatgctcatcatt	agaaagctgggttcagaaagttcctggaca
tb927.9.3580	aaaaagcaggctcgatgcatgtggcctctcc	agaaagctgggtctagtcagaatcactgtc
tb927.9.4080	aaaaagcaggctctatggcatcggaggcagc	agaaagctgggtttattcgcccctaactgc
tb927.9.5260	aaaaagcaggctcgatgctttccctatatgc	agaaagctgggtctaggaggcatcaatgga
tb927.9.5390	aaaaagcaggctctatgtctattgagaggct	agaaagctgggtttaatcatcgctaagttc
tb927.9.6470	aaaaagcaggctccatgacactggagggcgg	agaaagctgggtttaggccccttcgccaat
tb927.9.7110	aaaaagcaggctcgatggagggaagcactgc	agaaagctgggtttactgagtgtactgaac
tb927.9.7480	aaaaagcaggctcaatgaaccaaccgacaac	agaaagctgggtctacttggcacgtaggtg
tb927.9.8060	aaaaagcaggctccatgtacaaggagaccac	agaaagctgggtttaccaatcgtcatcatc
tb927.9.9290	aaaaagcaggctctatgacaatcgctgcaca	agaaagctgggtttaagcgcttgaggcgtg
tb927.9.9550	aaaaagcaggctccatgttggatcacacaat	agaaagctgggttcactgttgttgctgttg

Other Oligomers used

THT_fow- gtggcggccgctctagctagcactgaagcttaattc

THT_rev6- atagggcgaattgggtacacttacaagagatcattatagaagaacat

THT_rev5- atagggcgaattgggtacacttacgaaacgtcttcaattgtataa

THT_rev4- atagggcgaattgggtacacttacatacacactaacagcatactg

THT_rev-atagggcgaattgggtacacttacgatggaaaaacaaaagtatatat

THT_rev2-atagggcgaattgggtacacttaccaattacctccttcaaac

THT_rev1-atagggcgaattgggtacacttacaaacaacttcgtcgc

Aurora

Fow_caccatgcatcaccatcatcaccatcatcaccatcacatgaggtcaa ctgaggtc

Rev_attctctttccctgcagt

Caf40

Fow_caccatgcatcaccatcatcaccatcatcaccatcacatgcaccaaa cccaagcg

Rev_cgcactggcatcaccg

Fow_SNAPC-atgatgcctcctaaacaattggtgg

Rev_SNAPC-atagggcgaattgggtacacttacagcagtgtagtactttccttgagg

Fow_TRF4-atgatggacaatgacttcactggaaac

Rev_TRF4-atagggcgaattgggtacacttaccctctttgcgtactgtgcaat

Fow_TFIIA-atgatgtaccgtgagagcctgctgggt

Rev_TFIIA-atagggcgaattgggtacacttacccgcgagcgcttccgttt

Fow_U70k-gtggcggccgctctagctatgatgcagacccctggtga

Fwd_U24k-gtggcggccgctctagctatgatgacgtactgccccattacgc

Fow_U1c-gtggcggccgctctagctatgatgagtgatagttttgctttgg

Fow_u70k Topo-caccatgatgcagacccctggtga

Fow_U1a-gtggcggccgctctagctatgatggaggcgcttcgcgt

Polo-like kinase

Fow_caccatgatgcacgcaaccgctgagac

Rev_aatatcacggttttgtatgagctcgctg

Tb927.7.3730

Fow _gtggcggccgctctagctatgatggaaacgcaccaggat

Rev_ atagggcgaattgggtacacttacgtttgggctcatgaagttgg

Tb927.7.2680

Fow_gtggcggccgctctagctatgatgttccttcccgaacaatc

Rev_atagggcgaattgggtacacttacaaagggtacgtgtagggga

Tb927.8.6650

Fow_gtggcggccgctctagctatg atgacagaccaccaagtcat

Rev_ atagggcgaattgggtacacttacacctacacgcatggtgg

Tb927.5.1580

Fow_gtggcggccgctctagctatgatggcgagaaaccggaa

Rev_ atagggcgaattgggtacacttaccagagggtacacgggca

Tb927.7.5380

Fow_gtggcggccgctctagctatgatgagcttattttcttttac

Rev_atagggcgaattgggtacacttaccgtaagtgcagcaaatg

Tb927.5.285b

Fow_gtggcggccgctctagctatgatgacctcactgctgtcg

Rev_atagggcgaattgggtacacttacattagcatcattttccgtca

Tb927.7.3990

Fow_gtggcggccgctctagctatgatgcgacgtactttcagtcg

Rev_ atagggcgaattgggtacacttaccctggacaactcccctagtg

Tb927.11.16520

Fow_gtggcggccgctctagctatgatgaattattcaaccgggta

Rev_ atagggcgaattgggtacacttacagtgctactttcccctgtc

Tb927.11.10500

Fow_gtggcggccgctctagctatgatgaacgtcgatgatttcc

Rev_ atagggcgaattgggtacacttaccaggagctgcactcgg

Tb927.8.2780 rbp10

Fow_caccatgatgggagactcgatatcacc

Rev_ cattcgaaccggaggat

Trypanothion reductase

Fow_gtggcggccgctctagcatgtccaaggccttcgatttgg

Rev_atagggcgaattgggtacacttaccaggttagagtccggaagc

Z3CH11

Fow_caccatgatgcaccaccatcaccat

Rev_acaggagtccaagctcagc

Tgsgp fow_gctgctgtgttcggagagc

Tgsgp rev_ gccatcgtgcttgccgctc

Fow_atg_ggatcctaatacgactcactatagggagaccacaacggtttccctctagaaataatttgttaactttaaggaggggatccgccaccatgcgtggcggccgctctagc
Rev_myc_ctggaattcgccctttttattacagatcctcttcagagatgagtttctgctcaactttgtacaagaaagctgggtc
Lib_fow_gaaattaatacgactcactatagggagaccacaacggtttccctctagaaataatttttgtttaagaaggagatatacatatgcatcatcaccaccatcacgaaaaagcaggctc-3’
Lib_rev_ 5’-caccgctgagcaataactagctggaattcgccctttttatta gagaaagctgggt-3’

RNA probes

U1snRNA
aacucaccugcagugcgucaucacgaaagcgccgugcuugcaguaaacuuugagugggaacgaaagcgucccucc
PUF element
uaacauuguaccaugcauaacauuguaccaucag
VSG UTR repeat
cuugcuacuugacugauauauuuuaac
ARE Classical
uauuuauuuauuuauuuauuuau
ARE Non-classical
uuauuauuauuauuauuauuauu
Poly U
uuuuuuuuuuuuuuuuuuuuuuu

Poly C
CCCCCCCCCCCCCCCCCCCCCC
Mutant ARE
uucuucuucuucuucuucuucuu
THT1 Oligomer
cuggugagguauuguuugucucgcucg

Index II: Targets involved in immunity and infectious diseases

miRNA	Target	Pathways involving Target
hsa-miR-98	ACSL6	Fatty Acid Activation Fatty Acid Metabolism Fatty Acid β-oxidation I LPS/IL-1 Mediated Inhibition of RXR Function Mitochondrial L-carnitine Shuttle Pathway Stearate Biosynthesis I (Animals) Type II Diabetes Mellitus Signaling γ-linolenate Biosynthesis II (Animals)
hsa-miR-98	BCL2L1	Apoptosis Signaling B Cell Receptor Signaling CD27 Signaling in Lymphocytes Chronic Myeloid Leukemia IL-15 Signaling IL-8 Signaling Induction of Apoptosis by HIV1 JAK/Stat Signaling Lymphotoxin β Receptor Signaling
hsa-miR-195	PIK3R1	B Cell Receptor Signaling CCR3 Signaling in Eosinophils CD28 Signaling in T Helper Cells CTLA4 Signaling in Cytotoxic T Lymphocytes Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes FcγRIIB Signaling in B Lymphocytes G-Protein Coupled Receptor Signaling LPS-stimulated MAPK Signaling Leptin Signaling in Obesity Leukocyte Extravasation Signaling Myc Mediated Apoptosis Signaling NF-κB Activation by Viruses NF-κB Signaling NGF Signaling NRF2-mediated Oxidative Stress Response Natural Killer Cell Signaling Neuregulin Signaling Neuropathic PI3K Signaling in B Lymphocytes PKCθ Signaling in T Lymphocytes Role of NFAT in Regulation of the Immune Response Role of PI3K/AKT Signaling in the Pathogenesis of Influenza Role of Pattern Recognition Receptors in Recognition of Bacteria and VirusesT Cell Receptor Signaling
hsa-miR-193b	EIF2S1	Signaling Role of PKR in Interferon Induction and Antiviral Response Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses VEGF Signaling
hsa-miR-193b	PLA2G4D	Role of MAPK Signaling in the Pathogenesis of Influenza
hsa-miR-199a-3p	CXCL11	IL-17 Signaling IL-17A Signaling in Gastric Cells Pathogenesis of Multiple Sclerosis
hsa-miR-199a-3p	MET	FGF Signaling HGF Signaling
hsa-miR-27b	CXCL11	IL-17 Signaling IL-17A Signaling in Gastric Cells Pathogenesis of Multiple Sclerosis
hsa-miR-27b	FZD7	G-Protein Coupled Receptor Signaling Role of Wnt/GSK-3β Signaling in the Pathogenesis of Influenza
hsa-miR-27b	MAPK10	4-1BB Signaling in T Lymphocytes, B Cell Activating Factor Signaling, CCR5 Signaling in Macrophages CD27 Signaling in Lymphocytes CD28 Signaling in T Helper Cells Leukocyte Extravasation Signaling Role of MAPK Signaling in the Pathogenesis of Influenza Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
hsa-miR-27b	RFXAP	Primary Immunodeficiency Signaling
hsa-miR-27b	UBE2V1	NF-κB Signaling

hsa-miR-454	MET	FGF Signaling HGF Signaling
hsa-miR-454	RFXAP	Primary Immunodeficiency Signaling
hsa-miR-454	TIMP2	Leukocyte Extravasation Signaling
hsa-miR-4291	UBE2V1	NF-κB Signaling