UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Role of long non-coding RNAs in parasite differentiation

Mestrado em Biologia Molecular e Genética

Fábio Pereira Bento

Dissertação orientada por:
Doutora Luisa Miranda Figueiredo (Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa) e Prof. Doutor Jorge Marques da Silva (Faculdade de Ciências, Universidade de Lisboa)

2015

Acknowledgments

Em primeiro lugar, obrigado Luísa. Muito obrigado por toda a dedicação e paciência em todos os momentos desde que cheguei ao teu grupo. Agradeço-te pela chefe que és, por tudo o que me ensinaste, por me motivares e estimulares a ter opinião e a ser critico e me fazeres crescer como cientista. Obrigado por fazeres deste grupo uma família única, que me fez crescer não só como cientista mas também como pessoa. Foi um privilégio aprender contigo.

Agradeço ao Professor Jorge Marques da Silva pela disponibilidade ao longo do ano.

Muito obrigado Fabien, foste mais que um coorientador ao longo deste ano e desta tese. Foste incansável e o teu pupilo (ou "pupilho", seja lá a forma como o dizes) está eternamente grato por todo o teu apoio, dedicação e por tudo o que me ensinaste. Tornaste este projecto possível.

Obrigado a todos os elementos do laboratório que me acolheram e me fizeram sentir em casa, sempre disponíveis em todos os momentos. Obrigado Pena pela disponibilidade e sugestões, sem ti o laboratório é demasiado calmo. Leonor, agradeço-te por todo o apoio, mesmo depois de partir tanta coisa. Obrigado Helena e Daniel pela disponibilidade e ajuda naquilo que só vocês sabem fazer. Sandra agradeço-te pelas amostras de sangue e por todo o apoio. Obrigado Xico por toda a ajuda e disponibilidade sempre que precisei. Obrigado Idálio por todo o apoio e paciência com todas as perguntas que te fazia. Obrigado Margarida por todas as opiniões e sugestões. Filipa, a ti tenho que agradecer não só pelo apoio, mas também pela secretária e pelo ecrã extra (devo dizer que agora estou mal habituado). Agradeço ao mais recente elemento do grupo João (Johnny), também meu colega de curso. Obrigado ao João, Miguel, Eleonora e Jessica por toda a disponibilidade e ajuda.

Obrigado aos nossos vizinhos do MPrudêncio Lab que tornam ambiente no laboratório ainda mais único, especialmente à Carolina e à Marta por todo o apoio.

Agradeço a todos os meus colegas de curso com que me cruzei, especialmente ao Bruno e ao Jota por todas as conversas e todos os momentos únicos. Sem vocês o curso não era a mesma coisa.

Obrigado Vitor, Silvana e Raquel, com quem partilho a casa há 5 anos e me acompanharam ao longo de todo o meu percurso académico.

Agradeço a todos os meu amigos que estiveram lá sempre dispostos a apoiar-me (mesmo quando estavam fartos de me ouvir a falar de trabalho). Não vou listar nomes que vocês sabem quem são.

Quero dedicar todo este trabalho à minha família, a quem eu devo tudo na vida. Um obrigado muito especial aos meus pais que estiveram sempre presentes e com grande esforço e sacrifício tornaram tudo possível e fizeram de mim quem sou hoje. Estou eternamente grato por tudo. Obrigado.

Abstract

Trypanosoma brucei (T. brucei) is the parasite that causes African Human Trypanosomiasis (HAT), also known as sleeping sickness. During its complex life cycle, the parasite alternates between two different hosts, the insect vector and the mammalian host. To adapt and survive in these different environments, the parasite undergoes multiple developmental changes that require coordinated alterations in gene expression. In *T. brucei*, transcription control of protein-coding genes is scarce and most gene regulation depends on a post-transcriptional regulation. Slender-to-stumpy transition is triggered by an unknown stumpy induction factor (SIF) molecule, which parasites released via quorum-sensing mechanism. Recently, positive regulators of the quorum-sensing pathway have been identified in *T. brucei*, including an essential protein phosphate 1 (PP1) and stumpy-promoting RNA binding protein (RBP7).

Long non-coding RNAs (lncRNAs) are RNAs with more than 200bp that do not encode for protein. In many organisms, lncRNA play important roles in gene regulation, including at post-transcriptional level. Although, our laboratory has recently revealed existence of 946 lncRNA genes in the *T. brucei* genome, yet any function has been found. We hypothesized lncRNA to control and coordinate gene expression during stumpy formation in *T. brucei*.

The aim of my thesis was to identify lncRNAs that regulate the slender-to-stumpy transition. To that end, we established a reporter cell line, an *in vitro* culture condition and a gain- and loss-of-function approach to study the function of *T. brucei* lncRNAs in stumpy differentation.

We performed a bioinformatic analysis and selected four lncRNAs based on their genomic location, postulating these lncRNAs to regulate stumpy formation by modulating the expression of their nearby stumpy essential genes. After performing a phenotypic screen for stumpy formation and more detailed experiments, we identified one lncRNA that regulate the slender to stumpy transition of *T. brucei*. Indeed, while the overexpression of lncRNA5837 promotes stumpy formation, its depletion conversely prevents slender to stumpy transition. Interestingly, the overexpression of lncRNA5837 induces stumpy formation even in absence of external stimulus that promotes parasite differentiation. As a consequence, we identified a new positive regulator of stumpy differentiation *in T. brucei*.

Finally, using RNA FISH experiment, we observed that lncRNA5837 localized mainly in the nucleus of *T. brucei.* Strikingly, the slender to stumpy transition is associated with a relocalization of lncRNA5837 from the nucleoplasm into the nucleolus of stumpy forms.

Altogether, these results suggest that lncRNA5837 regulate the slender-to-stumpy transition through its function inside the nucleolus.

In this project, we identify and characterize the first function of a lncRNA in human deadly parasite, *T. brucei*. Indeed, the lncRNA5837 regulates the parasite differentiation from slender to stumpy forms, which is key process for transmission and infectivity of *T. brucei parasite*.

Keywords: Trypanosoma brucei; long non-coding RNA; stumpy form; differentiation; stumpy-reporter cell line

Resumo

O *Trypanosoma brucei* é um organismo eucariota unicelular responsável pela Doença do Sono. O parasita pertence à ordem Kinetoplastida, caracterizada por uma estrutura que contém DNA mitocondrial designada cinetoplasto. A distribuição da doença é condicionada pela localização do vector que transmite o parasita.

A doença é caracterizada por dois estadios: no primeiro, o parasita está restrito ao sangue e ao linfa e, no segundo, o parasita atravessa a barreira hematoencefálica e invade o sistema nervoso central, causando distúrbios no ciclo do sono. Esta doença endémica da África subsariana é letal se não for tratada. Atualmente, não existe ainda uma vacina e o seu tratamento, além de dispendioso, gera efeitos secundários indesejados.

Ao longo do seu complexo ciclo de vida, o parasita alterna entre o seu hospedeiro, um mamífero, e o seu vector, a mosca tsé-tsé. Aquando da picada da mosca tsé-tsé, o parasita entra na corrente sanguínea do hospedeiro e diferencia-se em *slender forms*. Os parasitas s*lender forms* reproduzem-se assexuadamente por fissão binária, estabelecendo e mantendo a infecção no hospedeiro. O parasita apresenta na sua superfície uma proteína designada glicoproteína variante de superfície (VSG). A rápida e periódica mudança da VSG, designada variação antigénica, permite ao parasita escapar do sistema imunitário do hospedeiro, o que leva a um combate pouco eficiente contra a infecção, já que o parasita consegue persistir durante anos no mesmo hospedeiro. Quando a população do parasita atinge grandes concentrações no sangue, um mecanismo de *quorum-sensing* permite a diferenciação de *slender forms* para *stumpy forms*. Os parasitas *stumpy forms* são a sua forma transmissível e pré-adaptada para sobrevivência no intestino da mosca tsé-tsé. Ao contrário dos *slender forms*, os *stumpy forms* diferenciam-se em *procyclic forms*, aquando da picada da mosca tsé-tsé, o que leva à transmissão do parasita para o seu vector.

T. brucei apresenta uma expressão génica pouco comum comparando com outros organismos eucariotas, uma vez que os genes estão organizados em unidades transcricionais policistrónicas (PTU) constituídas por vários genes não-relacionados funcionalmente e transcritos a partir de um promotor comum. Uma vez transcrito um RNA percussor policistrónico, o transcrito é processado em mRNA maduros monocistrónicos por reações de *trans splicing* e poliadenilação. À exceção dos genes de VSG, não existem evidências de regulação transcricional em *T. brucei*, motivo pelo qual a expressão génica é regulada ao nível pós-transcricional, particularmente por regulação da estabilidade e da tradução do mRNA. Em *T. brucei*, as estruturas secundárias contribuem para a estabilidade do mRNA e para o recrutamento de proteínas de ligação ao RNA (RBP). Além disso, as RBP podem ligar-se à região 3'UTR dos mRNA para regular a associação do mRNA com a maquinaria de tradução.

Os parasitas *stumpy forms* resultam da diferenciação irreversível dos *slender forms*. A transcrição e a síntese proteíca estão geralmente reprimidos em *stumpy forms*, reforçando o seu estado de células quiescentes. Os *stumpy forms* são importantes para a transmissão para o vector, bem como para a sua sobrevivência neste último. Além disso, os *stumpy forms* são igualmente importantes para o prolongamento da infecção no hospedeiro.

A diferenciação dos *slender forms* para os *stumpy forms* está relacionada com o aumento da densidade da população no hospedeiro. Esta diferenciação ocorre por comunicação parasita-parasita, através de um mecanismo de *quorum-sensing*, em que os *slender forms* secretam uma molécula, designada Factor de Indução de *Stumpy* (SIF). A molécula em causa acumula-se e estimula a diferenciação para os *stumpy forms* com o aumento do número de parasitas. Recentemente, foram identificados vários genes essenciais para a via de *quorum-sensing*, embora apenas tenham sido identificados reguladores positivos que atuam em diferentes etapas da via de sinalização de SIF, incluindo as fosfatases, as cinases e as proteínas envolvidas na regulação génica, como as RBP.

Os RNA longos não codificantes (lncRNA) são RNA com mais de 200 pares de base que não codificam proteínas. Os lncRNA atuam, em diferentes organismos de todos os reinos da vida, como reguladores de expressão génica, através da regulação da transcrição, da estabilidade de mRNA ou da eficiência da tradução, entre outros. Os lncRNA podem atuar *in cis* ou *in trans*, assim como em vários compartimentos celulares, nomeadamente no núcleo ou no citoplasma, para regular a expressão génica e a função. Recentemente, foram identificados, em *T. brucei*, 946 lncRNA, dos quais 567 foram identificados em apenas um estadio do ciclo de vida do parasita, sugerindo, assim, a existência de lncRNA específicos de um estadio. O objetivo desta tese é identificar e caracterizar o papel dos lncRNA na diferenciação dos *slender forms* para *stumpy forms*.

Para estudar a diferenciação para os *stumpy forms*, é necessário estabelecer uma linha celular que os assinale. Para tal, foram criados parasitas que expressam GFP especificamente em *stumpy forms*, devido ao 3'UTR de PAD1, uma proteína marcadora de stumpy forms, cujo 3'UTR permite a expressão génica específica em *stumpy forms*. Esta linha celular, designada GP1, foi caracterizada usando dois métodos para a formação de *stumpy forms in vitro*: meio complementado com metilcelulose e meio complementado com pCPT-cAMP, sendo o número de células na fase G1 foi relacionado com o número de células que expressam GFP. Os parasitas cultivados em meio com metilcelulose e/ou com pCPT-cAMP no final de dois ou três dias apresentam uma acumulação de células na fase G1, característica dos *stumpy forms*. A linha celular GP1 apenas expressa GFP quando a diferenciação dos *stumpy forms* é induzida, quer *in vitro* (meio complementado com metilcelulose e/ou pCPT-cAMP) quer *in vivo* (sangue recolhido seis dias após a infeção, quando a maioria dos parasitas são *stumpy forms*). Tal facto confirma que a GP1 é uma

linha celular marcadora de *stumpy forms* que pode ser usada para estudar a diferenciação destes.

Para selecionar lncRNA potencialmente envolvidos na formação dos *stumpy forms*, selecionámos lncRNA a jusante e a montante de nove genes essenciais para a formação de *stumpy forms*, que apresentaram resistência ao SIF *in vivo*. Assim, propomos que estes lncRNA regulam a expressão e/ou função dos genes essenciais para a formação de *stumpy forms*, identificando quatro lncRNA localizados imediatamente a jusante ou a montante de três dos nove genes essenciais para formação de *stumpy forms*. O lncRNA5388 e o lncRNA5389 estão localizados a montante da família de genes NEK, enquanto o lncRNA5837 está localizado a jusante da família de genes RBP7. Por último, o lncRNA6099 está localizado a jusante do gene YAK.

Para caracterizar os lncRNA na diferenciação dos *stumpy forms*, os lncRNA foram quantificados durante a formação de *stumpy forms in vitro*. Os níveis de transcritos de lncRNA não apresentaram alterações significativas durante a diferenciação para *stumpy forms*. Para testar a possibilidade de uma função dos lncRNA selecionados ao nível da formação de *stumpy forms*, determinámos o fenótipo na diferenciação para *stumpy forms* causado pela depleção ou sobre-expressão dos lncRNA. Contudo, as experiências apresentaram grande variabilidade e inconsistência, tornando-se difícil interpretar e concluir a função dos lncRNA na diferenciação dos *stumpy forms*.

Assim, a continuação do estudo incidiu apenas no lncRNA5837, visto que é o lncRNA localizado a jusante de genes essenciais para a diferenciação dos *stumpy forms*, assim como para a regulação génica, a família génica RBP7. Além disso, este lncRNA apresenta um perfil de expressão semelhante ao RBP7 durante a diferenciação para *stumpy forms*.

Durante a diferenciação *in vitro* para *stumpy forms*, a sobre-expressão de lncRNA5837, tal como a sobre-expressão de RBP7, resultou no aumento de células que expressam GFP, sugerindo isso que há promoção da diferenciação de *stumpy forms*. Por outo lado, a depleção de lncRNA5837, tal como a depleção de PP1, levou a que menos parasitas se diferenciarem para *stumpy forms*. Em condições que impedem a formação de *stumpy forms*, a sobre-expressão de lncRNA5837 é suficiente para induzir os *slender forms* a se diferenciarem em *stumpy forms*. Estes resultados são consistentes com o facto de a sobre-expressão de lncRNA5837 levar a uma diminuição da taxa de crescimento dos parasitas, característico da formação de *stumpy forms*, parasitas em fase G1.

A localização celular do lncRNA5837 foi determinada usando RNA FISH. O lncRNA5837 está localizado no núcleo de *T. brucei*. Em *slender forms*, é possível observar diversos sinais de lncRNA5837 no núcleo, um sinal mais intenso no nucléolo e um sinal (por vezes, dois) no nucleoplasma. No entanto, os *stumpy forms* apresentam apenas lncRNA5837 no nucléolo, ainda que aparentemente mais intenso.

Em suma, estes resultados sugerem que o lncRNA5837 atua na diferenciação para *stumpy forms*, através da sua função expressa no nucléolo de *T. brucei*.

Abbreviations

3'UTR – 3' untranslated region
5'UTR – 5' untranslated region
BSF – bloodstream form
bp – base pair
CCA – cis-aconitate
cDNA – complementary DNA
CDS – coding sequence
DAPI – 4',6-diamidino-2-phenylindole
DCL – Dicer-like enzyme
DNase I – deoxyribonuclease I
ESB – expression site body
FISH – fluorescence *in situ* hybridization
G1 – gap1 phase of cell cycle
G2 – gap2 phase of cell cycle
GFP – green fluorescence protein
HAT – Human African Trypanosomiasis
HMI-11 – Hirumi's modified Iscove's medium 11
Kb – kilobase
lncRNA – long non-coding RNA
M – mitosis phase of cell cycle
MAPK5 – mitogen-activated protein kinase 5
mRNA – menssenger RNA
ncRNA – non-coding RNA
NLS – nuclear localization signal
nt – nucleotide
OE-5388 – mutant that overexpress lncRNA5388
OE-5389 – mutant that overexpress lncRNA5389
OE-5837 – mutant that overexpress lncRNA5837
OE-6099 – mutant that overexpress lncRNA6099
PAD – Protein Associated with Differentiation
pCPT-cAMP – 8-(4-chlorophenylthio)-cAMP
PCR – polymerase chain reaction
PCF – procyclic form
PI – propidium iodide
Pol I – RNA polymerase I
Pol II – RNA polymerase II
Poly A - polyadenylated
PP1 – phosphatase protein 1
PTU – polycistronic unit
qPCR – quantitative real-time PCR
RBP – RNA binding protein
rDNA – ribosomal DNA
RNA-seq – RNA sequencing
RNAi – RNA interference
RNAi-5388 – mutant that knock-down lncRNA5388
RNAi-5389 – mutant that knock-down lncRNA5389
RNAi-5837 – mutant that knock-down lncRNA5837
RNAi-6099 – mutant that knock-down lncRNA6099
RNase A – ribonucelase A
RNase H – ribonuclease H
RPKM – reads per kilobase per million mapped reads
rRNA – ribosomal RNA
S – synthesis phase of cell cycle
s.e.m – standard error of the mean
seq – sequencing
SIF – Stumpy Induction Factor

SILAC – Stable Isotope Labeling of Amino Acids

siRNA – small-interfering RNA

SL – splice leader

TET – tetracycline

TOR4 –target of rapamycin kinase 4

tRNA – transfer RNA

VSG – variant surface glycoprotein

ZFK – zinc finger kinase

Table of Contents

ABSTRACT III

RESUMO V

ABBREVIATIONS IX

INTRODUCTION 1

TRYPANOSOMA BRUCEI 1

T. brucei life cycle *1*

Gene expression in T. brucei *2*

Stumpy differentiation *4*

LONG NON-CODING RNA 8

Non-coding RNAs in trypanosomes *9*

OBJECTIVES 10

METHODS 11

RESULTS AND DISCUSSION 14

CREATION OF A PLEOMORPHIC REPORTER CELL LINE FOR ANALYSIS OF STUMPY FORMATION 14

IN VITRO STUMPY FORMATION ASSAYS AND CHARACTERIZATION OF THE STUMPY-REPORTER CELL LINE 14

BIOINFORMATIC IDENTIFICATION OF CANDIDATE LNCRNAS IMPORTANT FOR SLENDER TO STUMPY TRANSITION 17

EXPRESSION PROFILE OF THE FOUR CANDIDATE LNCRNAS DURING STUMPY FORMATION 18

PHENOTYPIC SCREEN TO STUDY LNCRNA FUNCTION IN SLENDER TO STUMPY TRANSITION 19

ROLE OF LNCRNA5837 IN STUMPY FORMATION 21

GAIN- AND LOSS-OF-FUNCTION OF LNCRNA5837 22

LNCRNA5837 SUBCELLULAR LOCALIZATION 25

CONCLUSION AND FUTURE WORK 28

ANNEXES 30

REFERENCES 34

INTRODUCTION

TRYPANOSOMA BRUCEI

Trypanosoma brucei (*T. brucei*) is a protozoan parasite that causes N'gana in cattle and the fatal human disease, Human African trypanosomiasis (HAT), also known as sleeping sickness. This unicellular parasite is a member of the order Kinetoplastida, characterized by their mitochondrial DNA that forms a structure named kinetoplast. *T. brucei brucei* (*T. brucei* from here on) infects only livestock, while *T. brucei gambiense* and *T. brucei rhodesiense*, both causing sleeping sickness. Although 97% of cases of HAT are caused by *T. brucei gambiense*, *T. brucei rhodesiense* causes a more severe and acute illness. The distribution of the disease is governed by the distribution of the tsetse fly, the blood-feeding dipteran vector (*Glossina spp.*) that transmits the parasite[1].

The disease is characterized by a two stages progression: trypanosomes start to be restricted to the blood and lymph and, in a second stage, trypanosomes cross the blood-brain barrier to invade the central nervous system (CNS) (Fig. 1C), causing disturbance in sleep cycle, which gives the disease its name. Untreated patients can progress to a coma and organ failure, leading to death. Currently, a vaccine is not available and drugs for early-stage or late-stage are unavailable orally, frequently toxic and sometimes ineffective[2].

T. brucei life cycle

T. brucei parasite has a complex and sophisticated life cycle in two different hosts: a mammal and an insect vector, the Tsetse (Fig. 1).

After a bite of the tsetse fly, the growth-arrested metacyclic trypomastigotes parasites enter in the bloodstream of the mammal host. In the blood, the metacyclic forms differentiate quickly into long, slender trypomastigotes. Slender trypomastigotes reproduce asexually dividing by binary fission, thus establishing and maintaining a bloodstream infection. In the mammalian host, *T. brucei* evades the host immune system through a mechanism known as antigenic variation, which consists of periodic and rapid changes of the surface coat made of variant surface glycoproteins (VSGs)[3]. When the parasite population reaches high concentration in the blood, a quorum sensing-like mechanism elicits the differentiation of long slender forms into short stumpy forms, which are irreversibly arrested in their cell cycle.

Stumpy forms are the transmissible form pre-adapted for survival in the midgut of the tsetse vector. Stumpy forms, unlike slender forms that are rapidly killed in tsetse midgut, differentiate to a proliferative procyclic forms (PCF). At this stage the dense VSG coat is replaced by another dense coat of procyclins. Procyclic trypomastigotes initiate a long migration inside the tsetse host from the midgut to the salivary glands (blue line, Fig. 1B). Along the way, procyclic trypomastigotes undergo complex developmental changes and

divide asymmetrically to give rise to one long and one short epimastigote forms. Short epimastigote attaches to epithelial cells following arrival in the salivary glands. Attached epimastigotes are proliferative forms and differentiate into the non-dividing metacyclic trypomastigote to complete the cycle. Metacyclic trypomastigote expresses *de novo* the VSG coat and therefore are infective for a new mammalian host [4].

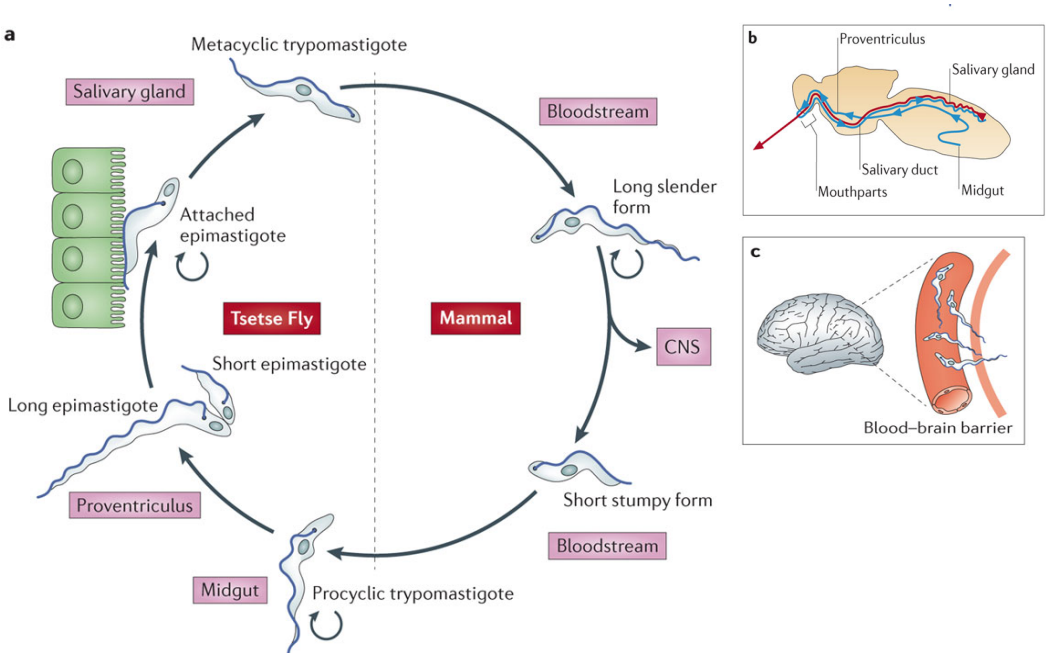


Figure 1: *T. bucei* life cycle. (a) The parasite during his life cycle faces many different environments: enters in the blood of mammalian host by a bite of the tsetse fly, pass to the midgut of the vector in a tsetse blood meal, and then migrate to the salivary glands of tsetse fly to enter in a new mammalian host. (b) Parasite movement inside the tsetse fly. The blue line represents the migration route of procyclic form from the midgut to the salivary gland and the red line indicates the migration route from the salivary gland to the mouthparts. (c) Parasites move out of blood vessels, penetrating the blood-brain barrier to enter the CNS. Adapted from Langousis *et al*., 2014[5].

Gene expression in *T. brucei*

T. brucei genome contains 26Mbp and is divided into 11 large, 3-5 intermediate and 100 mini-chromosomes. Genome architecture of trypanosomes is unusual compared with other eukaryotic cells in which genes are arranged polycistronic transcription units (PTUs)[6], long arrays of genes that are transcribed via a common and unique promoter.

Transcription and mRNA processing

Transcription of protein-coding genes in *T. brucei* differs from other eukaryotes, due to its unusual genome organization. Indeed, once long polycistronic precursor RNA (pre-RNA) is transcribed, the transcript is immediately processed into small mature mRNAs. Pre-RNAs are processed into mature monocistronic mRNAs by *trans* splicing reaction, which adds a 39 nucleotide spliced leader sequence to the 5' end, downstream of a pyrimidine rich motif, and

polyadenylation of the 3' end. Polyadenylation is coupled to the addition of cap 5' of the downstream ORF[7,8].

T. brucei is also the only known eukaryote with a multifunctional RNA polymerase (pol) I which synthesizes both ribosomal and mRNA. Pol I transcribes mRNAs encoding the proteins that cover the surface of the parasite, such as the VSG and procyclin coats[9].

Genes within a PTU are most often not functionally related. Except VSG and procyclin genes transcribed by Pol I[9], there is no evidence of transcription regulation in *T. bucei*. Hence, gene expression is controlled at a post-transcriptional level[10].

Post-transcriptional control of gene expression

Genome-wide analysis of mRNA abundance between *T. brucei* life stages shows that parasite differentiation is associated with large changes in gene expression[11–15]. Notably, it has been recently shown that 40% of the protein-coding gene transcripts are significantly regulated between mammalian and insect stages of *T. brucei*[16]. Moreover, a recent proteomic study using stable isotope labeling of amino acids (SILAC) has shown that 45% of *T. brucei* proteins change in abundance during parasite development[17]. Surprisingly, only a moderate correlation was found between the transcriptome and the proteome[17,18], suggesting the importance of post-transcriptional and translational mechanisms to control protein levels in *T. brucei*. Hence, in *T. brucei*, absence of transcriptional control leads to constitutive gene expression[10], which are regulated at a post-transcriptional level, notably by mRNA stability and translational control.

mRNA stability and translational control

In metazoan, mRNA stability is controlled by the action of microRNAs and RNA binding proteins (RBPs). In trypanosomes, mRNA turnover results mainly from the balance between RNA degradation (1) and mRNA stabilization by RBPs (2).

(1)*T. brucei* has two main pathways for mRNA degradation. The first pathway is initiated by deadenylation and then decapping resulting in 5'-3' exonuclease activity by a cytoplasmic exonuclease, XRNA, and 3'-5' degradation by a exonuclease complex, the exosome[10]. The second pathway involves degradation of highly unstable mRNAs dependent on XRNA, without the requirement for deadenylation[19,20].

(2) Secondary structures in 3'UTRs (3' untranslated region) contribute to mRNA stability and recruitment of RBPs. 3'UTR regulatory elements are binding sites for *trans*-acting RNA binding proteins to regulate mRNA stability.

Besides mRNA turnover, gene expression is controlled at translational level. In addition to regulate mRNA stability, RBPs can bind to 3'UTR of RNA to control association of the mRNA with the translation machinery[22,23]. For example, procyclic-specific genes in *T. brucei* have

regulatory regions in 3'UTR that control both translation efficiency and mRNA stability, which cause procyclic-specific genes to be unstable in bloodstream forms[24].

Finally, it has been demonstrated a correlation between codon usage and gene expression, suggesting that translational selection through codon *bias* is an important mechanism of differential gene expression in trypanosomes[25]

Stumpy differentiation

Stumpy forms

Stumpy forms are produced by irreversible differentiation of long proliferative slender forms[1]. While long and thin slender form measures on average 29 µm, stumpy forms are thick and short cells with no-free flagellum measuring about 18µm in length. Stumpy forms are cell-cycle arrested in G1/G0 and have a relatively short half-life estimated to be between 48 to 72 hours[26]. Stumpy forms do not switch VSG, in fact, VSG transcription in stumpy forms is suppressed below the limit of detectability[27]. Finally, it has been shown that transcription and protein synthesis are globally down-regulated in stumpy forms[28], reinforcing their *status* of quiescent cell.

The slender-to-stumpy transition

The slender-to-stumpy transition is a key process for the success of *T. brucei* infection and transmission mainly for three reasons: (1) to control parasite growth and host survival, (2) to regulate the host immune system and to "protect" the VSG repertoire of parasite and (3) to enable parasite transmission and infection inside the insect vector.

(1) To control parasite growth and host survival

Stumpy formation relies on a cell-density dependent mechanism. When the number of resident proliferating slender rise above a certain threshold, parasites sense their density and irreversibly differentiate into stumpy forms. At this time of infection, parasitaemia irremediably decrease as newly formed stumpy parasites are non-diving and non-differentiated slender parasites are eliminated by the host immune system through an antibody response against VSG antigen. However, residual slender forms that underwent antigenic variation will be able to escape the host immune response, proliferate and rise up once again the parasitaemia. Thus, the balance of slender and stumpy forms ensures the chronicity of *T. brucei* bloodstream infection, which is characterized by successive waves of parasitaemia. In contrast, a laboratory adapted cell line of *T. brucei*, called monomorphic strain, has lost the ability to form stumpy forms in response to cell-density. As consequence, mouse infection

with a monomorphic strain results in a uncontrolled parasite proliferation and consequent rapid death of the host[29].

Therefore, equilibrium between slender and stumpy forms within mammalian blood is the key of success for *T. brucei* infection. While slender forms establish and maintain the infection, stumpy formation controls parasitaemia at a viable level for the host.

(2) To regulate host immune system and "protect" the VSG repertoire

In the chronic phase of *T. brucei* infection, intermediate and stumpy forms represent 73%-97% of total parasites in the blood[30]. As consequence, stumpy cells are reducing the overall proportion of proliferative slender forms in the mammalian blood during chronic infections. As proliferative forms are the only forms undergoing productive antigenic variation, reducing their number would decrease the overall frequency of antigenic variation events[30]. Therefore, antigen repertoire would be preserved, herd immunity restricted and host reinfection frequency increased[1].

(3) To promote parasite transmission

Stumpy trypanosomes are pre-adapted forms for life (A), survival (B) and differentiation (C) in the tsetse midgut.

(A) Host environment changes drastically between mammalian blood and insect midgut, notably in their nutrient supplies. Unlike the mammalian blood, insect midgut is a glucose-poor and poorly oxygenated environment[31]. As consequence, energy production switches from glycolysis in slender form to proline metabolism in procyclic forms[32]. Stumpy forms are pre-adapted for metabolic changes as their mitochondria become more elaborated and ramified, developing large christae and expressing crucial enzymes for Krebs cycle, respiratory chain and oxidative phosphorylation function[33,34].

(B) Stumpy forms are more robust than slender forms to the host antibody response developed during *T. brucei* infection. Immunoglobulin-VSG complexes formed at the parasite cell surface are endocytosed twice rapidly by stumpy than slender forms[35] which promotes stumpy survival in the blood and increase time for parasite transmission.

Unlike slender forms, stumpy parasites resist and survive to acidic and proteolytic stress encountered inside the insect vector. Hallmarks of stumpy form, parasite regulates its internal pH to counterbalance the extreme acidic environment of the tsetse midgut and retains vestige of its VSG coat after proteolytic cleavage to "escape" the innate immunity of the insect vector[36].

(C) Stumpy forms are uniformly arrested in G1/G0 phase of the cell cycle, which ensures, when taken up by tsetse flies, a coordinated differentiation into procyclic forms and re-entry into the cell cycle. Hence, cell cycle arrest in stumpy form is crucial step for parasite

transmission capacity[37]. *T. brucei* respond to citrate or *cis*-aconitate (CCA) to initiate differentiation to procyclic forms when transmitted to the insect vector. A gene family, PAD (Protein Associated with Differentiation), encoding surface carboxylate transporters was recently identified, including PAD1 and PAD2 that were shown to be required for CCA perception[38]. PAD1 is expressed only in stumpy forms and PAD2 is also expressed in procyclic forms. This differential expression of PAD proteins between slender and stumpy forms ensures that only pre-adapted stumpy forms are able to perceive the differentiation signal. It was also demonstrated that PAD2 is released to cell surface only at 20°C, consistent with a previous study[39] that showed a increase sensitivity to CCA at 20°C. The differential localization of PAD2 at different temperatures ensures that only stumpy cells in the insect vector are able to receive the signal.

Although little is known about the molecular mechanism behind slender to stumpy transition, this transition is a crucial step during *T. brucei* life cycle because it helps controlling parasitaemia, prolonging host survival and promoting efficient parasite transmission.

Quorum sensing-like mechanism for stumpy formation

The slender to stumpy transition is related with increased cell density inside the mammalian host, but can occur independently of the host[40–42]. This transition is proposed to operate by a parasite-parasite signaling via a quorum sensing mechanism, in that slender forms secrete a soluble, low molecular weight, heat stable factor, termed Stumpy Induction Factor (SIF), which accumulates and stimulates stumpy formation as parasite numbers increase[41].

Cell permeable cAMP analogue, pCPT-cAMP, induces the formation of stumpy-like parasites. Treatment with pCPT-cAMP induces cell cycle arrest, PAD1 expression and promotes a synchronized differentiation to procyclic forms in response to CCA[43]. Although cells treated with this stumpy-inducing chemical do not develop a full stumpy morphology[43], treatment with pCPT-cAMP is a well characterized method to study stumpy formation *in vitro*.

Quorum sensing signaling pathway

Several genes essential for the quorum sensing signaling pathway were recently identified using a genome-wide RNAi screening strategy[44]. This screen identified components that positively act at different steps of the SIF signaling pathway, including cell signaling molecules, like phosphatases and kinases, and post-transcriptional gene regulators, such as RNA binding proteins[44,45].

Serine/threonine protein phosphatases of type 1 (PP1) and kinases YAK and NEK were all experimentally confirmed to be involved in *in vivo* stumpy formation. The authors

proposed that these phosphatases may be involved in deactivation of stumpy inhibitors. Protein kinases have been identified as major regulators of *T. brucei* cell cycle [46]. Thus it has been proposed that NEK and YAK may induce cell cycle arrest, as it has been previously shown in other organisms[45,47–49].

In trypanosomes, RNA binding proteins (RBPs) have been identified as key players in different life stages[50,51] due to their role in post-transcriptional regulation. RBP7 (representing two identical genes: RBP7A and RBP7B) was the only RBP experimentally demonstrated to be necessary for stumpy differentiation. Knock-down of RBP7 promotes upregulation of histone transcripts, reflecting the presence of proliferative slender forms. When overexpressed, RBP7 increased the abundance of others RBPs and transcripts associated with procyclic forms, showing a preparation to the next life stage[44]. RBP7 and other hypothetical proteins, identified in a recent genome wide screen for post-transcriptional regulators identification[52], should regulate genes essential for stumpy formation and differentiation to procyclic forms by post-transcriptional changes.

The RNAi strategy used in the previously mentioned study[44] only allowed the identification of positive regulators of slender to stumpy transition. Earlier studies identified negative regulators of stumpy formation such as a zinc finger kinase (ZFK), MAPK5 and a TOR kinase (TOR4)[53–55]. In all cases, RNAi of each gene drives to stumpy formation in pleomorphic cells. Both ZFK and MAPK5 RNAi increase stumpy formation *in vitro*, but MAPK5 RNAi phenotype was also seen *in vivo*. RNAi of TOR4 triggered stumpy formation also in monomorphic cells.

Markers for the stumpy life stage

In previous studies, different markers have been used to identify stumpy forms, including morphology, mitochondrial activation (expression of NADH dehydrogenase and dihydrolipoamide dehydrogenase), disappearance of the small subunit of ribonucleotide reductase, inhibition of growth and DNA synthesis, cell-cycle arrest in G1/G0 phase and the kinetics of transformation to procyclic forms[42,56,57]. However these methods are not ideal, some of the assays are subjective and time consuming or require population assays, rather than analysis of individual cells. Recently, PAD proteins were identified as transducers of the CCA differentiation signal. These proteins can act as a molecular marker for stumpy formation and a functional marker for transmission[38]. PAD1 is up-regulated at protein and mRNA levels in stumpy forms[13,38]. PAD1 mRNA start to be detected in intermediate forms[30,43], but protein levels has a more stringent control and protein expression is only found in stumpy forms[38]. *PAD1* 3'UTR is sufficient to control stumpy-specific gene expression. This has allowed the creation of a PAD1::GFP reporter strain in which GFP gene is under the control of a *PAD1* 3'UTR. GFP fluoresces only in stumpy forms[43].

Although, in trypanosomes, RBPs have been described as key players in the parasite life stages and important for mRNA stability and translation[21,50,51], no miRNAs or other non-coding RNA have been shown to be involved in developmental regulation of trypanosomes. As described in more detail below, the hypothesis of this thesis was that lncRNAs may be involved in *T. brucei* differentiation from slender to stumpy forms.

LONG NON-CODING RNA

One of the most surprising discoveries from the last decade has certainly been the finding that the vast majority of eukaryotic genome is transcribed into non-coding RNAs (ncRNAs). Indeed, it is now clear that organism complexity cannot be explained by the number of protein-coding genes that genomes encode, but in its non-coding component.

The most prominent examples of non-coding RNAs are transfer RNA (tRNA) and ribosomal RNA (rRNA), which are both crucial players in mRNA translation process. More recently, a new class of non-coding RNA has been identified as long ncRNAs (lncRNAs). The most common definition of lncRNAs refers to a transcript >200bp in length and that does not encode information for proteins.

LncRNAs can be classified based on their genomic location or/and their transcription orientation with respect to the nearest protein-coding genes: (1) sense or antisense, located on the same or opposite strand of the nearest protein-coding genes; (2) intronic or intergenic, located inside an intron of protein-coding gene or in the region between two protein-coding genes; (3) divergent or convergent, transcribed in divergent or convergent orientation as the nearest protein-coding genes [58,59].

LncRNAs have emerged as master regulators of gene expression in all kingdoms of life, by controlling transcription, mRNA stability or translation efficiency[60] (Fig. 2). For example, the lncRNA BACE1AS (BACE1 antisense RNA) increase the stability of its sense mRNA by inhibition of microRNA (miRNA) function. BACE1AS and BACE1 (beta-site APP-cleaving enzyme 1) mRNA form an RNA-RNA duplex, which mask miRNA-binding sites on BACE1 mRNA and prevents its degradation by miRNA-inducing silencing "pathway"[61].

LncRNAs can act *in cis* and *in trans* as well as within various cell compartments such as the nucleus and the cytoplasm, to regulate gene expression and function[62,63]. Promoter-associated RNA (pRNA) is a well-studied *cis*-acting nuclear lncRNA which regulates the expression of ribosomal genes in many eukaryotic cells. pRNA is 250-300 nucleotide transcript originating from the intergenic spacer region of rDNA gene units. pRNA recruits enzymes for heterochromatin formation and mediates transcription silencing of the downstream rRNA gene unit through *a cis* mechanism[64].

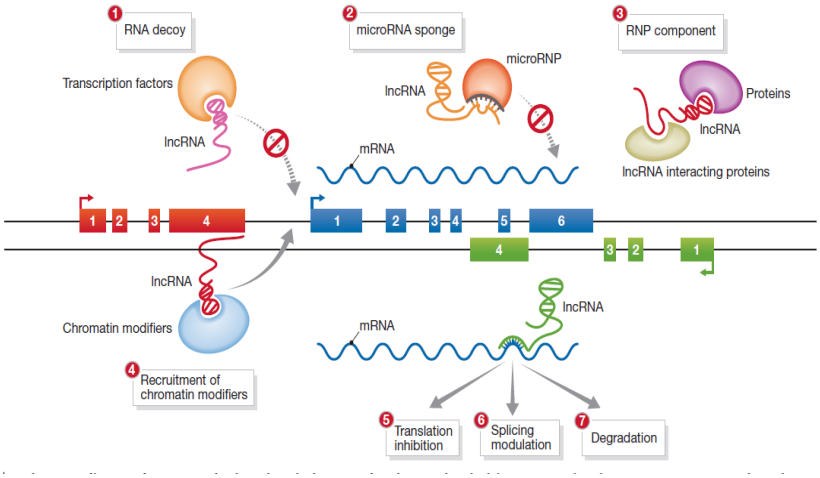


Figure 2: Mechanisms of lncRNA function. Studies have described several mechanisms by which lncRNAs regulate their targets: (1) lncRNAs can act as RNA decoys and prevent transcription factors to bind to their DNA targets; (2) at a post-transcriptional level, lncRNAs can act as miRNA sponges that prevent miRNA complexes to degradate their mRNA targets; (3) lncRNAs seem to interact with other proteins, acting as scaffolds of ribonucleoprotein complexes; (4) recruitment of chromatin-modifying complexes of DNA targets is a well characterized function of several lncRNAs; a few lncRNAs seem to act directly at their mRNA targets, including in translation (5), splicing (6) and degradation (7). Adapted from Hu et al., 2012[65].

Although lncRNAs are not unstable, their half-life are usually shorter than protein-coding RNAs[66]. Nuclear-localized lncRNAs are less stable than lncRNAs localized in the cytoplasm, suggesting that stability depends on the cellular localization of lncRNAs, rather than on their genomic classification[66]. The low stability of nuclear-localized lncRNAs is consistent with the model that these lncRNAs act as regulatory molecules, which require a rapid turned over to regulate dynamically gene expression.

LncRNAs are versatile and multifunctional molecules that regulate a myriad of cellular functions from gene expression[60] to more complex biological processes such as cell differentiation[67].

Non-coding RNAs in trypanosomes

Small non-coding RNAs such as small-interfering RNA (siRNA) have been identified in *T. brucei* and originated mainly from long double-stranded RNAs from the retroposons INGI and SLACS and from the CIR147 family of satellite-like repeats[68,69]. The parasite has a functional RNA interference (RNAi) pathway that relies on one single Argonaute protein and two distinct dicer-like enzymes (DCL). Surprisingly, *T. brucei* has compartmentalized its RNAi pathway between the nucleus and cytoplasm[70]. Nuclear RNAi pathway and siRNA are believed to be the guardian of genome stability in *T. brucei* by decreasing retroposon and repeat transcripts and potential retroposon hopping[71].

Unlike small non-coding RNAs, little is known about long non-coding RNA in *T. brucei*. Only a recent transcriptomic analysis has revealed existence of 103 lncRNAs with size ranging from 154–2229bp[15]. However, no experimental data has confirmed either existence or function of lncRNA in *T. brucei*.

Luisa Figueiredo laboratory has taken up the challenge to identify and characterize lncRNAs in *T. brucei*. RNA-seq was performed using two different life cycle stages of *T. brucei* (bloodstream and procyclic forms) and using four different RNA isolation procedures (Nuclear, cytoplasmic, poly A+ and poly A- RNA fractions). Bioinformatic analysis of the RNA-seq data identified 946 lncRNAs, trans-spliced and polyadenylated, with size ranging from 161-9190bp. Interestingly, 567 LncRNAs were identified in only one of the parasite stage (379 lncRNA in bloodstream form, 188 in procyclic form), suggesting the existence of stage-specific lncRNAs.

The discovery of lncRNA genes opens new avenues not only to study how gene expression is regulated in *T. brucei*, but also to understand the molecular mechanism behind parasite differentiation.

OBJECTIVES

The aim of this thesis is to study the role of lncRNAs in stumpy formation. For that, we first established in our lab the system that uses a previously published stumpy-reporter cell line and we validated it. Next, we used this assay to test the role of four lncRNAs in slender to stumpy transition and in gene expression during a differentiation process.

The main aims are:

1. To create a stumpy-reporter cell line
2. To identify potential lncRNA involved in stumpy formation
3. To quantify potential lncRNAs during stumpy formation
4. To localize lncRNA inside the cell
5. To determine the role of lncRNA during stumpy formation, using gain or loss of function approaches

METHODS

Strains and cultures. *T. brucei* AnTat 1.1 90:13 was used in this project. The plasmid p4231 containing a GFP-PAD1 3'UTR cassette was used to create a stumpy reporter cell line, named GP1 (AnTat 1.1 90:13 with GFP-PAD1 3'UTR). The plasmid integrates in β-tubulin locus and uses PolII transcription. A nuclear localization signal (NLS) is downstream of GFP gene to concentrate GFP protein into the nucleus and have a strong GFP signal. AnTat 90:13 was cultivated in HMI-11 with Neomycin 2,5μg ml^{-1} and Hygromycin 5μg ml^{-1} at 37°C and 5% CO_2. AnTat 90:13 with GFP-PAD1 3'UTR was cultivated in HMI-11 with Neomycin 2,5 μg ml^{-1}, Hygromycin 5μg ml^{-1} and Blasticidin 5μg ml^{-1} at 37°C and 5% CO_2. Both cell lines were maintained at $<5x10^5$ cells ml^{-1}.

***In vitro* stumpy formation.** We tested two different methods of *in vitro* stumpy formation: HMI-11 medium with 1.1% of Methylcellulose and HMI-11 medium with 100μM of 8-(4-chlorophenylthio)-cAMP (pCPT-cAMP). To prepare the medium with methylcellulose, 11g of Methylcellulose were resuspend in 366 ml of water and filled with HMI-11 1,6x concentrated up to 1L. Stumpy forms were induced by cultivating the cells (initial concentration: $2x10^5$ cells ml^{-1}) for two to three days at 37°C in methylcellulose-HMI-11 medium or HMI-11 with 100μM of pCPT-cAMP.

Transfection of cells with vectors.[72] For the transfection, $10x10^6$ cells were harvested by centrifugation (1800 rpm, 10 minutes at 4°C). The cell pellet was then resuspended in 100 μl of Roditi Buffer (90 mM $NaPO_4$, 5 mM KCl, 50 mM HEPES, 0.15 mM $CaCl_2$, pH7.3) with 2 μg of linearized vector, transferred to Amaxa cuvette and transfected with Amaxa Nucleofector® II using X-001 program. After transfection, cells were immediately transferred into 25 ml of pre-warmed HMI-11 medium. Parasites were plated in 24-well-plate using 1:1, 1:10 and 1:100 dilutions. After six hours incubation, selective drug was added to culture medium. Approximately 5-8 days post-transfection resistant clones became detectable and were transferred to fresh selective media and cell were kept $<5x10^5$ cells ml^{-1}.

RNAi-mediated gene knockdown. RNAi fragments were designed to target 300-400bp or 80% of the target gene. To knock-down target lncRNAs, fragments were amplified from AnTat 1.1 90:13 genomic DNA using forward primer containing BamHI restriction site and reverse primer containing HindIII restriction site for cloning into p2T7-177 plasmid[73] using In-Phusion HD cloning kit protocol. To knock-down PP1 gene, we used pALC14 plasmid, kindly provided by Keith Matthews[44]. The plamids were linearized with NotI and transfected into AnTat 1.1 90:13 by electroporation. Transfected cells with pALC14 plasmid were selected in the presence of 1μg ml^{-1} puromycin and cells with p2T7-177 were selected in the presence of 2,5μg ml^{-1} phleomycin.

Gene overexpression. To overexpress RBP7[44] and lncRNAs, we used a pDex577-Y plasmid that produces a high-level expression of a transgene by a tetracycline-inducible T7 promoter[74]. The DNA sequence of each gene was amplified from AnTat 1.1 90:13 genomic DNA using forward primer containing HindIII restriction site and reverse primer containing BamHI restriction site for cloning into pDex577-Y using In Phusion HD cloning kit protocol. Transfected cells were selected in the presence of 2,5µg ml^{-1} phleomycin.

Cell cycle analysis. $1x10^6$-$2x10^6$cells of BSF were harvested and washed with ice-cold PBS. Cells were resuspended in 1ml PBS/2mM EDTA and fixed by drop wise addition of 2.5ml of ice cold 100% ethanol while vortexing. After 1 hour at 4˚C, the cells were washed and ressuspended in 1ml of PBS with 2mM of EDTA. Subsequently,1µl of RNaseA (10µg $µl^{-1}$) and 1µl of propidiumiodide (1mg $µl^{-1}$) were added and cells were incubated 30 min at 37˚C. After PI staining, cells were analyzed by flow cytometry.

Flow Cytometry. $5x10^5$ cells were washed twice with TDB (5mM KCl, 80mM NaCl, 1mM $MgSO_4$, 20mM Na_2HPO_4, 2mM NaH_2PO_4, 20mM glucose, pH 7.4) and resuspended in 500µl of TDB. Percentage of GFP expressing GP1 cells were analyzed using a Becton Dickinson LSR Fortessa Flow cytometer and FlowJo software.

RNA extraction and cDNA synthesis. RNA was extracted from cell pellets using a culture of 20ml at $5x10^5$ cells ml^{-1}. Cells were harvested by centrifugation (1800 rpm, 10 min at 4°C), resuspended in TRIzol (Life Technologies) and RNA was extracted according to the manufacturer's instructions. RNA samples were quantified with NanoDrop 200 Spectrophotometer (Thermo Scientific). Purified RNA (2µg) was subjected to DNase treatment prior to cDNA synthesis. Reverse transcription was performed using SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) and random primers, accordingly to the manufacturer's protocol with 8µl of RNA (~300ng) in 20µl reaction volume.

Quantitative real-time PCR. qPCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) and specific primers for each target gene in 10µl reaction volume.

RNA FISH. RNA FISH probes were designed using Stellaris Probe Designer software available online. 1-$5x10^6$ cells were collected by centrifugation (1800 rpm, 10 min at 4°C), washed with 1xPBS (RNase free) and incubated at room temperature in Fixation Buffer (3.7% formaldehyde). After 10 minutes incubation, cells were washed twice with 1xPBS, resuspended in 500µl of PBS and transferred to a FluoroDish (World Precision Instruments, Inc). After 15-20 min, cells adhered and PBS was removed. To permeabilize cells, cells were immersed in 70% ethanol overnight. Ethanol was removed and Wash Buffer A (Biosearch technologies) was added to the cells for 5 minutes at room temperature. Then, Wash Buffer was removed, 200µl of Hybridization Buffer containing 4µl of RNA FISH probe was added

and cells incubated in the dark at 37°C overnight within a humidified chamber. Cells were washed using Wash Buffer A and incubated 30 minutes at 37°C in the dark. Finally, DNA of the parasite were stained using DAPI (5ng/ml DAPI in Wash Buffer A) during 30 min at 37°C in the dark. Zeiss Axiovert 200M wide-field microscope and Metamorph software was used to visualized and acquire images of several samples. Images were post produced using ImageJ.

RESULTS AND DISCUSSION

CREATION OF A PLEOMORPHIC REPORTER CELL LINE FOR ANALYSIS OF STUMPY FORMATION

In previous studies it was shown that PAD1 3' UTR is sufficient for stumpy-specific expression[43]. Therefore it is possible to create a stumpy-specific reporter cell line coupling PAD1 3'UTR to a reporter gene. This would provide a method to study stumpy formation *in vitro* or *in vivo*.

A fluorescent reporter gene, GFP, was coupled to PAD1 3'UTR to create a stumpy-specific reporter gene. This reporter construct was transfected into AnTat 1.1 90:13 pleomorphic cell line. Several clones were collected after transfection and one of the clones, named GP1, was selected after being tested for stumpy differentiation in HMI-11 with methylcellulose.

IN VITRO STUMPY FORMATION ASSAYS AND CHARACTERIZATION OF THE STUMPY-REPORTER CELL LINE

Induction of stumpy formation in *T. brucei* has been previously established *in vitro* using two different protocols (see materials and methods):

1. Culture of a pleomorphic strain of *T. brucei* in HMI-11 medium supplemented with 1.1% of methylcellulose.
2. Culture of a pleomorphic or monomorphic strain of *T. brucei* in HMI-11 medium supplemented with 100μM of the stumpy-inducing drug, 8-(4-chlorophenylthio)-cAMP (pCPT-cAMP)[43,44].

The suitability and efficiency of these two protocols were compared using the pleomorphic AnTat 1.1 90:13 strain and the stumpy-reporter cell line, GP1, of *T. brucei*. The two *in vitro* culture protocols were carried on during a maximum of three days, cells were collected at different time points and the number of stumpy cells were quantified using three different methods, by measuring:

1. the expression of PAD1 mRNA levels using qPCR.
2. the percentage of cells in G1 phase of the cell cycle, using flow cytometry.
3. the percentage of GFP-positive cells (stumpy-reporter cell line) using flow cytometry.

We observed that the stumpy-reporter cell line, GP1, and the parental cell line, AnTat 1.1 90:13, have similar growth rates, showing that the GFP-PAD1 3'UTR construct does not affect GP1 growth (Fig. 3A). After GP1 parasites were exposed to 100μM pCPT-cAMP for 24 hours, we observed an arrest in cell growth compared with control cell cultures without pCPT-cAMP (Fig. 3B). This was expected since pCPT-cAMP induces differentiation of slender forms to G1-arrested stumpy forms. Medium with methylcellulose mimics blood

environment, allowing a faster growth of GP1 cells in the first day. However, proliferation of GP1 parasites reached a stationary phase after one day in methycelluloase medium, suggesting formation of stumpy cells as observed in medium with pCPT-cAMP (Fig. 3B).

AnTat 1.1 90:13 and GP1 cells show the same cell cycle profile in different culture conditions (Fig. 3C, D). After two to three days in medium with methylcellulose and/or pCPT-cAMP, GP1 and AnTat 1.1 90:13 show an accumulation of 18-32% of cells in G1 phase compared with control cultures cultivated without methylcellulose and not exposed to pCPT-cAMP. Growth in medium with methylcellulose promoted higher levels of G1-arrested parasites than exposition to pCPT-cAMP (Fig. 3C, D). Stumpy-reporter cell line GP1 should express GFP only in conditions that promote stumpy formation. Indeed, only GP1 cells cultivated in medium with methylcellulose and/or with pCTP-cAMP express GFP, while no GFP expression is observed in cells cultivated with no supplemented medium (Fig. 3H).

To test the response of GP1 cell line to SIF signal *in vivo*, mice were infected with GP1 cell line and at 6 days post-infection parasites were collected from the blood, when trypanosomes are predominately stumpy forms[30]. As expected, parasites show an arrest in G1 phase and the majority of cells (82%) express GFP (Fig. 3G, H). As anticipated, GFP-positive cells are arrested in G1 phase and GFP-negative cells show a normal cell cycle profile, similar to GP1 cells cultivated in HMI-11 medium (Supplementary Fig. 1). Therefore this data shows that GP1 cell line has a stumpy-specific GFP expression both *in vivo* and *in vitro*.

PAD1 and GFP mRNA levels were quantified by qPCR and both PAD1 and GFP show the same expression pattern in cells treated with pCPT-cAMP: in the first 36 hours of pCPT-cAMP exposure, PAD1 and GFP mRNA levels progressively increase compared with the control cells without drug treatment, but after 48 hours with pCPT-cAMP both mRNA levels decrease compared with the 36 hours time point (Fig. 3E). A decrease in gene expression after 48 hours with pCPT-cAMP may be expected due to repression of transcription and translation during transition from slender to stumpy forms[28,75,76]. Although after a 24h treatment with pCPT-cAMP it is possible to see an increase in PAD1 mRNA, it is not possible to observe a G1 arrest (Supplementary Fig. 2). PAD1 and GFP (under control of PAD1 3'UTR) mRNA levels start to be detected early in differentiation to stumpy forms, while morphological changes are observed in the end of stumpy differentiation process, as previously described[30,43].

Based on the cell cycle profile, culture in medium with methylcellulose is the best method to differentiate stumpy trypanosomes and when both methods are combined more G1-arrested cells are observed (Fig. 3C, D). Although methylcellulose medium induces more cell cycle arrested cells than pCPT-cAMP, methylcellulose is a viscous medium that requires several washing steps to prepare the cells for further experimental assays. Moreover, *in vitro*

induction of stumpy formation in the presence of pCPT-cAMP is well characterized using stumpy-reporter cell lines[43]. Thus medium with 100μM pCPT-cAMP was the selected method to study stumpy formation.

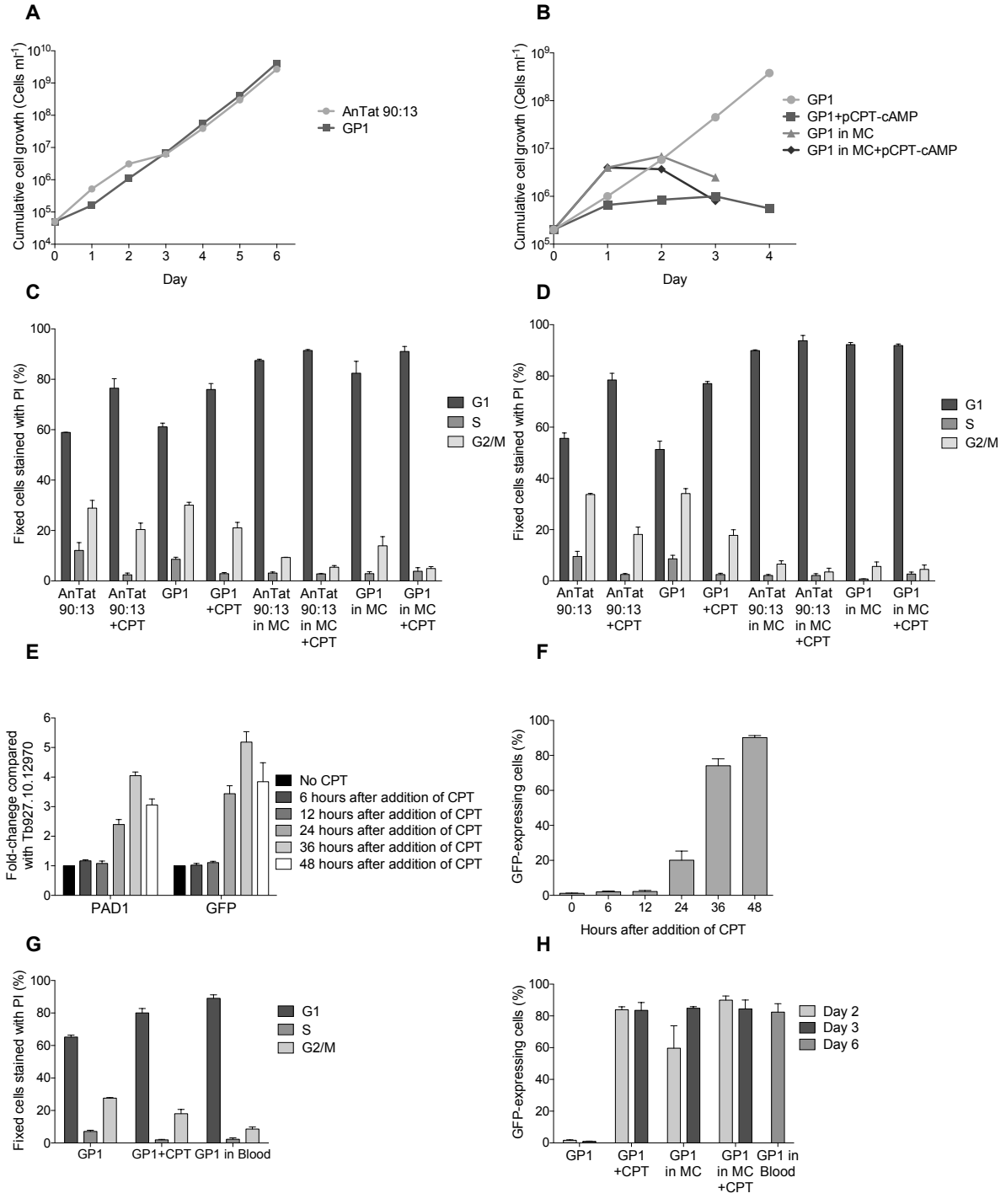


Figure 3: *In vitro* differentiation to stumpy forms: cell cycle profile and GFP expression. (A) Growth curve of GP1 and parental cell line, AnTat 90:13 cultivated in HMI-11 medium. **(B)** Cell growth of GP1 cultivated in HMI-11 medium (GP1) and in medium condition that promote stumpy formation: HMI-11 supplemented with pCPT-cAMP (GP1+pCPT-cAMP), supplemented with methylcellulose (GP1 in MC) and supplemented with methylcellulose and pCPT-cAMP (GP1 in MC+pCPT-cAMP). Cell cycle profile of AnTat 90:13 and GP1 cell line cultivated two days **(C)** or three days **(D)** in HMI-11 medium (AnTat 90:13 or GP1), HMI-11 medium with pCPT-

cAMP (CPT), HMI-11 medium with methylcellulose (MC) and in HMI-11 medium with pCPT-cAMP and methycellulose (MC+CPT) (n=3, mean ± s.e.m.). PAD1 and GFP mRNA levels **(E)** and percentage of cells expressing GFP **(F)** after addition of pCPT-cAMP (CPT) in the culture medium. Bars represents mean ± s.e.m; n=3. **(G)** Cell cycle profile of cells collected from blood 6 days post-infection (GP1 in blood) and cells cultivated two days in HMI-11 medium (GP1) and HMI-11 medium with pCPT-cAMP (GP1+CPT) (n=3, mean ± s.e.m.). **(H)** Percentage of cells expressing GFP after two and three days of culture in HMI-11 medium (GP1), HMI-11 medium with pCPT-cAMP (CPT), HMI-11 medium with methylcellulose (MC), HMI-11 medium with methylcellulose and pCPT-cAMP (MC+CPT) or cells collected from blood 6 days post-infection (n=3, mean ± s.e.m.).

Altogether, stumpy-reporter GP1 cell line responds to cell density *in vivo* and *in vitro* (in medium with methylcellulose) and differentiate with pCPT-cAMP, showing stumpy-specific GFP expression both *in vitro* and *in vivo*. Therefore, this data validates GP1 as a reliable stumpy-reporter cell line that allows us to easily follow stumpy differentiation and quantify the number of stumpy forms by measuring the percentage of GFP-positive cells.

BIOINFORMATIC IDENTIFICATION OF CANDIDATE LNCRNAS IMPORTANT FOR SLENDER TO STUMPY TRANSITION

The transcriptome profile of *T. brucei* slender and procyclic forms was analyzed by RNA sequencing (RNA-Seq) using a paired-end and strand-specific sequencing approach. Using bioinformatic analysis of the RNA-Seq data, Luisa Figuereido's laboratory identified 946 lncRNAs, in which 379 and 188 lncRNAs were exclusively identified in slender and procyclic forms, respectively.

Recently, a genome-wide RNAi screen identified 30 protein-coding genes essential for *T. brucei* to successfully differentiate in stumpy forms. Moreover, using a loss-of-function approach the authors validated experimentally that nine of these predicted genes showed resistance to SIF *in vivo*[44].

Should any of the *T. brucei* lncRNAs be important regulators of slender to stumpy transition, we hypothesize them to regulate the expression and/or function of the nine validated genes essential for stumpy formation. We identified four lncRNAs located directly upstream or downstream to three of the nine essential genes for stumpy formation (Fig. 4), suggesting that these lncRNAs could regulate the expression of stumpy essential genes via an *in cis* mechanism.

LncRNA5388 and lncRNA5389 are located downstream of *NEK* family genes (Tb927.10.5930 and Tb927.10.5940). LncRNA5837 is located upstream of *RBP7* family genes, but located within an upstream gene with unknown function (Tb927.10.12080). LncRNA6099 locates upstream of *YAK* gene. Moreover, except the lncRNA5389, all the other three lncRNAs overlap partially or completely with the annotated UTR of its nearby gene (Fig. 4).

To confirm that these lncRNAs were independent transcripts and not the result of alternative splicing from the UTR of an adjacent gene, the expression profiles of lncRNAs and its nearby genes were compared using different RNA-Seq data. None of the four lncRNAs showed a positive correlation with its nearby genes, confirming their independency as a transcript (Supplementary Fig. 3). Interestingly, a negative correlation was found between lncRNA6099 and YAK gene (Pearson correlation r = -0.82), suggesting a potential anti-regulation mechanism (Supplementary Fig. 3).

From our bioinformatic analysis, we have selected four lncRNAs and we postulated them to play an important role during slender to stumpy transition by regulating the expression of its nearby genes, essential for stumpy formation, via *in cis* mechanism.

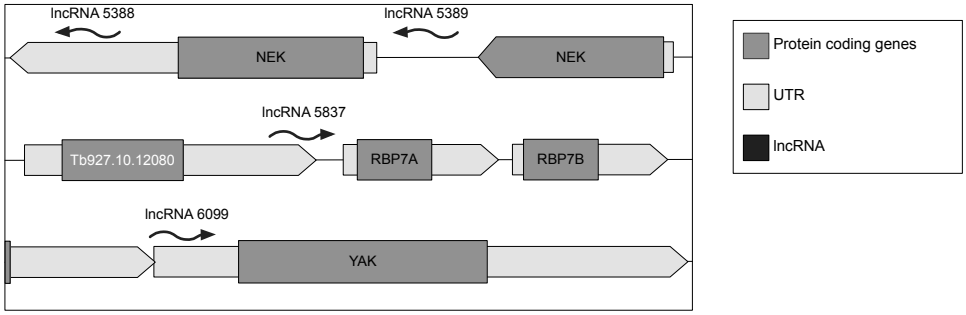


Figure 4: Genomic localization of candidate lncRNAs. LncRNA5388 is defined as 375bp in length and is located within the annotated 3'UTR of the NEK gene (Tb927.10.5930). LncRNA5389 is defined as 294bp in length and is located between the two NEK genes (Tb927.10.5930 and Tb927.10.5940) but does not overlap within their UTRs. LncRNA5837 is 425bp in length and locates upstream of RBP7 family genes. Moreover, 93% of lncRNA5837 overlap with the 3'UTR of an upstream gene, Tb927.10.12080, which encodes for a protein with an unknown function. LncRNA6099 is 273bp in length and localize partially with the annotated 5'UTR of YAK gene.

EXPRESSION PROFILE OF THE FOUR CANDIDATE LNCRNAS DURING STUMPY FORMATION

Should, any of our four candidates lncRNA regulate slender to stumpy transition, we expected to see changes in their transcript levels during stumpy formation, as described for two essential genes for stumpy formation: RBP7 and PP1[12].

Expression levels of candidate lncRNAs and their nearby stumpy essential genes were analyzed by qPCR during stumpy formation. pCTP-cAMP was used to induce slender to stumpy transition and RNA was collected 6, 12, 24, 36 and 48 hours after addition of pCPT-cAMP and in a control culture of trypanosomes slender forms were maintained $<5x10^5$ cells/ml.

During *in vitro* stumpy formation, we observed two different expression profiles for the stumpy essential genes: (1) a decrease of transcript levels for PP1 and RBP7 and (2) no changes in their mRNA levels for NEK and YAK (Fig. 5). PP1, RBP7 and YAK transcript levels reproduce microarrays data in the literature[12].

In the same way, we observed two expression profiles for the four selected lncRNAs: (1) decrease of transcript levels for lncRNA5837 and lncRNA6099 and (2) no changes in their transcript levels for lncRNA5388 and lncRNA5389 (Fig. 5). Interestingly, lncRNA5837 and its nearby gene RBP7 show a similar expression pattern during stumpy formation. While increasing in the first 6 hours, their transcript levels decrease during the following hours of stumpy formation (Fig. 5).

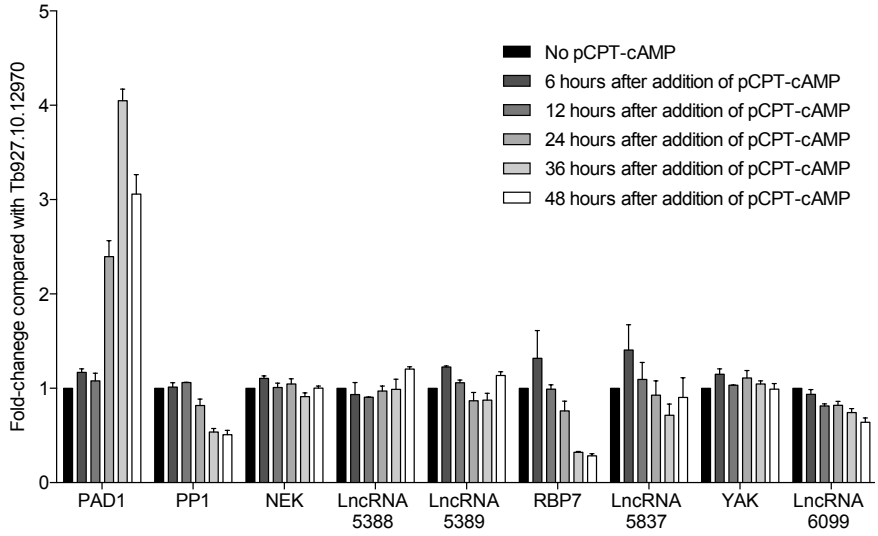


Figure 5: Transcript levels of candidate lncRNAs and nearby genes during stumpy formation. qPCR data was analyzed with delta-delta-Ct method, using control primers that target an endogenous gene (Tb927.10.12970), due to its stability during life cycle differentiation[46]. All the samples were normalized with GP1 sample (slender cells not exposed to pCPT-cAMP). Bars represents mean ± s.e.m.; n=3.

Although the expression of the candidate lncRNAs do not show significant changes during *in vitro* stumpy formation, it does not reflect a lack of function of the lncRNAs in stumpy formation. Indeed, lncRNA function can be controlled by their cellular localization, secondary structure or protein partners[77–81].

PHENOTYPIC SCREEN TO STUDY LNCRNA FUNCTION IN SLENDER TO STUMPY TRANSITION

We performed a phenotypic screen for stumpy formation to study the role of the four selected lncRNAs, using a gain- or loss-of-function approach with our reporter cell-line, GP1. Internal controls were included by adding RBP7 and PP1 mutant cell lines to the phenotypic screen. Indeed, RBP7 overexpression and PP1 knocked-down lead respectively to an increase and a decrease of stumpy formation[44], which provide positive controls for our phenotypic screen.

RNA interference or overexpression was induced or not with tetracycline and, after one day, stumpy formation was triggered by adding pCPT-cAMP in the culture medium. Two days after addition of pCPT-cAMP, GFP expression was quantified using flow cytometry.

As expected the overexpression of RBP7B (RBP7 #1) induces a small increase in the number of GFP expressing cells, confirming that RBP7 overexpression promotes stumpy formation (Fig. 6A). As well, knock-down of PP1 (PP1 #3) leads to a decrease about 50% of GFP expressing cells, showing that PP1 knock-down prevented slender to stumpy transition (Fig. 6B).

Although lncRNA mutant cells showed changes in the number of GFP expressing cells, our results suffered from variability, lack of reproducibility and consistency from the four different phenotypic screens carried on, making difficult any interpretations and conclusions. Different growth rates, leading to variable number of cells, between our lncRNA mutants cells might explain the variability in our phenotypic screen results. Indeed, recent data from Luisa laboratory (data not shown) suggest that the number of cells influences stumpy formation. Moreover, overexpression of RBP7 protein only led to 10% increase in GFP expressing cells comparing to non-induced conditions (Fig. 6; Supplementary Fig. 4), which undervalued previously published data[44]. In fact, we observed that our stumpy formation assay reached a maximal number of GFP expressing cells after 48h of pCPT-cAMP exposure (Fig. 3), which makes difficult to observe any increase in stumpy formation at this time point.

Although this phenotypic screen has not revealed a clear role of the four selected lncRNAs for stumpy formation, we gained more information in the experimental set-up as an early time point to analyze GFP signal will be more accurate to quantify stumpy formation.

Moreover, we decided to pursue our study and proceed for further experiments only with the lncRNA5837 as this lncRNA showed interesting features:

(1) It is located upstream to essential genes for stumpy formation, the RBP7 family genes

(2) Its expression profile during stumpy formation resembles the one of RBP7

(3) Its overexpression induces a small increase of stumpy formation as RBP7 overexpression

(4) It is the bigger lncRNA in length from the candidate lncRNAs that enable the design of probes for RNA FISH experiment.

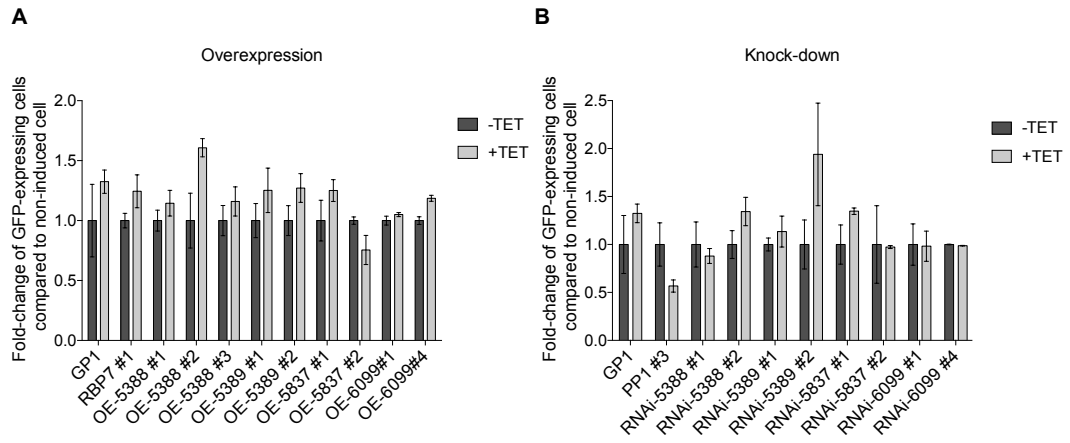


Figure 6: Phenotypic screen for stumpy formation using gain- and loss-of-function approach. Cells that overexpress RBP7 or candidate lncRNAs **(A)** and cells that knockdown PP1 or candidate lncRNAs **(B)** were

cultivated in medium with pCPT-cAMP and induced (+TET) or not (-TET) with tetracycline. Two days after, GFP expressing cells were quantified by flow cytometry and normalized to non-induced condition. (n=4, mean ± s.e.m.). OE – overexpression; RNAi – knock-down.

ROLE OF LNCRNA5837 IN STUMPY FORMATION

We decided to focus our study on the lncRNA5837 and to characterize in more detail the phenotypes associated with gain- or loss-of-function of lncRNA5837. Should, lncRNA5837 regulates slender to stumpy transition through *in cis* mechanism that controls RBP7 gene expression, we expected to see changes in RBP7 transcripts levels and/or in cell growth of mutant cells that overexpress or deplete lncRNA5837.

We had selected two independent clones of each overexpression and knock-down lncRNA5837 mutant cells and we included the positive controls, PP1 knock-down and RBP7 overexpression mutant cells to the study. First, we wanted to confirm that the mutant cells overexpress or knock-down the target gene. The mutant cells were induced or not with tetracycline during 24 hours, RNA was extracted and transcript levels quantified by qPCR (Supplementary Fig. 5).

Overexpression and knock-down mutant cells, induced with tetracycline, showed significant changes in transcript level of their target genes (Supplementary Fig. 5). While PP1 and lncRNA5837 knocked-down cells showed respectively 60% and 50% depletion, RBP7 and lncRNA5837 overexpression cells showed respectively 37 times and 8 to 9 times more expression of the target gene (Supplementary Fig. 5).

PP1 knock-down and RBP7 overexpression mutants affect the slender to stumpy transition and interestingly we showed that their cell growth relate to their phenotype in stumpy formation (Fig. 7). Overexpression of RBP7 leads to a decrease in the cell growth rate (Fig. 7), which relates its phenotype for promoting stumpy formation. Oppositely, PP1 knock-down leads to increase in the cell growth rate when compared to non-induced mutants cells (Fig. 7), which relates its phenotype for preventing stumpy formation and by consequent promoting proliferative slender forms. Similarly, overexpression of lncRNA5837 leads to a decrease in the cell growth rate while its knock-down increase the growth of cell when compared to the non-induced cells (Fig. 7). Moreover, overexpression of lncRNA5837 (clone #1) showed a leaky expression of lncRNA5837 in absence of tetracycline (Supplementary Fig. 5), which correlates the decrease of cell growth rate in the non-induced cells (Fig. 7). However, cell growth phenotype observed with lncRNA5837 mutant cells was not due to regulation of RBP7 or Tb927.10.12080 gene expression through *in cis* mechanism, since the overexpression or knock-down of lncRNA5837 does not affect expression of its nearby genes (Supplementary Fig. 5).

Therefore, lncRNA5837 mutant cells showed cell growth phenotype that could be due to an increase or decrease of stumpy formation, as it is observed in PP1 and RBP7 mutant cells.

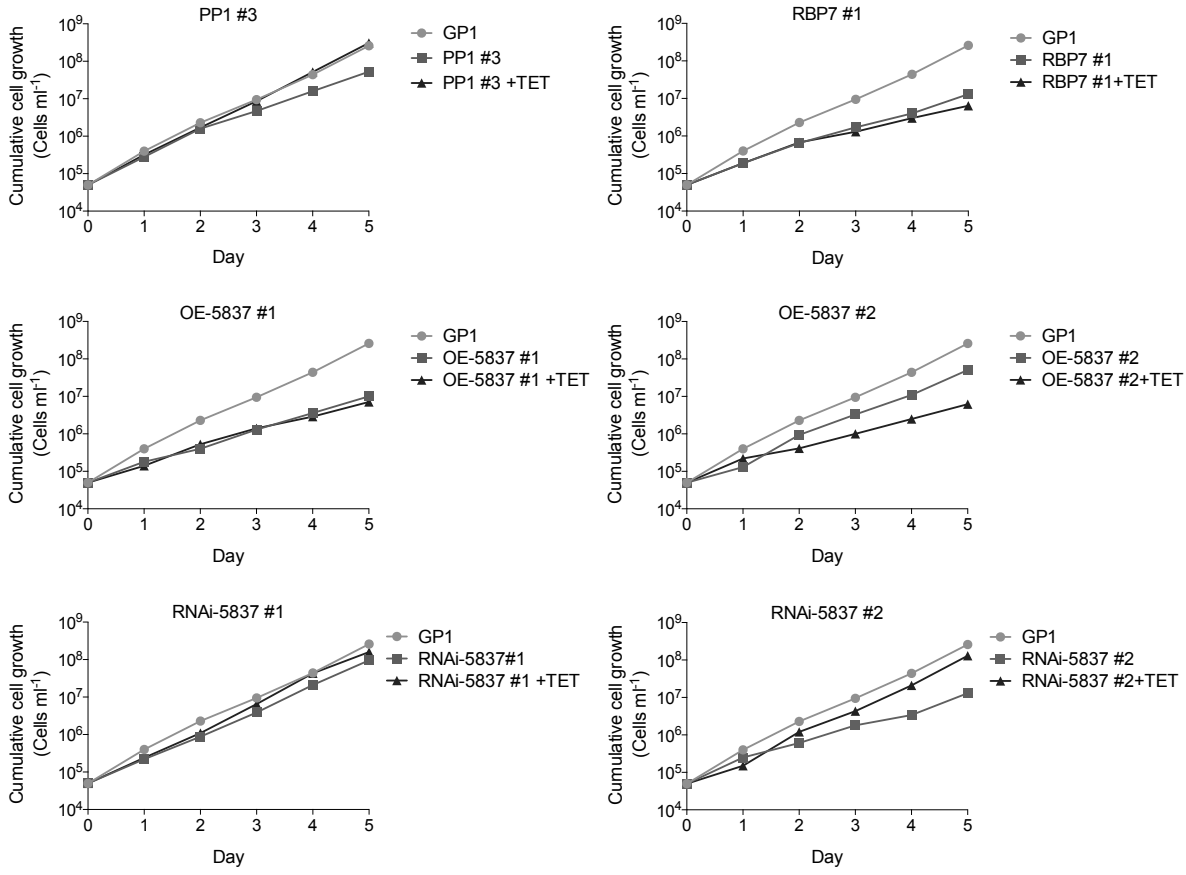


Figure 7: Overexpression and depletion of lncRNA5837 affect cell growth. *In vitro* growth of pleomorphic RNAi and overexpression cell lines targeting lncRNA5837, PP1 and RBP7. Cells were induced or not (1µg ml^{-1} tetracycline) and kept <5x10^5 cells ml^{-1}.

GAIN- AND LOSS-OF-FUNCTION OF LNCRNA5837

To study the function of lncRNA5837 in stumpy formation, the lncRNA was overexpressed and knocked-down in independent pleomorphic GP1 cell lines. On day 0, RNA interference or overexpression was induced or not with tetracycline, on day 1, stumpy formation was induced by adding pCPT-cAMP to the culture medium and, on days 2, 2.5 and 3, GFP expression was quantified using flow cytometry.

Control cell lines behaved as expected, PP1 knockdown resulted in a lower number of GFP expressing cells, indicating that PP1 depleted cells are less efficient at differentiating to stumpy form (Fig. 8A), while RBP7B overexpression promoted stumpy formation (Fig. 8C). Overexpression of lncRNA5837 (clone #1) promoted stumpy formation, mainly detected 36 hours after addition of pCPT-cAMP. However, a different clone of lncRNA5837 overexpression (clone #2) did not show the same phenotype as clone #1 (Fig. 8D), but

showed a decrease in growth rate (Fig. 7). This may be explained by overexpression of a non-functional product that only affects cell viability and growth. RNAi of the lncRNA5837 (clone #2) showed a reduced number of GFP expressing cells compared with parental cell line, suggesting that this mutant was less able to differentiate to stumpy forms. Surprisingly, a second clone (clone #1) revealed the opposing phenotype when we compared induced with non-induced conditions (Fig. 8B). The variability between RNAi clones indicate that more clones should be selected to confirm the phenotype.

Efficiency of RNAi was confirmed by performing qPCR and 3 days post-induction of RNAi, lncRNA5837 was depleted only by 50%, which could explain the small decrease observed in stumpy formation. Moreover, the non-induced RNAi clones show a leaky RNAi expression (Supplementary Fig. 6), which might explain the similar phenotype observed in induced and non-induced conditions.

Similarly, efficiency of overexpression was measured by qPCR. We observed that lncRNA5837 overexpression clones show a significant increase of lncRNA5837 after three days of induction with tetracycline and importantly do not show leaky expression (Supplementary Fig. 6).

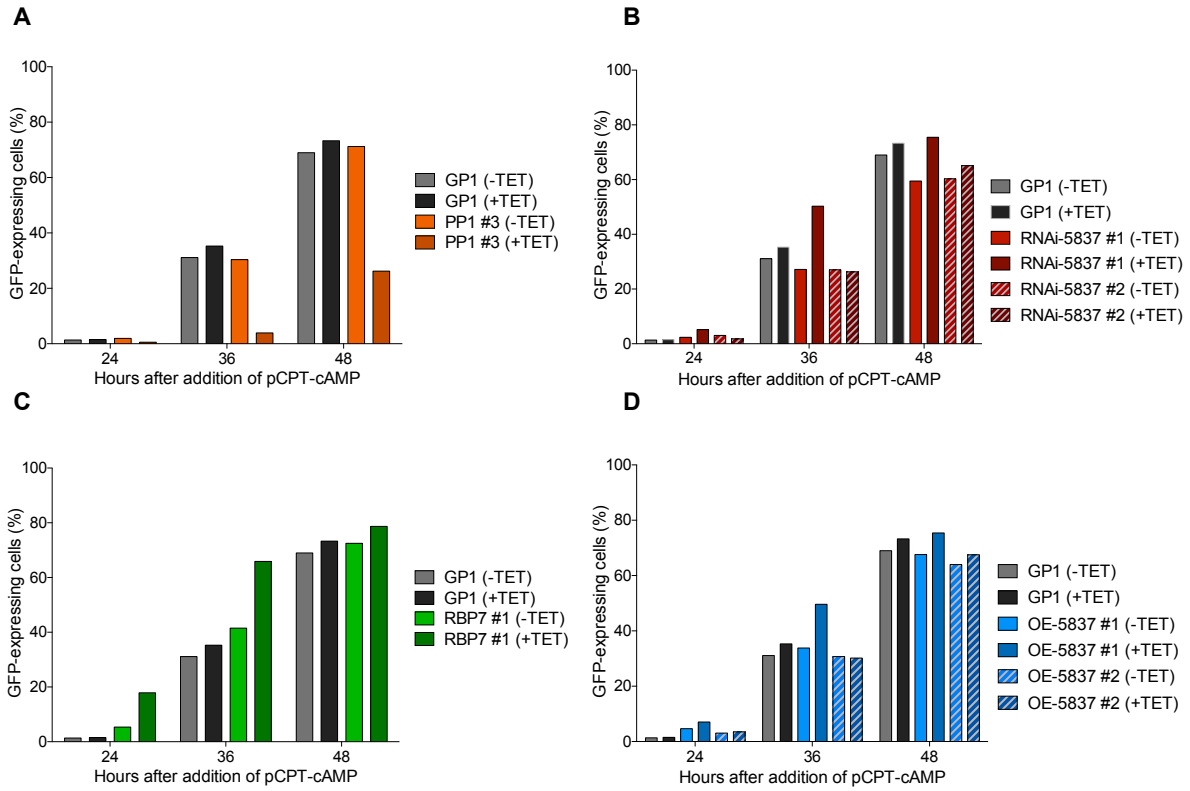


Figure 8: Role of lncRNA5837 in *in vitro* stumpy formation. Cells that knockdown **(A)** PP1 and **(B)** lncRNA5837, or overexpress **(C)** RBP7 and **(D)** lncRNA5837 were exposed to pCPT-cAMP and RNAi or overexpression being induced (1µg ml^{-1} tetracycline) or not (n=1). GFP expressing cell were quantified by flow cytometry 24, 36 and 48 hours after addition of pCPT-cAMP.

The increase of stumpy formation observed by overexpression of lncRNA5837 was only detectable in one of the two clones. This could be due to incorrect product of lncRNA overexpression. The overexpression plasmid (pDex-577) has an aldolase 3'UTR downstream of the transgene that stabilizes the transcript in bloodstream trypanosomes, which results in high mRNA of target transgene. However, aldolase 3'UTR is larger than lncRNA5837, which could affect lncRNA secondary structure, stability and function. Indeed, most lncRNAs showed lower transcript level and stability than protein-coding genes[66]. Next, the plasmid pDex-577 was modified to remove aldolase 3'UTR and lncRNA5837 was cloned into modified plasmid and used to transfect GP1 cell line. The new cell line, OE-5837 (clone #3), show a significant overexpression of lncRNA5837 and a decrease in growth rate compared to non-induced cells (Fig. 9A, B). Moreover, lncRNA5837 overexpression promotes stumpy formation (Fig. 9C) and more robust phenotype than previous overexpression clones. This suggests that lncRNA5837 is important for stumpy formation and that its function was being affected by aldolase 3'UTR.

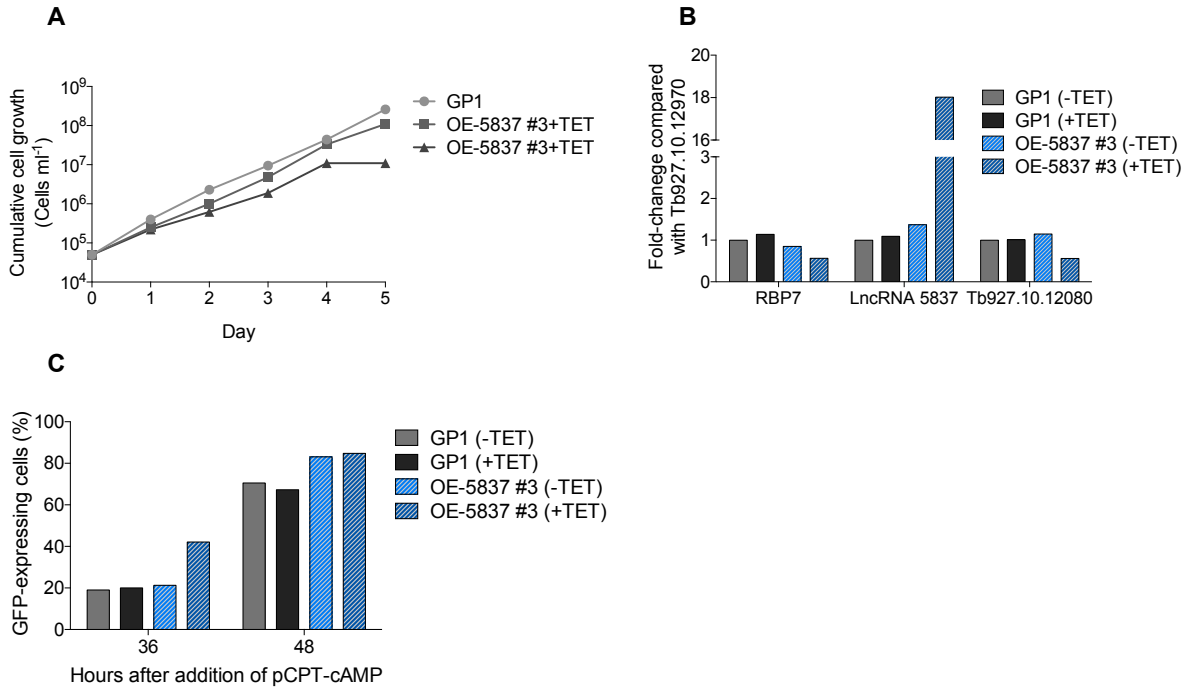


Figure 9: Characterization of the lncRNA5837 overexpression mutant using pDex-577 without Aldolase 3'UTR. (A) Growth curve of lncRNA5837 overexpression mutant, OE-5837 (clone #3). **(B)** Cells were induced (1µg ml^{-1} tetracycline) during 24 hours. RBP7, lncRNA5837 and Tb927.10.12080 were quantified by qPCR. Cells were normalized with endogenous gene Tb927.10.12970 and to non-induced parental cell line (GP1). **(C)** Percentage of GFP positive cells during *in vitro* stumpy formation with pCPT-cAMP (n=1).

Inside the mammalian host, slender to stumpy transition occurs at high cell density via quorum sensing pathway[41]. RBP7 overexpression alone is sufficient to promote premature cell cycle arrest in the absence of any stumpy-inducing exogenous stimuli[44]. Based on our previous result, we want to test if lncRNA5837 overexpression was enough to promote

stumpy formation without external stimuli. LncRNA5837 overexpression mutants were induced or not with tetracycline and maintained at a low concentration, <5x10^5 cells ml^{-1}. After five days, GFP expression was quantified using flow cytometry.

RBP7 and lncRNA5837 overexpression show an increase of GFP expressing cells compared to non-induced condition (Fig. 10). Therefore, these mutants are sufficient to promote stumpy formation without cell density or chemical stimulus. Moreover, this result confirms that growth defect observed in lncRNA5837 overexpression mutant (Fig. 9A) is due to stumpy differentiation.

Although phenotype caused by lncRNA5837 depletion was not as robust as PP1 depletion, our results of lncRNA5837 overexpression support the hypothesis that lncRNA5837 has a role as positive regulator of stumpy formation.

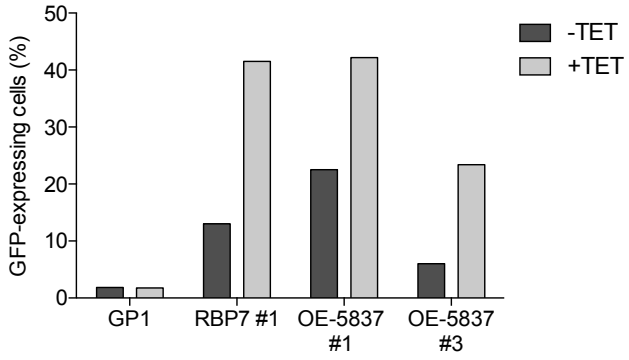


Figure 10: Overexpression of lncRNA5837 promotes stumpy formation without chemical or density stimulus. Percentage of GFP positive cells 5 days in non-induced or induced conditions (1µg ml^{-1} tetracycline). Cells were kept <5x10^5 cell ml^{-1} (n=1). Overexpression of lncRNA5837 using pDex-577 (OE-5837 #1) or modified pDex-577 without aldolase 3'UTR (OE-5837 #3).

LNCRNA5837 SUBCELLULAR LOCALIZATION

Recent studies have showed that human lncRNAs have stronger nuclear localization than mRNA[77,82]. Indeed, previous reports showed the involvement of lncRNA in nuclear functions such as epigenetic regulation of gene expression, alternative splicing and transcription initiation[83]. Therefore, subcellular localization of *T. brucei* lncRNA5837 will provide more information on its cellular function. We performed RNA FISH to determine the subcellular localization of lncRNA5837 and compare for changes its localization in slender and stumpy forms of *T. brucei,* using GP1 reporter cell line.

LncRNA5837 localizes mainly in the nuclear compartment of slender forms in which we observed an intense signal inside the nucleolus and an additional less intense signal in the nucleoplasm (Fig. 11). However, multiple foci can be also observed inside the nucleoplasm of slender forms (Supplementary Fig. 7). Interestingly, in stumpy forms the nucleoplasm

signals disappear completely and a more intense and unique signal is observed inside the nucleolus (Fig. 11;Supplementary Fig. 7). RNA FISH signals inside the nucleoplasm and in the nucleolus disappear when cells are treated with RNase A, validating that the RNA FISH signals are RNA specific (Fig. 11).

Interestingly, the accumulation of lncRNA5837 inside the nucleolus is observed during slender to stumpy transition as well as in lncRNA5837 overexpression mutants, two situations in which stumpy formation is promoted, suggesting that lncRNA5837 play a role in promoting stumpy formation through its function inside the nucleolus of *T. brucei*.

However, knock-down of lncRNA5837 results only in the partial disappearance of RNA FISH signals. Indeed, while the nucleoplasm signals disappeared completely, the nucleolus signal remained intact (Fig. 11). This result is in agreement with the qPCR results (Supplementary Fig. 5) in which only 50% depletion have been observed in lncRNA5837 knock-down mutants. Therefore, the remaining signal of lncRNA5837 inside the nucleolus could explain the small and inconsistent phenotype for stumpy formation observed in lncRNA5837 knock-down mutants.

LncRNA5837 localizes mainly in the nuclear compartment of *T. brucei* and it is relocated and restricted inside the nucleolus during the slender to stumpy transition. Altogether, these results suggest that lncRNA5837 promotes stumpy formation through its function inside the nucleolus of *T. brucei*.

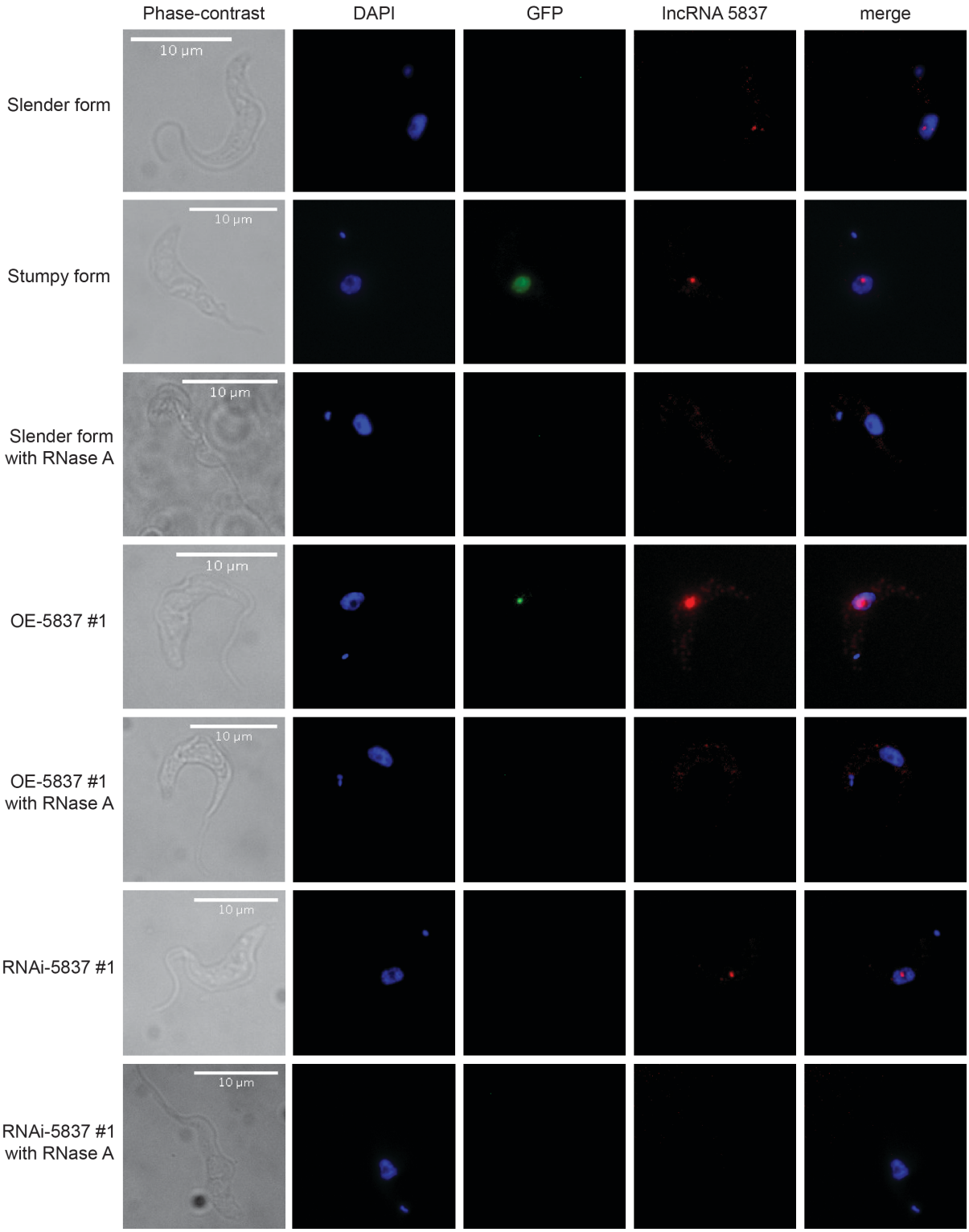


Figure 11: LncRNA5837 subcellular localization. RNA FISH was used to study cellular localization of lncRNA5837 in slender, stumpy forms and slender forms overexpressing or knocking-down lncRNA5837. Stumpy forms were collected from *in vitro* culture at medium cell density, in which a low percentage of cells were stumpy forms. LncRNA5837 overexpression (OE-5837 #1) and depletion (RNAi-5837 #1) mutants were induced with 1µg ml^{-1} tetracycline. Slender forms and lncRNA5837 mutants were treated or not with RNase A before the hybridization of the RNA FISH probes.

CONCLUSION AND FUTURE WORK

From yeast to humans, lncRNAs have been identified and shown to function in a large variety of biological processes, including gene expression, post-transcriptional regulation[83], nuclear organization[84] and cell differentiation[67]. In this project we aimed to identify lncRNAs that play a role in stumpy formation in *T. brucei*.

First, we created a *T. brucei* pleomorphic reporter cell line that expresses GFP specifically in stumpy forms. We validated the use of this cell line as an efficient tool to study and quantify stumpy formation both *in vitro* and *in vivo*.

Second, using a bioinformatic approach, we selected four lncRNAs as candidate regulators of stumpy formation, based on their genomic proximity to genes essential to stumpy formation that could act via a *cis* regulatory mechanism. In the future, additional approaches could be used to select lncRNAs important to the stumpy formation, in order to include trans-acting lncRNAs. To that end, the genome-wide RNAi screen[44] used to identify protein-coding genes essential for stumpy formation could be applied to identify essential lncRNAs for the slender to stumpy transition.

From our initial screen of selected lncRNAs, we characterized in more detail lncRNA5837. The most important observation from this thesis is that overexpression of lncRNA5837 increases stumpy formation and, conversely, its knockdown inhibits the slender to stumpy transition under *in vitro* conditions. Interestingly, even in cultures without any exogenous stimulus of stumpy differentiation, overexpression of lncRNA5837 is sufficient to promote differentiation to stumpy forms, mimicking the RBP7 overexpression phenotype[44]. Therefore, lncRNA5837 appears to be a positive regulator of the quorum sensing signaling pathway.

Although the depletion of lncRNA5837 promotes a decrease in stumpy formation, this phenotype could not be reproduced with a second lncRNA5837 knock-down clone, which could be explained by an insufficient depletion method causing an inconsistent phenotype. Therefore, in the future it would be necessary to explore other approaches that provide a more efficient depletion of lncRNA5837, such as small hairpin RNA (shRNA), antisense LNA Gapmers or knockout of the lncRNA5837 from the genome.

Additionally, it would be important to confirm the effect *in vivo* of both lncRNA5837 mutant parasites on stumpy formation. Using a rodent model, we could quantify the number of GFP expressing parasites during *T. brucei* infection, comparing lncRNA5837 mutant and wild-type parasites.

Using RNA FISH we observed that the endogenous lncRNA5837 localizes mainly in the nucleus of *T. brucei*. Interestingly, subcellular localization of lncRNA5837 changes during slender to stumpy transition. While in slender forms, lncRNA5837 resides in the nucleoplasm and in the nucleolus, in stumpy forms, it relocalizes entirely to the nucleolus. Although these

observations need to be further validated and especially quantified, they are consistent with a regulatory function of lncRNA5837 in gene expression or nuclear organization of *T. brucei*. Besides, they suggest that lncRNA5837 is functionally associated with the nucleolus and with the gene reprogramming associated with slender to stumpy differentiation.

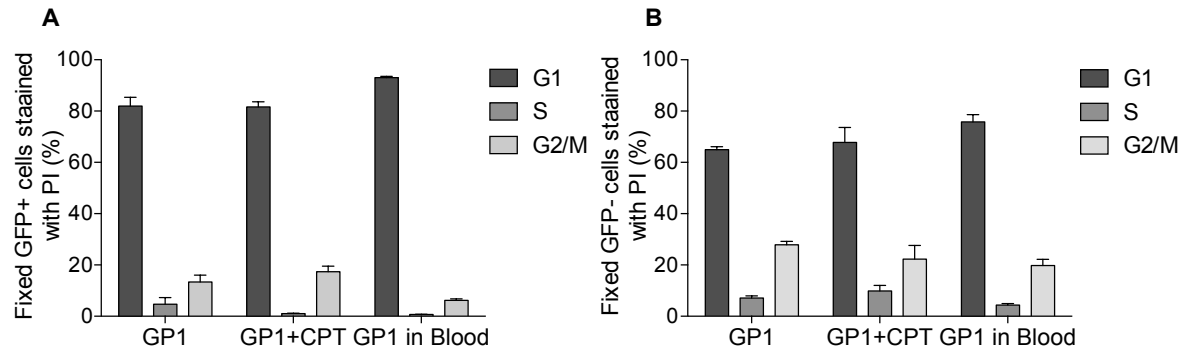
On the other hand, overexpression or depletion of lncRNA5837 does not affect the expression of its nearby gene, RBP7. As a consequence, lncRNA5837 does not control stumpy formation via a *cis*-acting regulatory mechanism of *RBP7* RNA expression. However, additional experiments will be required to exclude a regulatory mechanism of lncRNA5837 on RBP7 expression. Indeed, it would be important to study RBP7 protein levels in mutant parasites that overexpress or deplete lncRNA5837 using *T. brucei* cell line in which RBP7 protein is tagged.

To understand the molecular mechanism behind the role of lncRNA5837 in stumpy formation, it would be important to identify the protein partners and/or DNA binding sites of lncRNA5837. To that end, using biotinylated DNA probes, we could perform RNA pull-down assays recently adapted for lncRNA such as CHIRP (chromatin isolation by RNA purification) or RAP (RNA antisense purification) techniques.

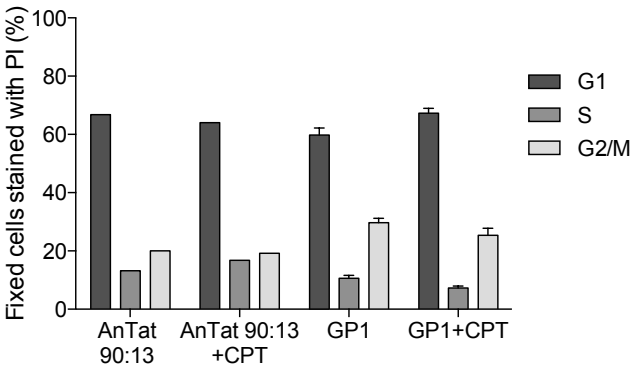
Nucleolar proteins, including Pol I, could be potential protein partners of lncRNA5837 and as such be identified in the RNA pull-down assays. Interestingly, Pol I shows a similar subcellular localization than lncRNA5837 in slender forms (Fig. 10), residing in two discrete regions of the nucleus: in the nucleolus and in the nucleoplasm[85]. Additionally, Pol I and lncRNA5837 interaction could be confirmed using immunofluorescence with Pol-I antibody coupled with RNA FISH to detect lncRNA5837.

In summary, in this project, we provide the first evidence for the function of a lncRNA in human deadly parasite, *T. brucei*. Interestingly, this lncRNA acts as positive regulator to promote *T. brucei* life cycle progression from proliferative slender forms to quiescent stumpy forms. Therefore, our study reveals that lncRNA, as in other eukaryotic cells, plays crucial role in biology of *T. brucei*, by controlling the parasite differentiation. Even though more studies are needed to demonstrate that lncRNA-5837 is a key inducer of stumpy differentiation, this work collected the first pieces of evidence of such mechanism.

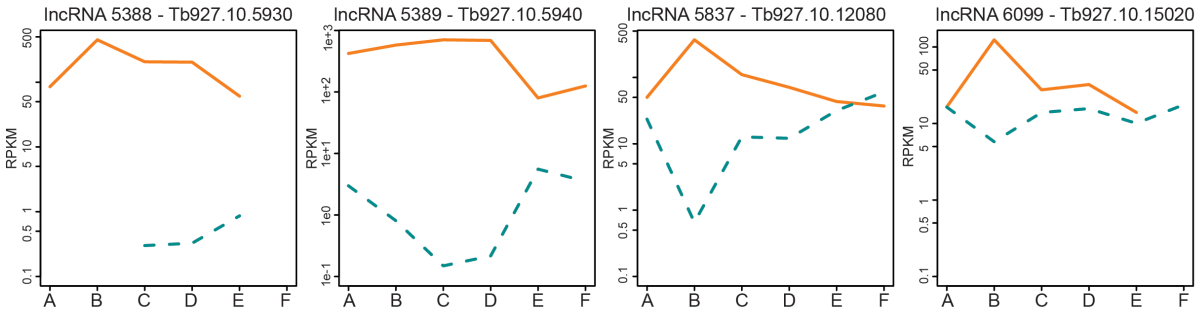
ANNEXES



Supplementary figure 1: Cell cycle profile of cells differentiated *in vitro* or collected from blood at day 6 post-infection. (A) Cell cycle profile of fixed GFP-positive cells stained with PI. **(B)** Cell cycle profile of fixed GFP-negative cells stained with PI.

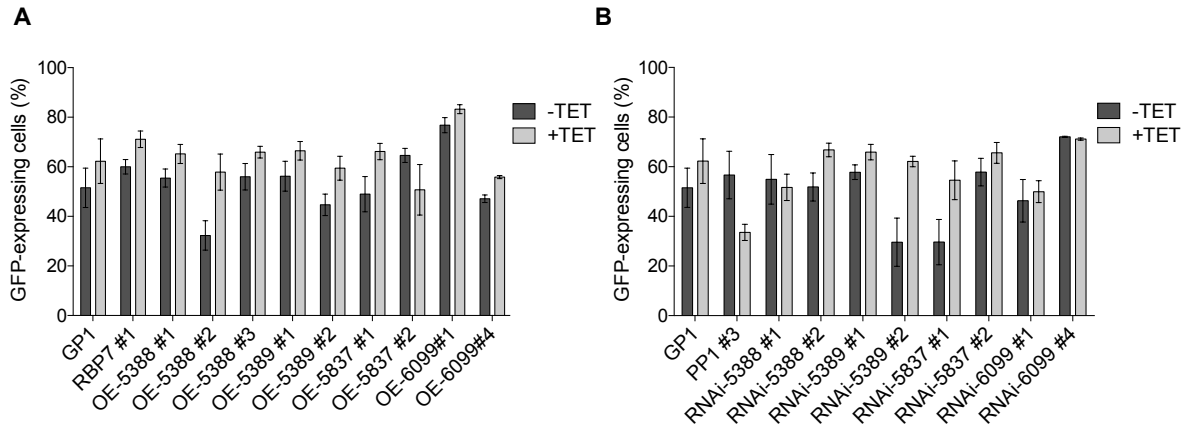


Supplementary figure 2: Cell cycle profile one day after addition of pCPT-cAMP. Cell cycle profile of AnTat 90:13 and GP1 cell line cultivated one day in HMI-11 medium (AnTat 90:13 or GP1) or in HMI-11 medium with pCPT-cAMP (CPT).

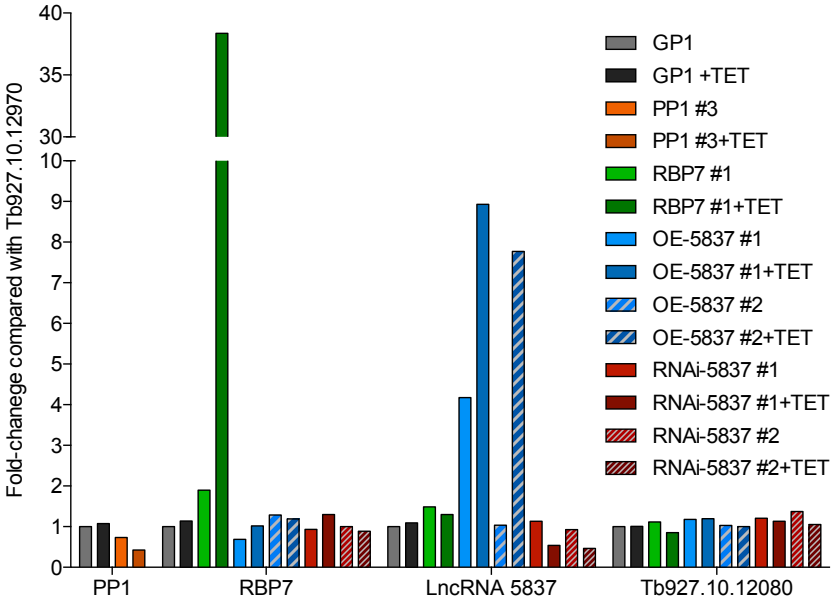


Supplementary figure 3: Correlation plot between candidate lncRNAs and nearby genes. Orange lines correspond to lncRNAs and dashed blue lines are protein-coding genes. A and B are respectively bloodstream forms and procyclic forms RNA sequenced samples. D to F are samples collected from infected mice with *T. brucei*. Pearson correlation r > 075 or r < -0-75 for positive correlation or negative correlation, respectively. LncRNA5388 – Tb927.10.5930 (Pearson correlation r = -0.43). LncRNA5389 – Tb927.10.5940 (Pearson

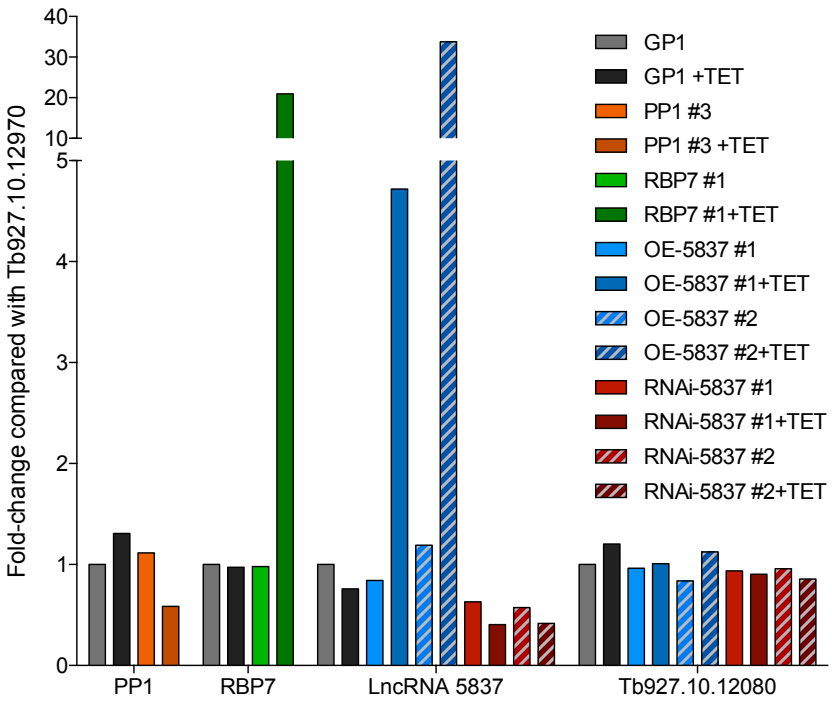
correlation r = -0.96). LncRNA5837 – Tb927.10.12020 (Pearson correlation r = -0.65). LncRNA 6099 – Tb927.10.15020 (Pearson correlation r = -0.82).



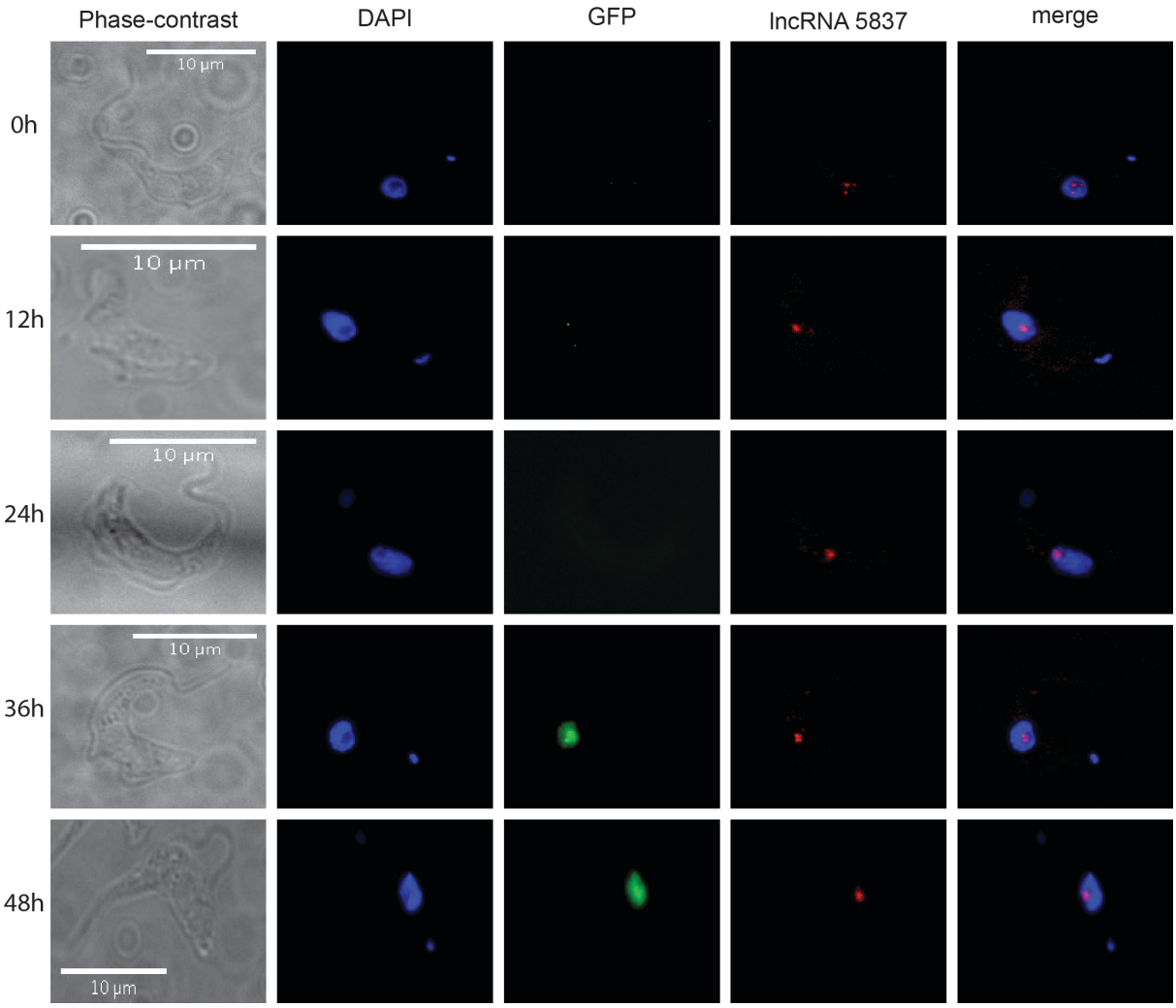
Supplementary figure 4: Phenotypic screen for stumpy formation using gain- and loss-of-function approach. Cells that overexpress RBP7 or candidate lncRNAs **(A)** and cells that knockdown PP1 or candidate lncRNAs **(B)** were cultivated in medium with pCPT-cAMP and induced (+TET) or not (-TET) with tetracycline. Two days after, GFP expressing cells were quantified by flow cytometry. (n=4, mean ± s.e.m.). OE – overexpression; RNAi – knock-down.



Supplementary figure 5: Transcript levels of PP1, RB7 and lncRNA5837. Cells were induced (1µg ml^{-1} tetracycline) during 24 hours. PP1, RBP7, lncRNA5837 and Tb927.10.12080 were quantified by qPCR (n=1). Cells were normalized with endogenous gene Tb927.10.12970 and to non-induced parental cell line (GP1).



Supplementary figure 6: Transcript levels of PP1, RBP7 and lncRNA5837. Cells were induced (1µg ml^{-1} tetracycline) during three days. PP1, RBP7, lncRNA5837 and Tb927.10.12080 were quantified by qPCR. Cells were normalized with endogenous gene Tb927.10.12970 and to non-induced parental cell line (GP1).



Supplementary figure 7: LncRNA5837 subcellular localization during *in vitro* stumpy formation. GP1 cell line was exposed to pCPT-cAMP and cells were collected after 12, 24, 36 and 48 hours. Stumpy forms express GFP in the nucleus. Percentage of GFP expressing cells were quantified by flow cytometry: 1,29%, 13,2%, 68,4% and 90,9% of GFP positive cells 12, 24, 36 and 48 hours after addition of pCPT-cAMP, respectively.

REFERENCES

1. MacGregor, P., Szöőr, B., Savill, N. J. & Matthews, K. R. Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nat. Rev. Microbiol.* **10,** 431–8 (2012).

2. Kennedy, P. G. E. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurol.* **12,** 186–194 (2013).

3. Pays, E. Regulation of antigen gene expression in Trypanosoma brucei. *Trends Parasitol.* **21,** 517–520 (2005).

4. Vickerman, K. Developmental cycles and biology of pathogenic trypanosomes. *Br. Med. Bull.* **41,** 105–114 (1985).

5. Langousis, G. & Hill, K. L. Motility and more: the flagellum of Trypanosoma brucei. *Nat. Rev. Microbiol.* **12,** 505–18 (2014).

6. Tschudi, C. & Ullu, E. Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. *EMBO J.* **7,** 455–463 (1988).

7. Ullu, E., Matthews, K. R. & Tschudi, C. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol. Cell. Biol.* **13,** 720–725 (1993).

8. Matthews, K. R., Tschadi, C. & Ullu, E. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* **8,** 491–501 (1994).

9. Günzl, A. *et al.* RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei. *Eukaryot. Cell* **2,** 542 – 551 (2003).

10. Clayton, C. E. Life without transcriptional control? From fly to man and back again. *EMBO J.* **21,** 1881–1888 (2002).

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

12. Kabani, S. *et al.* Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. *BMC Genomics* **10,** 427 (2009).

13. Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J. D. & Clayton, C. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics* **10,** 495 (2009).

14. Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. & Cross, G. a M. Genome-wide analysis of

mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* **38,** 4946–4957 (2010).

15. Kolev, N. G. *et al.* The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. *PLoS Pathog.* **6,** 1–15 (2010).

16. Nilsson, D. *et al.* Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS Pathog.* **6,** e1001037 (2010).

17. Gunasekera, K., Wüthrich, D., Braga-Lagache, S., Heller, M. & Ochsenreiter, T. Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. *BMC Genomics* **13,** 556 (2012).

18. Urbaniak, M. D., Martin, D. M. a & Ferguson, M. a J. Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. *J. Proteome Res.* **12,** 2233–2244 (2013).

19. Li, C.-H. *et al.* Roles of a Trypanosoma brucei 5'->3' exoribonuclease homolog in mRNA degradation. *RNA* **12,** 2171–2186 (2006).

20. Schwede, A. *et al.* The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Res.* **37,** 5511–5528 (2009).

21. Estévez, A. M. The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res.* **36,** 4573–4586 (2008).

22. Paterou, A., Walrad, P., Craddy, P., Fenn, K. & Matthews, K. Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. *J. Biol. Chem.* **281,** 39002–39013 (2006).

23. Walrad, P., Paterou, A., Acosta-Serrano, A. & Matthews, K. R. Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathog.* **5,** 1–10 (2009).

24. Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S. & Matthews, K. 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.* **34,** 5312–5324 (2006).

25. Horn, D. Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics* **9,** 2 (2008).

26. Turner, C. M., Aslam, N. & Dye, C. Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. *Parasitology* **111,** 289–300 (1995).

27. Amiguet-Vercher, A. *et al.* Loss of the mono-allelic control of the VSG expression sites during the development of Trypanosoma brucei in the bloodstream. *Mol. Microbiol.* **51,** 1577–1588 (2004).

28. Pays, E. *et al.* Abrupt RNA changes precede the first cell division during the differentiation of Trypanosoma brucei bloodstream forms into procyclic forms in vitro. *Mol. Biochem. Parasitol.* **61,** 107–114 (1993).

29. Turner, C. M. The use of experimental artefacts in African trypanosome research. *Parasitol. Today* **6,** 14–17 (1990).

30. MacGregor, P., Savill, N. J., Hall, D. & Matthews, K. R. Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host Microbe* **9,** 310–318 (2011).

31. Vickerman, K., Tetley, L., Hendry, K. a & Turner, C. M. Biology of African trypanosomes in the tsetse fly. *Biol. Cell* **64,** 109–119 (1988).

32. Hannaert, V., Bringaud, F., Opperdoes, F. & Michels, P. Kinetoplastid Biology and Disease. *Kinetoplastid Biol. Dis.* **2,** 1–30 (2003).

33. Priest, J. W. & Hajduk, S. L. Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. *J. Bioenerg. Biomembr.* **26,** 179–191 (1994).

34. Vickerman, K. Polymorphism and Mitochondrial Activity in Sleeping Sickness Trypanosomes. *Nature* **208,** 762–766 (1965).

35. Engstler, M. *et al.* Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes. *Cell* **131,** 505–515 (2007).

36. Nolan, D. P., Rolin, S., Rodriguez, J. R., Van Den Abbeele, J. & Pays, E. Slender and stumpy bloodstream forms of Trypanosoma brucei display a differential response to extracellular acidic and proteolytic stress. *Eur. J. Biochem.* **267,** 18–27 (2000).

37. Matthews, K. R. & Gull, K. Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J. Cell Biol.* **125,** 1147–1156 (1994).

38. Dean, S., Marchetti, R., Kirk, K. & Matthews, K. R. A surface transporter family conveys the trypanosome differentiation signal. *Nature* **459,** 213–217 (2009).

39. Engstler, M. & Boshart, M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. *Genes Dev.* **18,** 2798–811 (2004).

40. Seed, J. R. & Sechelski, J. B. Mechanism of long slender (LS) to short stumpy (SS) transformation in the African trypanosomes. *J. Protozool.* **36,** 572–577 (1989).

41. Vassella, E., Reuner, B., Yutzy, B. & Boshart, M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J. Cell Sci.* **110 (Pt 2,** 2661–2671 (1997).

42. Reuner, B., Vassella, E., Yutzy, B. & Boshart, M. Cell density triggers slender to stumpy

differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol. Biochem. Parasitol.* **90,** 269–280 (1997).

43. MacGregor, P. & Matthews, K. R. Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in Trypanosoma brucei. *Nucleic Acids Res.* **40,** 7705–17 (2012).

44. Mony, B. M. *et al.* Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature* **505,** 681–685 (2014).

45. Mony, B. M. & Matthews, K. R. Assembling the components of the quorum sensing pathway in African trypanosomes. *Mol. Microbiol.* **96,** 220–232 (2015).

46. Jones, N. G. *et al.* Regulators of Trypanosoma brucei Cell Cycle Progression and Differentiation Identified Using a Kinome-Wide RNAi Screen. *PLoS Pathog.* **10,** (2014).

47. Souza, G. M., Lu, S. & Kuspa, a. YakA, a protein kinase required for the transition from growth to development in Dictyostelium. *Development* **125,** 2291–2302 (1998).

48. Moniz, L., Dutt, P., Haider, N. & Stambolic, V. Nek family of kinases in cell cycle, checkpoint control and cancer. *Cell Div.* **6,** 18 (2011).

49. Fry, A. M., O'Regan, L., Sabir, S. R. & Bayliss, R. Cell cycle regulation by the NEK family of protein kinases. *J. Cell Sci.* **125,** 4423–4433 (2012).

50. Clayton, C. E. Networks of gene expression regulation in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **195,** 96–106 (2014).

51. Kolev, N. G., Ullu, E. & Tschudi, C. The emerging role of RNA-binding proteins in the life cycle of Trypanosoma brucei. *Cell. Microbiol.* **16,** 482–9 (2014).

52. Erben, E. D., Fadda, A., Lueong, S., Hoheisel, J. D. & Clayton, C. A genome-wide tethering screen reveals novel potential post-transcriptional regulators in Trypanosoma brucei. *PLoS Pathog.* **10,** e1004178 (2014).

53. Vassella, E. *et al.* Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of Trypanosoma brucei. *Mol. Microbiol.* **41,** 33–46 (2001).

54. Pfister, D. D. *et al.* A mitogen-activated protein kinase controls differentiation of bloodstream forms of Trypanosoma brucei. *Eukaryot. Cell* **5,** 1126–1135 (2006).

55. Barquilla, a. *et al.* Third target of rapamycin complex negatively regulates development of quiescence in Trypanosoma brucei. *Proc. Natl. Acad. Sci.* **109,** 14399–14404 (2012).

56. Tyler, K. M., Matthews, K. R. & Gull, K. The bloodstream differentiation-division of Trypanosoma brucei studied using mitochondrial markers. *Proc. Biol. Sci.* **264,** 1481–1490 (1997).

57. Breidbach, T., Ngazoa, E. & Steverding, D. Trypanosoma brucei: In vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. *Exp. Parasitol.* **101,** 223–230 (2002).

58. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol* **10,** 925–933 (2013).

59. Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **81,** 145–166 (2012).

60. Bonasio, R. & Shiekhattar, R. Regulation of Transcription by Long Noncoding RNAs. *Annu. Rev. Genet.* **48,** 433–455 (2014).

61. Faghihi, M. A. *et al.* Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* **11,** R56 (2010).

62. Moran, V. a., Perera, R. J. & Khalil, A. M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.* **40,** 6391–6400 (2012).

63. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: Cellular address codes in development and disease. *Cell* **152,** 1298–1307 (2013).

64. Mayer, C., Schmitz, K. M., Li, J., Grummt, I. & Santoro, R. Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. *Mol. Cell* **22,** 351–361 (2006).

65. Hu, W., Alvarezxs-Dominguez, J. R. & Lodish, H. F. Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep.* **13,** 24–26 (2012).

66. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22,** 885–898 (2012).

67. Savić, N. *et al.* lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs. *Cell Stem Cell* **15,** 720–34 (2014).

68. Djikeng, a, Shi, H., Tschudi, C. & Ullu, E. RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA* **7,** 1522–1530 (2001).

69. Tschudi, C., Shi, H., Franklin, J. B. & Ullu, E. Small interfering RNA-producing loci in the ancient parasitic eukaryote Trypanosoma brucei. *BMC Genomics* **13,** 427 (2012).

70. Patrick, K. L. *et al.* Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote Trypanosoma brucei. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 17933–17938 (2009).

71. Patrick, K. L. *et al.* Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient Trypanosoma brucei. *Mol. Microbiol.* **67,** 435–447 (2008).

72. MacGregor, P., Rojas, F., Dean, S. & Matthews, K. R. Stable transformation of pleomorphic bloodstream form Trypanosoma brucei. *Mol. Biochem. Parasitol.* **190,** 60–62 (2013).

73. Wickstead, B., Ersfeld, K. & Gull, K. Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of Trypanosoma brucei. *Mol. Biochem. Parasitol.* **125,** 211–216 (2002).

74. Kelly, S. *et al.* Functional genomics in Trypanosoma brucei: A collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.* **154,** 103–109 (2007).

75. Brecht, M. & Parsons, M. Changes in polysome profiles accompany trypanosome development. *Mol. Biochem. Parasitol.* **97,** 189–198 (1998).

76. Capewell, P. *et al.* Regulation of Trypanosoma brucei Total and Polysomal mRNA during Development within Its Mammalian Host. *PLoS One* **8,** e67069 (2013).

77. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16,** 1–16 (2015).

78. Johnsson, P., Lipovich, L., Grandér, D. & Morris, K. V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* **1840,** 1063–71 (2014).

79. Lu, Z. J. *et al.* Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res.* **21,** 276–285 (2011).

80. Zhu, J. J., Fu, H. J., Wu, Y. G. & Zheng, X. F. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci. China Life Sci.* **56,** 876–885 (2013).

81. Ferre, F., Colantoni, a. & Helmer-Citterich, M. Revealing protein-lncRNA interaction. *Brief. Bioinform.* 1–11 (2015). doi:10.1093/bib/bbv031

82. Ayupe, A. C. *et al.* Global analysis of biogenesis, stability and sub-cellular localization of lncRNAs mapping to intragenic regions of the human genome. *RNA Biol.* **6286,** 00–00 (2015).

83. Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14,** 699–712 (2013).

84. Quinodoz, S. & Guttman, M. Long noncoding RNAs : an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* **24,** 651–663 (2014).

85. Navarro, M. & Gull, K. A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature* **414,** 759–763 (2001).