THE UNIVERSITY *of* EDINBURGH

RNA Editing in Trypanosomes: a Tale of Two Ligases

Laura Jeacock

School of Biological Sciences

Thesis submitted for a degree of Doctor of Philosophy

University of Edinburgh

August 2013

Abstract

Uridylyl insertion/deletion mRNA editing is essential for mitochondrial gene expression in *Trypanosoma brucei* and governed by multi-protein complexes called editosomes. The final step in each cycle of this post-transcriptional process is that of re-ligating the edited mRNA fragments. The ~20S RNA editing core complex contains two RNA editing ligases, REL1 and REL2, located, respectively, in a deletion and an insertion subcomplex. While REL1 is clearly essential for RNA editing, REL2 knockdown by RNAi has not resulted in a detectable phenotype. To explain these findings, alternative scenarios have been suggested: (a) REL2 is not functional *in vivo*; (b) REL1 can function in both insertion and deletion editing, whereas REL2 can only function in insertion editing; (c) REL1 has an additional role in repairing erroneously cleaved mRNAs.

To further investigate respective functions of the two RELs this study used three complimentary approaches: (i) genetic complementation with chimeric ligase enzymes, (ii) deep sequencing of RNA editing intermediates after ligase inactivation, and (iii) evolutionary analysis.

In vivo expression of two chimeric ligases, providing a REL2 catalytic domain at REL1's position in the deletion subcomplex and a REL1 catalytic domain at REL2's position in the insertion subcomplex, did not rescue the growth defect caused by REL1 ablation. Although the results were not fully conclusive they suggest that it is the specific catalytic properties of REL1 rather than its position within the deletion subcomplex that makes it essential.

In order to identify *in vivo* substrates of REL1, specific editing intermediates

that accumulated after genetic knockdown of REL1 expression were captured by 5' linker and deep sequenced using Ion Torrent and Illumina technology. Analyses of such unligated editing intermediates with bespoke bioinformatics tools suggest that REL1 functions in deletion editing as expected, but also in the repair of miscleaved mRNAs, implying a novel role for this ligase. Neither role can be fulfilled by REL2, at least not with sufficient efficiency. Sequencing data also suggest that either REL1 is not involved in ligation of addition editing substrates, or that REL2 in this case can fully compensate for loss of REL1.

REL1, REL2 and KREPA3 sequences were subjected to analysis using MEGA5 and the HyPhy package available on the Datamonkey adaptive evolution server. Results indicated that all three editosome genes are under much stronger purifying than diversifying selective forces. In general this selection pressure to conserve protein sequence increased from KREPA3 to REL2 to REL1, suggesting a requirement to maintain catalytic function for both ligases.

Taken together, these experiments reveal a novel function for REL1 during RNA editing, providing a rationale for its essentiality. Deductively, the results also suggest REL2, which was previously thought to be non-essential, may still be required by the cell at its position in the addition subcomplex. Evolutionary analysis suggests that the RELs and KREPA3 are under the same evolutionary forces to maintain their respective functions in RNA editing.

Declaration

Three of the five cell lines used in experiments illustrated in Chapter 2 were generated by Matthew Spencer, who also constructed the initial growth curve.

The sequences obtained from high-throughput Ion Torrent and MiSeq sequencing were generated by Angie Fawkes at the Western General hospital and by the GenePool facility (University of Edinburgh), respectively. Reads were filtered using a pipeline devised by Al Ivens (CIIE), who also generated graphs used in Chapter 3.

The qRT-PCR data for sequencing was generated by Caroline Dewar.

I declare that all material presented in this thesis is my own work, unless otherwise stated, and has not been submitted for any other degree previously.

Laura Jeacock

Acknowledgements

There have been many people who have helped me to completion of this thesis. I want to thank my principle supervisor, Dr. Achim Schnaufer, who has provided guidance and constructive criticism along the way, and has pushed me to completion of this project. I also want to express my thanks to my secondary supervisor, Prof. David Tollervey, and to Dr. Sander Granneman who helped greatly in the initiation of the sequencing strategy and provided the 5' RNA linker used in these experiments.

I want to thank my fellow lab mates, past and present: Carol Hunja, Caroline Dewar, Dr. Matt Gould and Sinclair Cooper for their continuing support and guidance, as well as their company in the office and choices of afternoon YouTube videos. My thanks also go to the whole of the Matthews lab group, Robbie Carloni of the Interthal group, and Dr. Yvonne Adams of the Rowe group, for their project advice and helpfulness. I would also like to thank Drs. Pegine Walrad, Yvonne Adams, Paula MacGregor and Paul Hunt in particular, for their support when times got tough, and for invaluable advice. I would also like to extend my gratitude to Dr. Joanne Thomspon for being a supportive and organised post graduate committee member.

I would also like to say a big thank you to Thanasis Kousathanas who suggested that my project questions could also be answered by evolutionary analysis, and who kindly proof read Chapter 4 for me, and to Al Ivens who devised the RNA editing intermediate pipeline and help with this analysis.

My thanks go out to Alena Zíková, who kindly invited me over to learn how to pour and run glycerol gradients in České Budějovice and to her and the rest of her lab who

made me feel so welcome.

Outside of Ashworth there have been too many supportive people to mention personally including my close friends and partner, Jesse Harrison. I am truly fortunate to be surrounded by such exceptional people and I extend my gratitude to every one of them.

I would like to dedicate this thesis to my late father, Bob Jeacock, the only person outside this field of biology who I could imagine reading this, and the man who would have given me grief if I hadn't finished.

Abbreviations

ºC	degrees Celsius
%	per cent
α (as prefix)	anti
β	beta
μ	micro (10^{-6})
5' PO_4	5' monophosphate
3' OH	3' hydroxyl
A	adenosine
A6	adenosine triphosphate synthase subunit 6
AAT	animal African trypanosomiasis
AGV	antigenic variation
Ak (as prefix)	akinetoplastic
AMP	adenosine monophosphate
ApoL1	apolipoprotein L1
ATP	adenosine triphosphate
BBB	blood brain barrier
BCP	1-bromo-3-chloropropane
bp	base pairs
BSA	bovine serum albumin
BSF	bloodstream form
c (as prefix)	centi (10^{-2})
cAMP	cyclic-adenosine monophosphate
CBP	camodulin binding peptide
CD	catalytic domain

cDNA	complementary deoxyribonucleic acid
cKO	conditional knockout
CNS	central nervous system
CO	cytochrome oxidase
CR	G versus C strand biased gene
Ct	cycle threshold value
Cyb	apocytochrome B
da	dalton
DEPC	diethylpyrocarbonate
Dk (as prefix)	dyskinetoplastic
dN	number of non-synonymous mutations
dNTP	deoxyribonucleotide
DNA	deoxyribonucleic Acid
dS	number of synonymous mutations
DTT	DL-Dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESAG	expression site associated gene
et al	“and others”
EtBr	ethidium bromide
ExoUase	U-specific exonuclease
FBS/FCS	foetal bovine serum/foetal calf serum
FE	fully edited
g	gramme
G4.18	neomycin
GAP	guide-ribonucleic acid associated protein

gGAPDH	glycosomal glyceraldehyde dehydrogenase
gRNA	guide-ribonucleic acid
HAT	human African trypanosomiasis
Hb/Hp	haemoglobin/haptoglobin receptor
HRP	horseradish peroxidase
ID	interaction domain
IgG	Immunoglobulin-G
K (as prefix)	kilo (10^3)
k (as prefix)	kinetoplast
KPAF	kinetoplast polyadenylation and unridylylation Factor
KPAP1	kinetoplast poly(A) polymerase 1
KREL	kinetoplast ribonucleic acid ligase
KREN	kinetoplast ribonucleic acid editing endonuclease
KREPA	kinetoplast ribonucleic acid editing protein A
KREPB	kinetoplast ribonucleic acid editing protein B
KREH	kinetoplast ribonucleic acid editing helicase
KRET	kinetoplast ribonucleic acid editing 3′ terminal uridylyl transferase
KREX	kinetoplast ribonucleic acid editing U-specific exonuclease
l	litre
LSU	large subunit
MEAT1	mitochondrial editosome-like complex associated TUTase 1
MERS1	mitochondrial editing ribonucleic acid stability factor 1
MOPS	3-(N-morpholino) propanesulfonic acid
MRB1	mitochondrial binding complex 1

mRNA	messenger-ribonucleic acid
MRP	mitochondrial ribonucleic acid binding protein
m	metre
m (as prefix)	milli (10^{-3})
M	molar
mt (as prefix)	mitochondrial
MURF	maxicircle unidentified reading frame
n (as prefix)	nano (10^{-9})
ND/ NADH	nicotinamide adenine dinucleotide dehydrogenase
nt	nucleotides
OB	oligonucleotide/oligosaccharide binding
P38	protein of 38 kilo-daltons
P93	protein of 93 kilo-daltons
PAP	peroxidase Anti-Peroxidade
PBS	phosphate-buffered Saline
PCF	procyclic Form
PCR	polymerase chain reaction
PE	partially edited
pH	measure of acidity or basicity of a solution
POLIB	DNA polymerase 1B
PPi	pyrophosphate
PPR	pentatricopeptide repeat
Q-PCR	quantitative- polymerase chain reaction
r (as prefix)	ribosomal
RBP16	ribosomal binding protein of 16 kilo-daltons

RECC	ribonucleic acid editing core complex
REL	ribonucleic acid editing ligase
RNAi	ribonucleic acid interference
RPS12	ribosomal protein 12
RT	reverse transcription
RT-PCR	reverse transcription – polymerase chain reaction
S	Svedburg unit(s)
S/C	subcomplex
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHAM	salicylhydroxamic acid
SIF	stumpy induction factor
SSC	saline-sodium chloride
SSU	small subunit
T4Rnl	T4 bacteriophage ribonucleic acid ligase
TAC	tripartate attachment complex
TAO	trypanosome alternative oxidase
TAP	tandem affinity purification
TbRGG	*Trypanosoma brucei* protein with RGG domain
TBST	tris buffer saline, with tween 20
tet	tetracycline
TEV	tobacco etch virus
TEVCB	tobacco etch virus cleavage buffer
TNFα	tumour necrosis factor α
tRNA	transfer- ribonucleic acid

TUTase	3′ terminal uridylyl transferase
U (after number)	unit
Us	uridylyl subunits
UE	unedited
UMSBP	universal minicircle sequence-binding protein
V	volts
VAT	variable antigen type
VSG	variable surface glycoprotein
v/v	volume per volume
WHO	world health organisiation

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 African trypanosomes and trypanosomiasis

African, or salivarian, trypanosomes (Genus *Trypanosoma*) are the causative agents of sleeping sickness in humans and nagana, dourine and surra in domesticated livestock. The collective clinical manifestations of trypanosomiasis are referred to as human African trypanosomiasis (HAT) and animal African trypanosomiasis (AAT), respectively, although mechanically transmitted salivarian trypanosomes (such as *Trypanosoma vivax, T. evansi* and *T. equiperdum*) have escaped their evolutionary origins in Africa to be transmitted globally, independently of tsetse flies (Jordan, 1986).

Species of trypanosome differ, not only in their pathogenic abilities, but also in their host specificity. When HAT and AAT are described they refer to distinct, but occasionally overlapping, species of disease organisms. For example, *T.brucei brucei, T. brucei rhodesiense, T. vivax, T congolense. T. evansi, T. equiperdum* and *T. simiae* cause AAT and *T. brucei rhodesiense* and *T. brucei gambiense* cause HAT (Jordan, 1986). *T. brucei sp* are limited in geographical distribution by their tsetse fly vectors, which are restricted to sub-Saharan Africa (See Figure 1.1 for HAT specific distribution). As a consequence, the spread of human settlement and agriculture in this area has been markedly influenced by their presence. HAT is predominantly a disease affecting rural communities, however tsetse flies have adapted to urban environments (Courtin *et al.*, 2009).

HAT still remains one of the 13 neglected tropical diseases, affecting people across some 36 endemic sub-Saharan countries today (http://www.who.int/mediacentre/factsheets/fs259/en/).

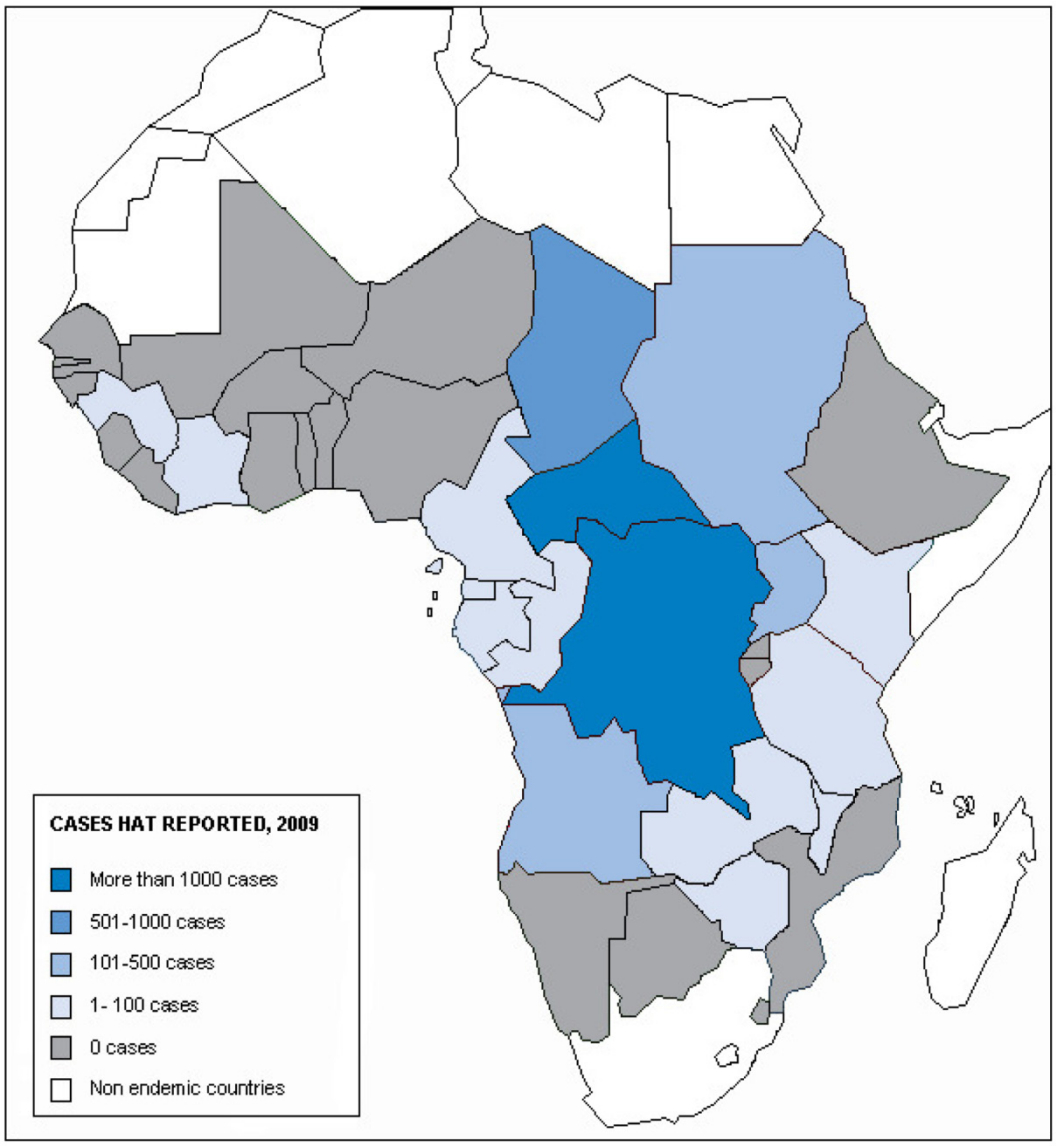


Figure 1.1. Cases of African trypanosomiasis reported from 2000 to 2009.

The distribution of *T. brucei* unlike mechanically transmitted trypanosome species is governed by the availability of tsetse fly hosts. Areas in sub-Saharan Africa most affected are those with prolonged periods of civil unrest, such as the Democratic Republic of Congo, the Central African Republic and Chad.

Reproduced from (Simarro *et al.*, 2011).

Historically, the worst afflicted countries are the Democratic Republic of Congo, Sudan and Angola, where re-emergence of the disease is closely associated with conflict and civil unrest (Ford, 2007). More recently, the worst afflicted countries are the Democratic Republic of Congo, the Central African Republic and Chad (Simarro *et al.*, 2011). The number of new cases dropped below 10,000 in 2009; a drop of 63% of reported cases since 2000. In 2009, 19 countries, of the original 36, reported no disease and in 2010 less than 8,000 new cases were reported (http://www.who.int/mediacentre/factsheets/fs259/en/).

HAT is a product of two *Trypanosoma brucei* subspecies; *T. brucei rhodesiense* and *T. brucei gambiense*, which are distributed in East and West Africa respectively and largely segregated geographically by the rift valley, only coexisting within Uganda where the subspecies are still separated geographically (Welburn *et al.*, 2001). *T. b rhodesiense* is zoonotic and the movements of cattle provide a constant source of focal outbreaks (Fèvre *et al.*, 2001). AAT may be found worldwide, since some (sub)species have evolved means of transmission that are independent of the tsetse fly vector (Brun *et al.*, 1998- see Section 1.2.2). AAT and HAT are constraints to both social and economic development in Africa; however the persistence of AAT is the greatest burden as unlike HAT, it is widespread throughout Africa, whereas HAT is focally distributed (Jordan, 1986).

The decrease of HAT and AAT cases around the turn of the 20th century coincided with the deaths of several ungulates and cattle around Africa due to rinderpest epidemics, which were thought to cause an initial decrease in tsetse numbers due to loss of an important reservoir host (Jordan, 1986; Rogers and Randolf, 1988). However, soon after rinderpest had left its mark, the number of HAT

cases exploded, spurring the excessive slaughter of game animals and bush clearing in an attempt to control the disease. The 1901 to 1910 epidemics of HAT in Uganda were thought to be caused by *T. b. gambiense* and therefore thought to be largely down to the migration of people, either through forced or voluntary movements. It is likely that abandonment of farm land and subsequent encroachment of tsetse-infested vegetation had its effect (Jordan, 1986), however in 2004, a study by Fèvre *et al* demonstrated that these outbreaks were due to *T. b. rhodesiense*. It was also speculated that AAT outbreaks were triggered by the restocking of infected cattle to replace those lost in the rinderpest pandemics (Fèvre *et al.*, 2004). The most recent outbreak highlighted by the WHO was in 1970, after the near disappearance of the disease in the 1960s. This was due to the subsequent relaxation of control efforts. (http://www.who.int/mediacentre/factsheets/fs259/en/)

1.1.1 Diagnosis and current control of trypanosomiasis

The differential diagnosis of HAT into *T. b gambiense* or *T. b rhodesiense,* and determining the stage of infection are essential to effective treatment of trypanosomiasis. Whilst clinical symptoms and serological or molecular methods give indirect diagnosis of the subspecies of *T. brucei,* parasitological identification of trypomastigotes in blood (finger prick) or cerebral spinal fluid (lumbar puncture) is key to identifying the stage of infection (Wastling and Welburn, 2011). There has been extensive research into potential biomarkers of subspecies and indicators of the stage of infection for HAT. PCR-based methods of identification have been reviewed by Gibson, (2009).

The diagnosis of HAT is based on screening and subsequent confirmation of parasitemia and identification of the stage of infection (Malvy and Chappuis, 2011). The card agglutination test for trypanosomiasis (CATT) provides a means of serological screening, and is regularly used in countries of endemicity (Magnus *et al.*, 1978). Stage specific diagnosis must be made from parasitological confirmation from blood and cerebral spinal fluid (CSF) samples. This is achieved through analysing blood and lumbar puncture smears microscopically, the former of which may require microhaematocrit centrifugation due to low sensitivity (Woo, 1970). The diagnosis of *T. b. rhodesiense* is slightly different from that of *T. b. gambiense*, since its diagnosis is also based upon clinical symptoms and history of exposure (Malvy and Chappuis, 2011).

History has shown previously that relaxing of control methods, usually during periods of civil unrest and often with large population movements of people and livestock, leads to epidemics, even with relatively low number of tsetse present. Therefore, the most important aspects of control are that of organisation and commitment for vector control and trypanosome treatment to be effective against the burden of AAT and HAT (Seed, 2001).

1.1.2 The infectivity and pathogenesis of *T. brucei*

Only two subspecies of *T. brucei* typically infect humans. Humans have innate immunity to *T. b. brucei* due to the presence of a trypanolytic high density lipoprotein bound to a human specific lipoprotein called apolipoprotein L1 (ApoL1) (Vanhamme *et al.*, 2003). ApoL1 is taken up the parasite via endocytosis, and from within produces pores within the lysosome, rendering the organelle susceptible to osmotic swelling that eventually lyses the cell (Pérez-Morga *et al.*, 2005; Pays *et al.*, 2006). Human infective *T. b. rhodesiense* resists lyses due to the presence of a serum-resistance-associated gene (SRA); a truncated variable surface glycoprotein (VSG) that can neutralise the effects of ApoL1 through physical interaction. The expression of the SRA gene in *T. brucei* is sufficient on its own to confer human infectivity (Van Xong *et al.*, 1998; Vanhamme *et al.*, 2003; Oli *et al.*, 2006). *T. brucei gambiense*, however, lacks an SRA gene, but evades TLF through a mutation accumulated in its Hp/Hb (haptoglobin/haemoglobin) receptor (Kieft *et al.*, 2010; Capewell *et al.*, 2011). It is entirely possible that the spread of *T. b gambiense*, which makes up to 90% of total HAT cases, in West Africa was favoured by the emergence of ApoL1 variants (with G1 and G2 mutations) in the human population of this region, which made up 38% and 8% of the population for G1 and G2 variants, respectively. ApoL1 mutations strongly correlate with an increased risk of renal disease by an unknown mechanism, which may prevent the mutations from reaching fixation. This population is naturally resistant to *T. b rhodesiense* (Genovese *et al.*, 2013).

Living extracellularly presents the problem of immunological attack by the hosts' defences. To overcome this *T. brucei* exhibits clonal antigenic variation

(AGV) (Turner, 1999; Matthews *et al.*, 2004). This is the process by which the VSG protein coat is switched to another isoform to produce a distinct variable antigen type (VAT). Proliferation of these new VATs requires a new specific immune response, which takes time, allowing a new wave of parasitemia to take hold. This switch is pre-emptive and not induced by antibody production and a new and distinct VAT presents itself approximately every 1000 doublings (Turner and Barry, 1989; Becker *et al.*, 2004).

VSGs are only expressed when positioned within an active expression site (Borst, 1986). The clonal phenotypic expression of VSGs can be modulated since only one of the 10 - 20 VSG expression sites (ESs) is transcribed at any one time, whilst the rest are silenced. Whilst VSGs in silent ESs may be activated, VSGs at subtelomeric regions may also switch by conversion (Hertz-Fowler *et al.*, 2008). Gene conversion events can occur due to the VSGs homologous flanking sequences which allows VSG switching that is independent of sequence (Horn and Barry, 2005; Boothroyd *et al.*, 2009). Conversion events may occur simply with a gene switch, but the VSG expressed may not be completely replaced leaving a hybrid VSG gene at an active ES (Bernards *et al.*, 1981). Often mosaic genes are formed from segmental conversions of pseudogenes and silent VSG gene families, which would not be able to be expressed through a simple switch (Thon *et al.*, 1989; Roth *et al.*, 1991; Barbet and Kamper, 1993). The presence of multiple ES's allows the potential switching between two VSGs, of which there are <1000 at sub-telomeric locations. For these aforementioned reasons, and due to the polycistronic arrangement of genetic material in *T. brucei,* infections may span many months or years without the hosts' immune system clearing them (Borst, 1986; Johnson *et al.*, 1987; Berriman *et al.*, 2005; Horn

and McCulloch, 2010; MacGregor *et al.*, 2012).

1.1.3 The limitations of current available chemotherapeutics

The clinical presentation of sleeping sickness may manifest from two separate stages of the disease. The first (early or S1) stage is haemolymphatic and the second (late or S2) stage occurs when the parasites cross the blood brain barrier (BBB) and enter the central nervous system (CNS) causing meningoencephalitis and sleep and behavioural disturbances. Human infections can lay latent for many years before any symptoms emerge (Barrett *et al.*, 2003). S2, if left untreated may lead to coma and death (Malvy and Chappuis, 2011). The pathology of African trypanosomiasis is thought to arise from an uncontrolled type 1 cell response, acting via TNFα and the immune molecules actions upon macrophages (Magez *et al.*, 1999). Another feature of the immune response is a great increase in the IgM levels, both specific and non-specific to trypanosomes (Vincendeau and Bouteille, 2006). It was once thought that HAT was fatal in all cases, if not treated. However, more recent studies suggest that *T. b. gambiense* was not 100% fatal, and did not always follow the classical progression (S1 - S2) of HAT to neurological involvement. This suggests that either human hosts were able to clear infection, or could live as asymptomatic hosts (Jamonneau *et al.*, 2012). Since no vaccine is available, it means that clear diagnosis of subspecies, prophylaxis of cattle hosts and chemotherapeutics are the only way to treat this disease. HAT diagnosis tends to be late, so chemotherapeutic treatment is reliant on the efficacy of S2 treatment (Wastling and Welburn, 2011).

Drug efficacy, practicality and affordability remain great obstacles in the

control of sleeping sickness, since tsetse fly control through pesticide spraying, targets and traps will never be a long term solution. Only four drug combinations are available for treatment of HAT (as discussed by Steverding, 2010). All of these chemotherapeutics have to be administered intravenously under hospital stay, which is far from ideal, since many people in tsetse endemic regions do not have easy access to health facilities. Most of these treatments cause adverse effects in patients. For example, melarsoprol, given for S2 infection in *T. b. rhodesiense*, kills 5 - 10% of patients that receive it (Legros *et al.*, 2002; Balasegaram *et al.*, 2006). The safest S2 treatment is Eflornithine-Nifurtimox (NECT) combination therapy, but is only available for the treatment of *T. b. gambiense* (Priotto *et al.*, 2009).

Drug resistance is a continuing problem and is associated with a variety of parasitic responses to selection pressures. Resistance can be readily seen in the field and has two main themes; firstly, parasites show reduced drug uptake and secondly, parasites exhibit cross-resistance to arsenicals and diamidines (Williamson and Rollo, 1959; Damper and Patton, 1976; Frommel and Balber, 1987; Mäser *et al.*, 2003). For Melarsoprol and Pentamidine resistance, this is because both types of compounds are taken up by the same P2 adenosine transporter, which is deficient in resistant cell lines and so reduces uptake of these compounds into the cell (Carter and Fairlamb, 1993; Carter *et al.*, 1995; de Koning *et al.*, 2000). Recent studies have also identified the presence of an aquaporin (AQP) 2 that may ordinarily have a role in osmoregulation or glycerol transport, which is also responsible for conferring cross-resistance to Pentamidine and Melarsoprol (Baker *et al.*, 2013).

The ideal solution to the problem of trypanosomiasis control would involve the systematic development of new oral therapeutics; however, drug discovery

remains limited due to low economic input (Croft *et al.*, 2005).

1.1.4 The lifecycle of *T. brucei*

T. brucei is termed pleomorphic, as it exists in many morphological forms through the course of its lifecycle (see Figure 1.2). In the mammalian host there exists two main bloodstream forms (BSFs): a proliferative slender form and a quiescent stumpy form, named due to their diverse physical appearance under a microscope (Vickerman, 1965; Tasker *et al.*, 2000).

Slender forms proliferate to increase or establish parasitemia, but are removed by immune lysis. Thus infection is self limiting, and the parasite may actually exploit their host's immune response to ensure a chronic infection and to increase the chances of being picked up by a tsetse vector (Pays *et al.*, 2001). The formation of stumpy trypomastigotes is thought to be density dependent, and it has been suggested that stumpy induction factor (SIF), a molecule produced by the parasite, mediates cell density sensing in the blood stream via the cAMP pathway (Vassella *et al.*, 1997). The differentiation of slender to stumpy cells is not due to the host's immune response, as immunodepressed hosts may still generate stumpies (Matthews and Gull, 1994). Nonetheless, the density dependent (utilising positive feedback) control of stumpy cell differentiation and the arrest of cell cycle division at phases G0 and G2 of the cell cycle maintains the parasitemia below a threshold that would be otherwise fatal to their host (Vassella *et al.*, 1997). As well as being a way of limiting parasite burden, the formation of stumpies prepares the trypanosome for passage into the tsetse fly midgut and for differentiation to procyclic forms (PCFs).

The additive effects of differentiation and AGV contribute to the infectiveness of the parasite (Vassella *et al.*, 1997; Barry and McCulloch, 2001; Lythgoe *et al.*, 2007; MacGregor *et al.*, 2012).

Once the stumpy parasites are taken up in a blood meal they are taken to the midgut. Here they develop into PCFs within 24 hours at the posterior of the midgut. This differentiation can be reproduced *in vitro,* and is apparent after 8 - 10 hours (Roberts *et al.*, 2000; Fenn and Matthews, 2007). After approximately 10 days of midgut establishment PCFs escape through the peritotrophic membrane via the anterior of the midgut to the proventriculous and colonises this area as elongated proventricular PCFs. After subsequent cell cycle arrest the parasites then migrate to the salivary glands as non-dividing proventricular mesocyclics and proceed to attach themselves to the microvilli of the salivary glands via their flagellum, where they divide as epimastigotes. Once the parasites disengage from the salivary gland cells and change to metacyclic form, the tsetse fly becomes infective and can inject these parasites into their mammalian hosts upon their next blood meal. The whole process from ingested stumpy to injected metacyclic takes 15 to 35 days (as described by Roberts *et al.*, 2000; Sharma *et al.*, 2009).

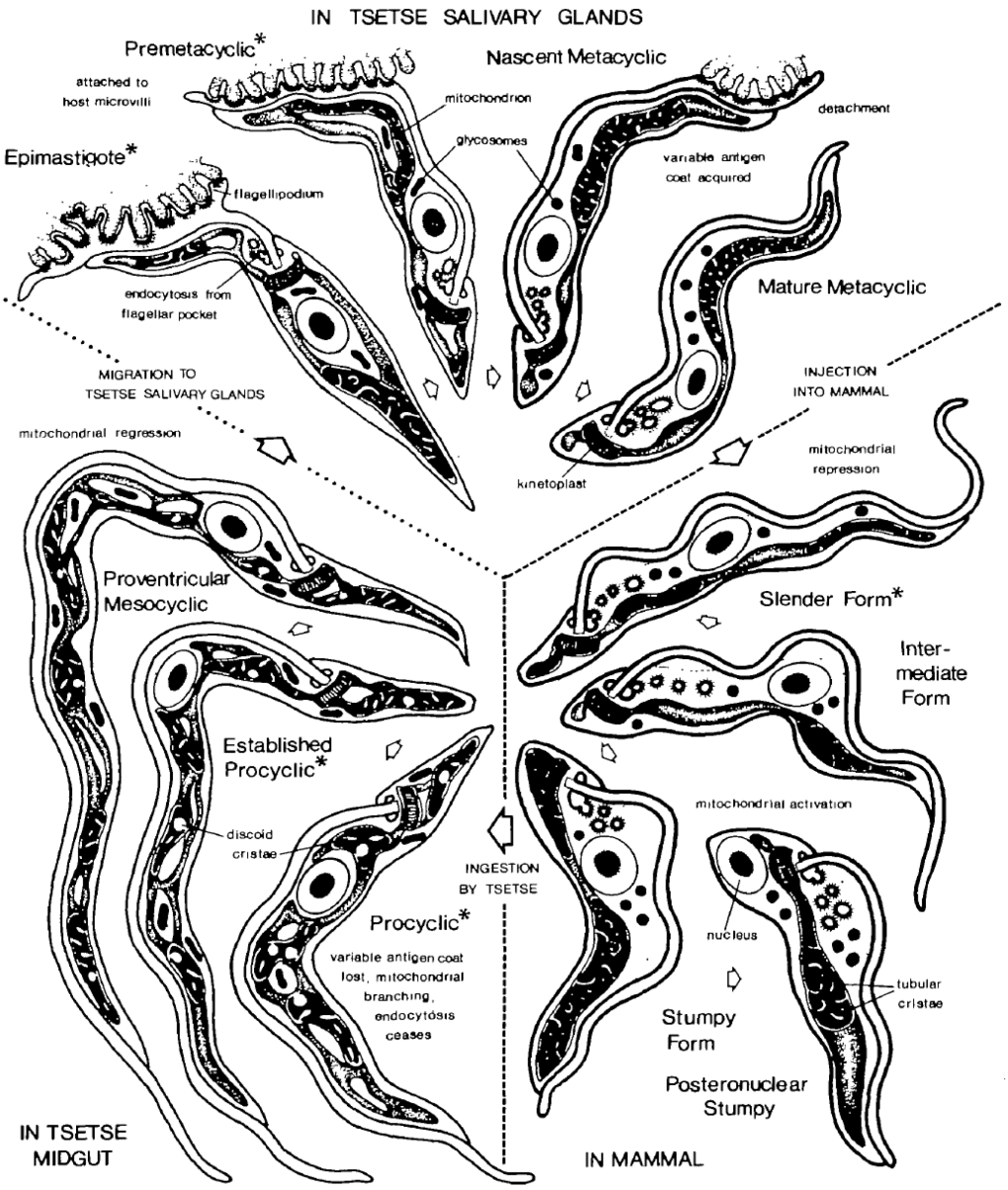


Figure 1.2. The lifecycle of *T. brucei* in the mammalian and insect hosts.

The mammalian host is infected upon the bite of an infected tsetse fly, where infective metacyclic stage trypomastigotes are injected into the bite wound. Bloodstream form trypomastigotes (BSFs) divide as slender forms in the mammalian host and differentiate to stumpy forms at high density, which are subsequently picked up by a feeding tsetse fly. Once inside the midgut, procyclic cells (PCFs) escape the peritrophic membrane and colonise the salivary glands where they replicate as epimastigotes and differentiate to mammalian infective metacyclics. The cell cycle is closely linked to kinetoplast repositioning and mitochondrial remodelling.

Reproduced from (Vickerman, 1985).

The different morphological forms illustrated in Keith Vickerman's lifecycle diagram are brought about in part by kinetoplast repositioning in relation to the nucleus through the elongation of microtubules from the posterior end of the cell (Matthews *et al.*, 1995). Central to the lifecycle of *T. brucei* is the process of metabolic and mitochondrial remodelling that equips the parasite for two very different host environments, which will be discussed in more detail in the next Section.

1.2 The unique mitochondrion of *T. brucei*

Within the mammalian bloodstream, glucose is abundant, and so trypanosomes make use of this through glycolysis (see Michels *et al.*, 2006). In *T. brucei* the first seven steps of glycolysis occur within a specialised peroxisome-like organelle, the glycosome. The last three steps, in the cytosol, produce pyruvate, which is excreted and transported to the mitochondrion (Opperdoes, 1987).

In PCFs, the branched mitochondrion (see Figure 1.3 A) allows for the utilisation of more carbon sources than just glucose that are abundant in the tsetse fly midgut, such as threonine and proline (Cross *et al.*, 1975; Coustou *et al.*, 2003). Here glycolysis produces acetyl Co-enzyme A (acetyl-CoA) and subsequently acetate, via degradation of pyruvate, within the mitochondria (see Tielens and Van Hellemond, 1998; Michels *et al.*, 2006). Unusually, acetyl-CoA doesn't feed the tri carboxylic acid (TCA) cycle, even though all TCA enzyme are present, rather this enzyme is utilised for *de novo* fatty acid synthesis via an unusual elongase system (Durieux *et al.*, 1991; van Weelden *et al.*, 2003; Lee *et al.*, 2006). The TCA cycle is not active as a cycle, but is fed via glutamate (derived from scavenged L-proline) to produce succinate as its end product (see Besteiro *et al.*, 2005). The phosphopyruvate generated from glycolysis can also re-enter the glycosome to create succinate directly (see Michels *et al.*, 2006). PCFs display classical electron transport chains, similar to mammals, where glycerol-3-phosphate produced in the glycosome shuttles electron to the mitochondrion and respiratory complexes I and II provide electrons to ubiquinone, which can then be transferred to the complexes III and IV. Unusually, trypanosomes have a SHAM sensitive, plant like, trypanosome alternative oxidase (TAO), which is the only terminal oxidase that BSF *T. brucei* possess and is vital to

aerobic respiration in the parasite (Clarkson *et al.*, 1989; McIntosh, 1994; Chaudhuri and Hill, 1996). In PCFs, an alternative branch leads to the TAO, the terminal oxidase. Complex V acts as an ATP synthase in the final step, exploiting the proton-motif force (see Van Hellemond *et al.*, 2005; Besteiro *et al.*, 2005).

The structure and function of the mitochondrion of bloodstream form trypanosomes (BSFs) is unique (see Figure 1.3 B). The much reduced organelle lacks the key components required for oxidative phosphorylation and TCA cycle, therefore, BSFs rely on glycolysis for ATP production, which indicates adaptation to a glucose rich environment (see Van Hellemond *et al.*, 2005; Tielens and Van Hellemond, 2009). As with PCFs, oxygen is the final electron acceptor. To maintain a mitochondrial membrane potential BSFs must employ respiratory complex V, the ATP synthase, as an ATPase, where a proton gradient is generated from the hydrolysis of ATP (Schnaufer *et al.*, 2005).

Several studies have been undertaken to elucidate the differential expression of mitochondrial subunits between BSFs and PCFs. Although it had previously been thought that complex 1 (NDH1) was absent in *T. brucei*, both complex 1 and an alternative complex 1 (NDH2) are present in both lifecycle stages, although the formers' proton pumping activity is uncertain. This is why it was assumed complex 1 was active but non-essential to cell metabolism and growth, since the two are most likely functionally redundant. In this scenario NADH reducing agents are shuttled from the glycosome to mitochondria by glycerol-3-phosphate, which shows the importance of this organelle in both lifecycle stages (Panigrahi *et al.*, 2008; Panigrahi *et al.*, 2009; Verner *et al.*, 2011; Surve *et al.*, 2012; Verner *et al.*, 2013).

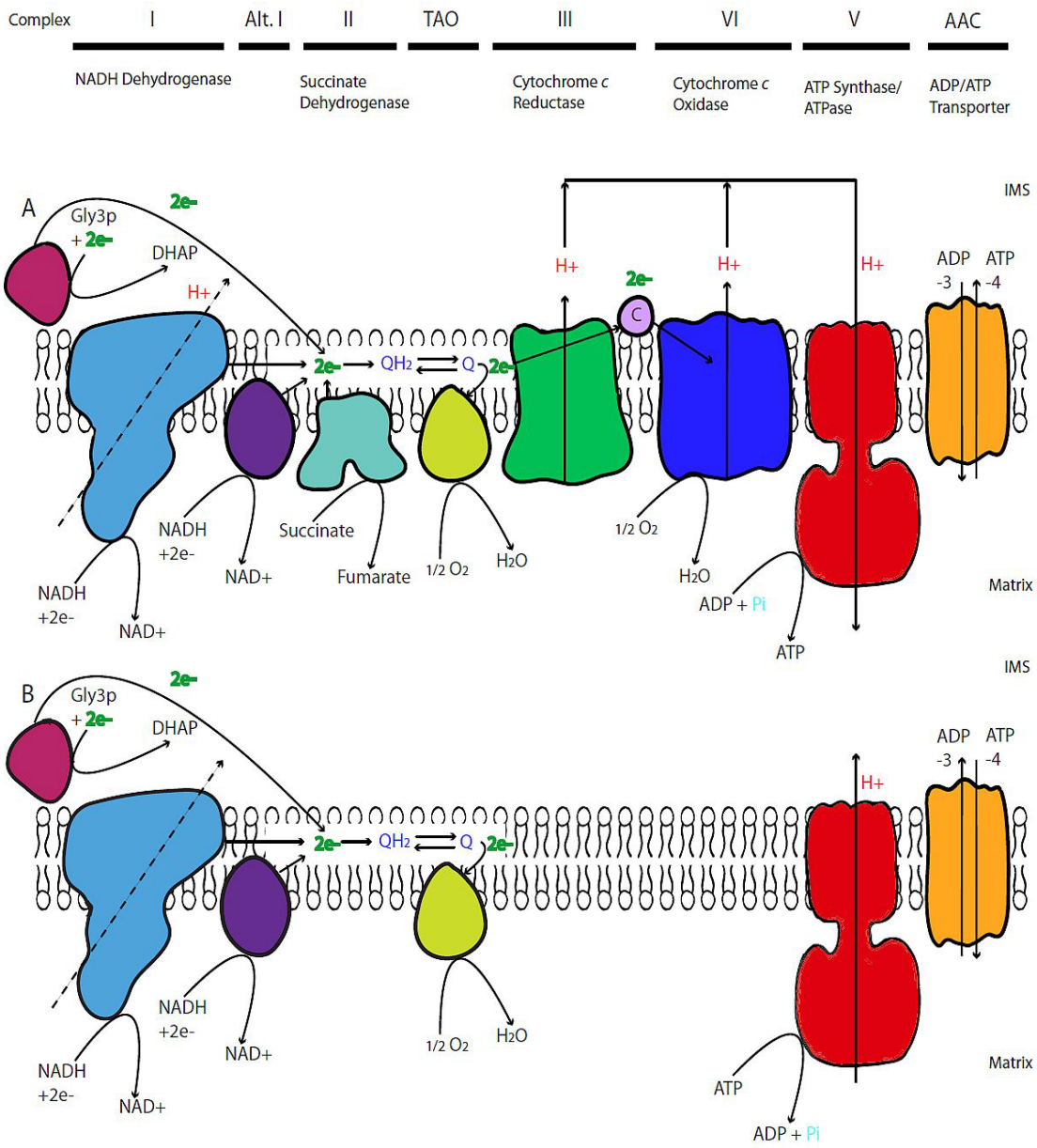


Figure 1.3. Expression and function of mitochondrial complexes in PCF and slender BSF *T. brucei.*

Slender BSF (B) trypanosomes have a reduced, sack like mitochondrion, which has fewer cristae and respiratory chain subunits, in comparison to PCF cells (A). Slender BSFs lack electron transport chain components, and complex V works as an ATPase, generating a proton gradient through the hydrolysis of ATP. PCFs have a mitochondrion more similar to mammals than BSFs, in that they have a working electron transport chain, and their complex V works as an ATP synthase. Uncertain processes are indicated by a dashed line. Complex 1 (NDH1) contains the subunits encoded for by NADH dehydrogenase subunit 7 (ND7). This transcript is focussed upon in Chapter 3.

Abbreviations: TAO - trypanosome alternative oxidase, Gly3p – glycerol-3-phosphate, IMS - intermembrane space and Alt.I – Alternative complex 1(NDH2).

Figure 1.3 illustrates NDH2 facing the matrix, as postulated by Coustou *et al.*, 2008; Tielens and Van Hellemond, 2009, however, it has also been suggested that this complex may face the intermembrane space (IMS). This more unlikely scenario was concluded indirectly, because the matrix is impermeable to NADH and addition of NADH to permeated mitochondria caused an increase in oxidative activities (Verner *et al.*, 2013). Although the *T. brucei* mitochondrion is an obligate aerobic organelle, the parasite only partially oxidises its metabolic substrates, the reasons for this is not known (Tielens *et al.*, 2002).

Since different substrate environments dictate the type of energy metabolism employed by the parasite, it can be assumed that they have evolved a degree of metabolic flexibility, which has proven useful in establishing themselves in new host environments (Tielens and Van Hellemond, 2009).

1.2.1 The kinetoplast and the kDNA network

Trypanosomes contain a single mitochondrion per cell, and the mitochondrial DNA is organised as a structure called the kinetoplast. In *T. brucei* this mitochondrial (mt) or kinetoplast (k) DNA makes up to 20% of the total DNA content of the cell, and is arranged into thousands of concatenated, highly heterogeneous minicircles (of ~1 kb in length) and 40 to 50 homogeneous maxicircles (20 to 40 kb in length). Minicircles encode uniquely *trans*-acting guide (g)RNAs that bind through Watson-Crick base-pairing to maxicircles that encode pre-edited and never edited mitochondrial transcripts (Englund and Marini, 1980; Hajduk *et al.*, 1986; Benne *et al.*, 1986; Simpson, 1987; Simpson *et al.*, 1987; Simpson and Shaw, 1989; Stuart *et al.*, 1989; Blum *et al.*, 1990; Pollard *et al.*, 1990; Sturm and Simpson, 1990a; Shapiro and Englund, 1995). The arrangement of maxicircles and minicircles, and the transcripts that they encode are summarised in Figure 1.4. This order of mitochondrial genes is present in all species of trypanosomatids investigated, but pre-mRNAs are edited to different degrees (Hajduk and Ochsenreiter, 2010).

Both maxi- and minicircle DNA is transcribed through the action of mtRNAP, a polymerase that closely resembles at T7 phage RNA polymerase (Grams *et al.*, 2002; Hashimi *et al.*, 2009).

For a single transcript, all corresponding gRNAs must be maintained lest the mature transcript be lost. To ensure that minicircle classes are not lost during random segregation events, they are catenated when not segregating and each of the smaller minicircles are interlocked (without being supercoiled) to an average of three

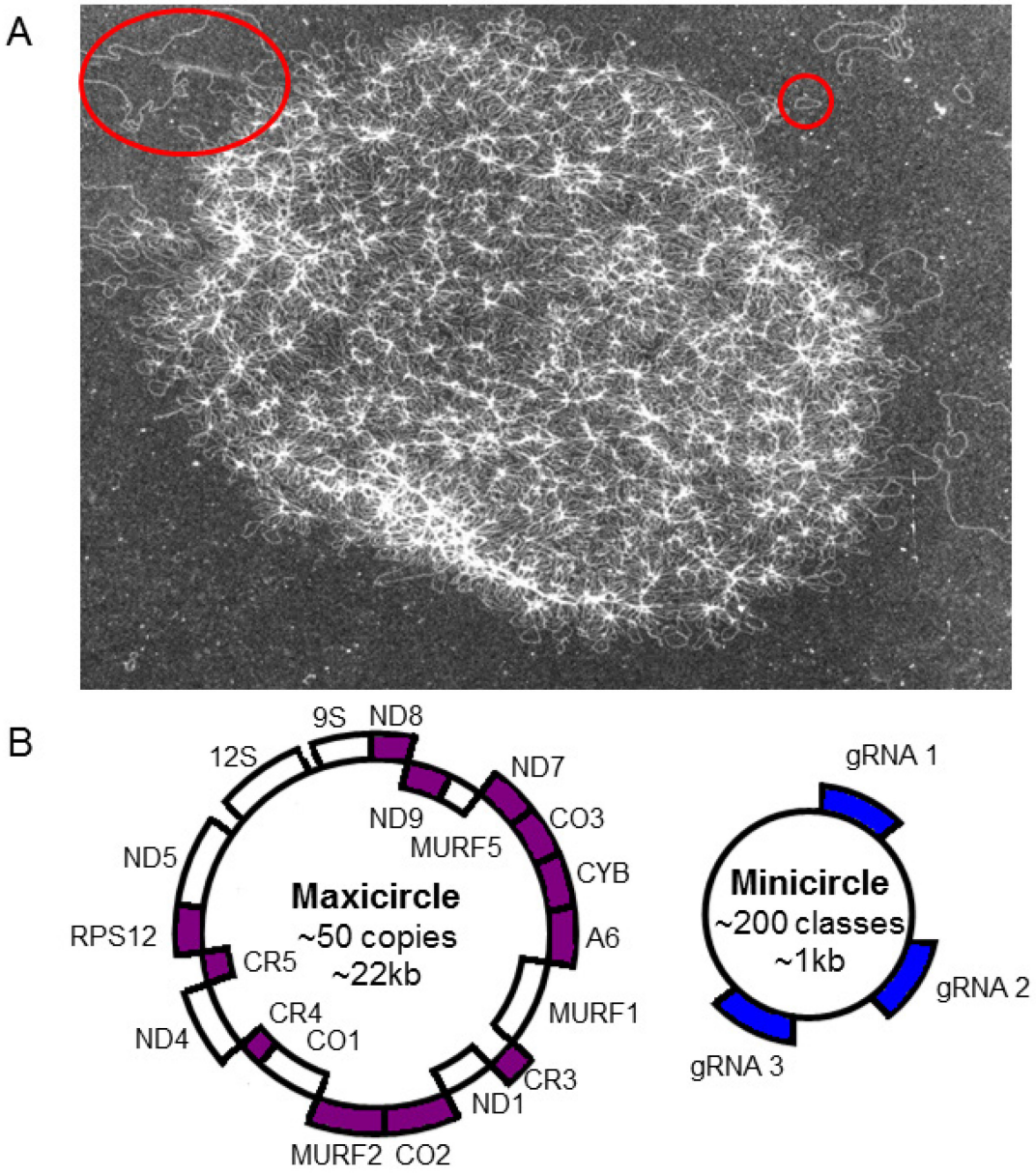


Figure 1.4. The kinetoplast network of mitochondrial DNA molecules.

As with the rest of the *T. brucei* genome, mitochondrial transcripts are transcribed into polycistronic units before post-transcriptional cleavage and subsequent kDNA specific editing. A) Highlighted in red are a maxicircle (left) and minicircle (right) separated from the kDNA network. B) Schematics of the maxicircle and minicircle DNA. Maxicircles cryptogenes (i.e. genes the transcripts of which undergo editing) are highlighted in purple. The degree of editing in *T. brucei* for each mitochondrial transcript is summarised in Table 1.1.

neighbouring ones (Borst, 1991; Drew and Englund, 2001; Liu *et al.*, 2005; Jensen and Englund, 2012). It should be stressed; however, that the exact manner in which minicircles and maxicircles entwine is not well understood (Shapiro, 1993). Minicircles exist in over 200 different classes, of which there are multiple minicircles each with 3 - 4 separate gRNA transcripts coexisting alongside. This presents the cell with approximately 10, 000 separate gRNAs (Steinert and Van Assel, 1980; Blum *et al.*, 1990; Pollard *et al.*, 1990; Sturm and Simpson, 1990a). The sequences of these gRNAs are especially heterogenous in *T. brucei* and are directly correlated with the extensiveness of RNA editing that takes place to produce mature mitochondrial transcripts (Stuart and Feagin, 1992).

All minicircles have a conserved region, which also contain their origin(s) of replication (Birkenmeyer *et al.*, 1987). Minicircle replication requires a number of proteins such as topoisomerase II, UMSBP, POLIB, p38, p93 and various mitochondrial DNA (mtDNA) polymerases to aid in the initiation of transcription, as well as maintenance of the kDNA network and subsequent segregation of catenated and daughter minicircles (Wang and Englund, 2001; Liu *et al.*, 2006; Liu and Englund, 2007; Milman *et al.*, 2007; Li *et al.*, 2008; Liu *et al.*, 2009; Bruhn *et al.*, 2010; Liu *et al.*, 2010).

In general gRNAs are transcribed from minicircles and pre-mRNAs from maxicircles. Unusually, COII gRNA required for its editing is contained within the 3' end of the primary transcript, and works in *cis* but not *trans*. Most gRNAs work in *trans,* since they are transcribed elsewhere and are shuttled to the editosome (Clement *et al.*, 2004; Golden and Hajduk, 2005).

1.2.2 Living without mitochondrial DNA

Rather unusually, cells exist that have a reduced, or completely absent, kinetoplast and are termed dyskinetoplastic (Dk) or akinetoplastic (Ak) respectively. Ak (*T.* evansi) and Dk trypanosomes (*T. equiperdum* and *T. evansi*) can be found in the wild. These parasites have lost the ability to differentiate into insect stage PCFs and can only be transmitted through biting flies or venereally, which in turn has allowed them to leave the tsetse fly belt in sub-Saharan Africa (Hoare, 1937; Tobie, 1951; Riou and Saucier, 1979; Brun *et al.*, 1998). Examples of naturally occurring *T. evansi* lack maxicircle DNA, and have only a single gRNA class (Borst *et al.*, 1987; Songa *et al.*, 1990). Ak *T. brucei*, which is lacking in all kDNA and associated mRNAs has been created through treatment with acriflavine, a DNA intercalator that chemically induces kDNA loss over passage history (Stuart, 1971; Stuart and Gelvin, 1980).

Detailed characterisation of a number of *T. evansi* and *T. equiperdum* isolates by Lai *et al.*, 2008 confirmed various degrees of kDNA loss, including partial or complete maxicircle deletions and minicircle homogenisation. They proposed that the latter may have been the result of lack of genetic exchange within the tsetse fly vector, in turn a consequence of being locked within the mammalian host (Lai *et al.*, 2008).

The loss of kDNA in trypanosomes will ultimately have downstream consequences on mitochondrial biogenesis. The A6 subunit of complex V appears to be the only editing product ultimately required in BSFs, and its loss can be compensated for by a mutation in the γ subunit of F_1 (Bhat *et al.*, 1990; Schnaufer *et*

al., 2005). In the case of *T. brucei,* a single L262P mutation within the C terminal of the γ-subunit of F1 complex V was sufficient to allow complete kDNA loss on treatment with acriflavine (Dean *et al.,* 2013). This mutation caused the uncoupling of F_1 and F_O, rendering complex V an obligatory ATPase from a facultative ATP synthase. It was already known that petite-negative yeast can exist without mtDNA and still maintain a functioning mitochondrion, as long as they exhibited certain mutations in the F_1 moiety of complex V. These allow the generation of a membrane potential by acting as an ATPase in conjunction with electrogenic exchange of ATP^{4-} / ATP^{3-} by the ATP/ADP carrier (Clark-Walker *et al.*, 2000) (see Figure 1.3).

This aside, lab-induced Dk cells have maintained functional (in the presence of a gRNA substrate) editosomes, since editosome genes are nuclear encoded, even though the RNA editing process is redundant in these cells (Schnaufer *et al.*, 2002; Domingo *et al.*, 2003).

1.2.3 Complex editing in *T. brucei*

RNA editing in *T. brucei* is an essential and extensive post-transcriptional process with 12 of the 18 mitochondrially encoded protein-coding transcripts being edited (Estévez and Simpson, 1999). For a list of these transcript and the subunits they encode for, see Table 1.1. The process of editing (Figure 1.5) is essential since it creates a functional open reading frame from pre-mRNA maxicircle transcripts, through the use of gRNA templates, and in doing so produces a mature transcript suitable for translation (Estévez and Simpson, 1999). RNA editing was first described by Benne *et al.*, (1986), as the process where 4 uridylyl (U) residues were added to the COII transcript of *T. brucei* and *Crithidia fascilata.*

There have been three models of the editing process discussed in the literature. Firstly that of transesterification, put forward separately by Blum *et al* and Cech *et al* (1991). In this model the site of editing is determined by the 3' of the gRNA and the gRNA oligo-(U) tail is the source of uridylyls inserted in editing (Blum *et al.*, 1991; Cech, 1991). This has since been dismissed as the number of added uridylyls does not seem to be dictated by the gRNA, although RNA cleavage appears to be gRNA directed (Seiwert and Stuart, 1994; Kable *et al.*, 1996; Seiwert *et al.*, 1996). Also, the discovery of other catalytic components involved in editing, as discussed in Section 1.3, has also undermined this model. A second model is that of cleavage-ligation, where by the editing site is predicted by the pre-mRNA, and is directed by gRNA templates throughout cleavage and addition and deletion of uridylyls, before the rejoining of the mature transcript by a ligase (Blum *et al.*, 1990; Pollard *et al.*, 1992). The third model of editing involved cleavage and ligation through the formation of gRNA-mRNA chimeras (Sollner-Webb, 1991). As with the

cleavage ligation model, the editing site is determined by the pre-mRNA, however after the initial cleavage gRNA-mRNA chimeras are formed and the subsequent number of uridylyls added is determined after the subsequent cleavage of the chimera. The finding of gRNA-mRNA chimeras *in vitro* editing assays suggested that their formation uses the same cleavage-ligation activities as RNA editing itself, in a gRNA dependent manner since the 5' monophosphate born of endonucleolytic cleavage could be a substrate for the 3' end of the gRNAs (Blum and Simpson, 1992; Harris and Hajduk, 1992; Koslowsky, *et al.*, 1992b; Read *et al,* 1992a; Seiwert *et al.*, 1994; Rusché *et al.*, 1995; Piller *et al.*, 1996; Seiwert *et al.*, 1996)

Although chimeras can be detected i*n vitro,* which makes the last model appealing, chimeras are thought to be scarce *in vivo* and can detected only through PCR (Stuart *et al.*, 1997). The first direct evidence for the involvement of a ligase in the production of editing intermediates was provided by Sabatini and Hajduk (1995), who also showed that chimera formation and ligase activity *in* vitro could be inhibited by addition of pyrophosphate, in a dose dependent manner. At present, the nuclease-ligase model of RNA editing is the preferred one, as many other catalytic components involved in editing have since been discovered. These will be discussed in Section 1.3.

Table 1.1. The transcripts involved in RNA editing.

Mitochondrial transcript	Respiratory complex/ function	No. of U insertions/ U-deletions	Length of edited mRNA (nt)
ND1	Complex I	Not edited	
ND3		210/13	452
ND4		Not edited	
ND5		Not edited	
ND7		553/89	1,238
ND8		259/46	574
ND9		345/20	649
Cyb	Complex III	34/none	1,151
COI	Complex IV	Not edited	
COII		4/none	663
COIII		547/4	1 969
A6	Complex V	447/28	811
RPS12	Ribosomal protein S12	132/28	325
MURF1	Unknown function	Not edited	
MURF2		26/4	1,111
MURF5		Not edited	
CR3		148/13	299
CR4		325/40	567
9S rRNA	SSU ribosomal	RNA 30 oligo uridylation	
12S rRNA	LSU ribosomal	RNA 30 oligo uridylation	

12 of the 18 mitochondrial transcripts are edited, some of these extensively so. ND NADH-ubiquinone oxidoreductase subunits 1-9; Cyb: apocytochrome b; CO: cytochrome oxidase subunits I–III; A6: ATP synthase subunit 6; S12: small subunit ribosomal protein 12; MURF: maxicircle unidentified reading frame; CR: G versus C-strand biased genes no. 3 and 4; SSU: small subunit; LSU: large subunit.

Reproduced from Göringer, 2012.

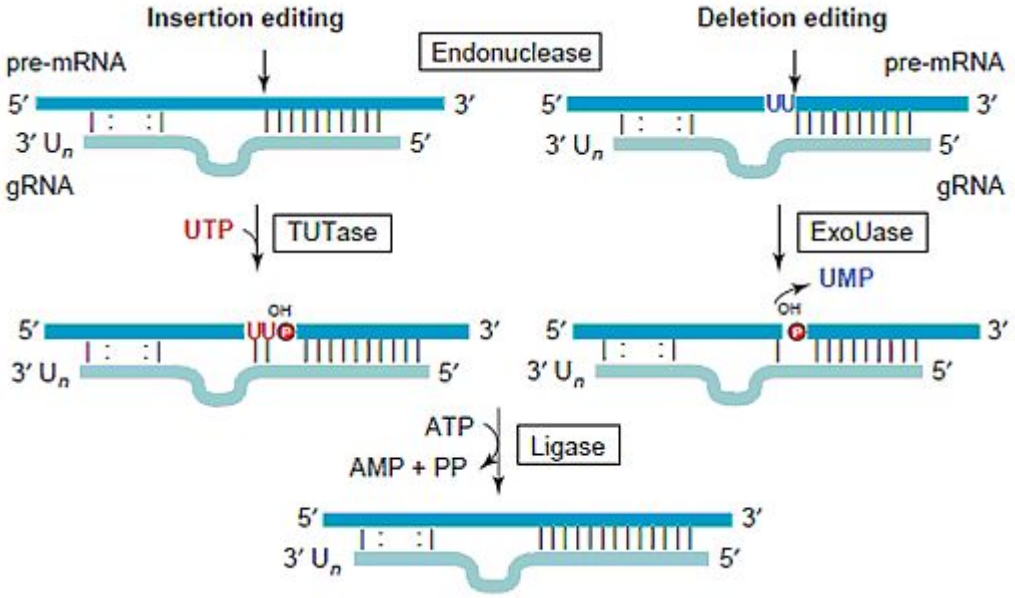


Figure 1.5. The RNA editing process.

In the editing process pre-mRNA mitochondrial transcripts (dark blue) and processed in a 3' to 5' direction by Watson-Crick (dashes) or G:U (colons) base pairing with gRNA templates (light blue). Endonucleolytic cleavage by an editing endonuclease occurs at the first mismatched base pair after the gRNA-pre-mRNA anchor duplex. Uridylyls are either added by a TUTase or deleted by an exoUase, in accordance to the gRNA sequence. All fragments are ligated by an RNA editing ligase (REL). Several rounds of editing and many gRNA templates are required to fully edit the pre-mRNA.

Reproduced from Stuart *et al.*, 2005.

The editing process itself, using the nuclease-ligation model (Figure 1.5) requires the following three steps: endonucleolytic cleavage at an internal editing site, addition or deletion of uridylyls via an RNA editing terminal uridylyl transferase (TUTase) or a U- Specific exonuclease (exoUase) respectively, and finally ligation of the edited transcript fragments by an RNA editing ligase (REL). Addition editing is a more common event that deletion editing (see Stuart *et al.*, 2005).

Editing occurs generally in a 3' to 5' direction, whereby the sequential annealing of gRNAs by Watson-Crick and G:U base pairing at a stretch of 10-15 nucleotide long complementary anchor sequence generates the next available editing site. Endonucleolytic cleavage occurs at the first mismatched nucleotide pair. Editing occurs in different blocks, or domains throughout the transcript (Decker and Sollner-Webb, 1990). It has been suggested that some kind of higher order structure is required for substrate recognition and anchoring of the gRNA-mRNA. This may suggest a proofreading step (Golden and Hajduk, 2006).

The population of mtDNA transcripts in its steady state include fully, partially and unedited mRNAs (Simpson and Shaw, 1989). Of all these transcripts, partially edited ones are the most abundant (Koslowsky *et al.*, 1991). Editing is not a perfect process, however, and misediting often occurs (Sturm and Simpson, 1990a; Sturm *et al.*, 1992). For comprehensive reviews of RNA editing see Stuart *et al.*, 2005 and Aphasizhev and Aphasizheva, 2011.

RNA editing is developmentally regulated independent of gRNA abundance (Koslowsky *et al.*, 1992a). Complex 1 subunits and RPS12 are preferentially edited in BSFs , whereas Cyb and COII are only edited in PCFs (Feagin *et al.*, 1985; Feagin

and Stuart, 1985; Jasmer *et al.*, 1985; Feagin *et al.*, 1986; Feagin *et al.*, 1987; Feagin *et al.*, 1988; Koslowsky *et al.*, 1990; Read *et al.*, 1992b; Souza *et al.*, 1992). However, A6 was shown to be constitutively edited (Bhat *et al.*, 1990). This tight regulation of mt transcripts is very different in the genus *Leishmania,* where kDNA is constitutively expressed across both its lifecycle stages (Nebohácová *et al.*, 2009).

The BSF mitochondrion, being the site of RNA editing, makes the organelle an attractive target for drug discovery and design. Several of the enzymes involved in RNA editing, including RNA editing ligase 1 (REL1) show promise as a potential drug targets, since their inhibition caused cell death in BSF *T. brucei* following a loss of detectable fully edited mRNA (Schnaufer *et al.*, 2001; Wang *et al.*, 2003a; Trotter *et al.*, 2005; Carnes *et al.*, 2005; Salavati *et al.*, 2006; Law *et al.*, 2007; Babbarwal *et al.*, 2007; Tarun Jr. *et al.*, 2008; Guo *et al.*, 2008; Ernst *et al.*, 2009; Guo *et al.*, 2012; Lerch *et al.*, 2012; Carnes, *et al.*, 2012b).

In particular, the essential REL1 has been the focus for structural analyses, and subsequent virtual and compound screening with the intention of discovering an inhibitor. To date, naphthalene- and azo-dye-based inhibitors have had a suboptimal effect on the active site of the ligase, but important screening pipelines have been established (Amaro *et al.*, 2008a; Amaro *et al.*, 2008b; Durrant *et al.*, 2010). Such approaches in design of novel chemotherapeutics are important, since there are a limited number of drugs available and resistance has been described. To add to this, it is unclear how some readily available drugs act upon the parasite and there can be side effects to their administration (Williamson and Rollo, 1959; Damper and Patton, 1976; Frommel and Balber, 1987; Carter and Fairlamb, 1993; Carter *et al.*, 1995; de Koning *et al.*, 2000; Legros *et al.*, 2002; Matovu *et al.*, 2003; Croft *et al.*, 2005;

Balasegaram *et al.*, 2006; Steverding, 2010).

1.2.4 The evolution of RNA editing

It is thought that the unique and extensive U addition/deletion editing first arose in the bodonids, which are a class paraphyletic with Trypanosomatida within the phylum Euglenozoa. This would suggest that RNA editing is an ancestral process within kinetoplastids (Deschamps *et al.*, 2011). Since the bodonid, *Trypanoplasma borreli*, displays this same kind of deletion/insertion editing, this biological process was suggested to be between 500 and 700 million years old (Fernandes *et al.*, 1993; Lukeš *et al.*, 1994; Speijer, 2006). The need to modify mitochondrial transcripts post-transcriptionally can also be seen in diplonemids, which frequently use trans-splicing (for a comprehensive comparison see Lukeš *et al.*, 2005).

A possible linkage to lifecycle complexity and selection pressures acting upon the parasite to produce and maintain RNA pan editing is being debated, with positive and purifying pressures and neutral evolution being implicated. These theories can be separated by three themes: (i) editing on the way out, (ii) maintenance of kDNA and mt gene expression and (iii) protein diversification. Theories ii and iii can also be thought of as editing "on the way in" since they suggest that pan-edited evolved from never edited transcripts.

Editing on the way out (i) suggests that pan editing existed in the last common ancestor of the extant kinetoplastids, and over time during several instances pan edited genes were replaced with partially and never edited transcripts due to a selection pressure to reduce the numbers of mutations. In this model the never-edited

flanking regions of maxicircles facilitated homologous recombination in a manner that replaced sequences with reverse transcripts of more fully edited ones. This has been further linked to the loss of minicircle classes and dyskinetoplasty (Landweber and Gilbert, 1994; Maslov *et al.*, 1994). However, in some lineages minicircle classes remain diverse, which suggests that RNA editing has remained an important process. To add to this, dozens of nuclearly encoded and essential editosome proteins are required just to edit 12 mitochondrial transcripts in T. brucei (Lukeš *et al.*, 2009).

Maintenance of kDNA and mt gene expression (ii) suggests that pan RNA editing evolved in parasites with two hosts from a limited process when a pre-existing RNA machinery was already in place, to provide a proofreading system to counteract mutations and ensure mitochondrial gene expression when selection pressure is lax. In this way the more fragile mitochondrial kDNA is fragmented and tightly regulated, and thus its loss, when not in use for mt biogenesis, would be prevented. In this manner editing could either become fixed by genetic drift and subsequently become essential, or fragmentation of the kDNA genome could be a result of a positive selection pressure to prevent loss of mitochondrial subunit expression (Covello and Gray, 1993; Cavalier-Smith, 1997; Speijer, 2006). The presence of many overlapping gRNAs suggests that RNA editing may be prone to errors and a proofreading system is in place (Pollard *et al.*, 1990). This would provide a mechanism (through the exploitation of an already existing RNA processing machinery) to prevent the loss of genes that are not essential in the bloodstream form and, are therefore not expressed highly or essential, which would be required for the parasite to become tsetse fly infective (Lukeš *et al.*, 2009). The role of neutral evolution in RNA editing has been discussed at length by Lukeš and

colleagues. They have postulated that the correct RNA editing machinery was already in place before and the editing process co evolved with the gRNA templates, which were produced from gene duplication events aimed to neutralise mutations accumulated over time. This editing system would be biased against gRNA loss, and so, contrary to what had been discussed before, editing site would evolve alongside their gRNA templates in a manner that produced no real selective advantage, in a "unidirectional ratchet-like expansion" (Lukeš *et al.*, 2005; Lukeš *et al.*, 2009; Gray, 2012). This model constructive neutral evolution suggests that complexity in editing arises in the absence of positive selection (Gray, 2012).

Protein diversification (iii) encompasses the idea that two proteins can be coded from a single gene. RNA editing is governed by guide RNA templates. The heterogeneity of *T. brucei* minicircles reflects the extent and complexity of RNA editing in this organism (Sturm and Simpson, 1990a). Indeed the number of unique gRNAs exceeds that of known editing sites (Corell *et al.*, 1993; Hong and Simpson, 2003; Ochsenreiter *et al.*, 2007). In addition, alternative gRNAs for some mRNAs may result in alternative editing products, through the production of alternative reading frames, and potentially govern the diversity of proteins derived from RNA editing (Abraham *et al.*, 1988; Decker and Sollner-Webb, 1990). The most well studied alternatively edited protein is one that is derived from the COIII transcript, called Alternatively Edited Protein 1 (AEP-1). This protein is truncated compared to the full length protein, at 214 amino acids long, and contains five transmembrane domains. It is thought that this protein is responsible for kDNA maintenance and has a particular role in segregation. This is because it localises to the Tripartite Attachment Complex (TAC) between basal body and kinetoplast, and its

construction of an AEP-1 dominant-negative cell line leads to kinetoplast mis-segregation (Ochsenreiter and Hajduk, 2006; Ochsenreiter *et al.*, 2008). This evidence, however, is indirect since the protein could not be detected by mass spectrometry (Hajduk and Ochsenreiter, 2010). The theory of alternative RNA editing evolving as a mechanism to create further genetic diversity is not a new one, however, the idea of two proteins from one gene remains controversial. This is because it is thought that the potential losses from production of alternative proteins may outweigh the benefits of producing the mitochondrial subunits intended for translation (Landweber *et al.*, 1993; Ochsenreiter *et al.*, 2008; Lukeš *et al.*, 2009).

Although RNA editing has been shown to be widespread if not ubiquitous among kinetoplastids (as reviewed by Roy *et al.*, 2007; Lukes *et al.*, 2009) this thesis will focus on *T. brucei.*

1.3 The structure and function of the 20S core editosome

RNA editing is governed by nuclearly encoded and self-assembling multi-protein editing machines called editosomes of which there may be two distinct conformations, a stable 20S and a less stable 35-40S complex. In the trypanosome editing field, the S (Svedberg unit) typically refers to the sedimentation rate of editosome particles on a 10 – 30 % glycerol gradient (Pollard *et al.*, 1992; Corell *et al.*, 1996). A list of editosome components can be found in Table 1.2, and a protein interaction map can be found in Figure 1.6 A. Editosome components are nuclearly encoded, but enter the mitochondrion due to a targeting sequence on the protein, where they are assembled into functional complexes (Rusché *et al.*, 2001; Panigrahi *et al.*, 2001; Panigrahi *et al.*, 2003). Functional 20S editosomes can assemble within the mitochondria in the absence of any RNA substrate, as concluded from the study of chemically induced *T. brucei* and naturally occurring *T. evansi* Dk cells (Domingo *et al.*, 2003).

The core 20S editosome complex can be further organised into a deletion and an insertion subcomplex (Schnaufer *et al.*, 2003; Panigrahi *et al.*, 2006; Golas *et al.*, 2009). RNA editing ligase 1 (REL1), a U-specific exonuclease (ExoUase, KREX2) and kinetoplast RNA protein A2 (KREPA2) make up the deletion subcomplex, and RNA editing ligase 2 (REL2), a 3′ terminal uridylyl transferase (TUTase, KRET2) and kinetoplast RNA protein A1 (KREPA1) make up the insertion subcomplex (Palazzo *et al.*, 2003; Schnaufer *et al.*, 2003). The two subcomplexes were hypothesised to be joined by a substantial RNA substrate binding site, supposedly bridging the two complementary addition and deletion catalytic editing

Table 1.2. Protein Components of the 20S Editosome.

Current name	Former name(s)	Function	Motif
KREPA1	*Tb*MP81, LC-1, band II	Interaction	OB-fold zinc finger
KREPA2	*Tb*MP63, LC-4, band III	Interaction	OB-fold zinc finger
KREPA3	*Tb*MP42, LC-7b, band VI	Interaction	OB-fold zinc finger
KREPA4	*Tb*MP24, LC-10	Interaction	OB fold?
KREPA5	*Tb*MP19	Interaction	OB fold?
KREPA6	*Tb*MP18, LC-11, band VII	Interaction	OB fold
KREN1	*Tb*MP90, KREPB1	Deletion	RNase III, dsRBM, U1-like
KREPB1	*Tb*MP67	Endonuclease	RNase III, dsRBM, U1-like
KREN2	*Tb*MP61, LC-6a, KREPB3	Insertion	RNase III, dsRBM, U1-like
KREPB4	*Tb*MP46, LC-5	Interaction	RNase III, Pumilio, U1-like
KREPB5	*Tb*MP44, LC-8	Interaction	RNase III, Pumilio, U1-like
KREPB6	*Tb*MP49, LC-7c	Interaction	U1-like
KREPB7	*Tb*MP47	Interaction	U1-like
KREPB8	*Tb*MP41	Interaction	U1-like
KREL1	*Tb*MP52, LC-7a, band IV	Ligase	Ligase, tau, K
KREL2	*Tb*MP48, LC-9, band V	Ligase	Ligase, tau, K
KREX1	*Tb*MP100, LC-2	ExoUase	5’3’ exo,endo/exo/phos
KREX2	*Tb*MP99, LC-3, band I	ExoUase	5’3’ exo,endo/exo/phos
KRET2	*Tb*MP57, LC-6b	TUTase (editing)	NT, PAP-core, PAP-assoc
KREH1	*Tb*mHel61p	Helicase	Helicase

Editosomes consist of many catalytic and stabilising proteins, also important in the binding of ss and later dsRNA, endonucleolytic cleavage, U addition and deletion, and 5’ – 3’ ligation.

Reproduced from Carnes and Stuart, (2008a).

activities (Göringer, 2012). The assembly of the editosome complex has revealed two distinct subcomplex positions of REL1 and REL2, inferring a differential role for these ligases (which will be discussed later in Section 1.3.5). The manner of organisation also allows certain catalytic activities to be confined to the editosome that would normally be detrimental to the cell if left unconstrained (Stuart *et al.*, 2005).

It is likely that all components of the editosome machinery may have originated from RNA repair enzymes that existed in an RNA-protein world which predated a DNA world, since the protein machinery would be capable of proofreading and repairing RNA molecules (Ho *et al.*, 2004).

1.3.1 Core interactive proteins

Kinetoplast RNA editing proteins (KREP)A proteins are key interactive proteins responsible for editosome integrity. Several studies have suggested that the disruption of KREPA proteins result in the loss of functioning editing complexes and disruption of editing through loss of editosome integrity. This was revealed through a shift in editosome size, or a repression of endonucleolytic cleavage and subsequent RNA editing events (Drozdz *et al.*, 2002; Huang *et al.*, 2002; Salavati *et al.*, 2006; Law *et al.*, 2007; Tarun Jr. *et al.*, 2008; Law *et al.*, 2008; Guo *et al.*, 2008). Some KREPA proteins contain a zinc finger domain used for the binding and recognition of RNA and all contain an interactive oligonucleotide/oligosaccharide binding (OB fold) domain responsible for editosome integrity through protein-protein binding (Schnaufer *et al.*, 2003; Schnaufer *et al.*, 2010; Park *et al.*, 2012b). If either of these

domains are interrupted in KREPA3 the editosome integrity is compromised, resulting in a loss of complexes and a complete disruption of editing in BSFs and partial loss of editosomes in PCF, although this may be a function of knock-down efficiency, rather than function (Guo *et al.*, 2008). KREPA3 may also exhibit some catalytic activity, is involved in gRNA-mRNA processing and is capable of excising uridylyls *in vitro*, although this remains a controversial viewpoint as no recognisable catalytic motif has been found and regular deletion activity remains in its absence (Brecht *et al.*, 2005; Law *et al.*, 2008; Guo *et al.*, 2008; Niemann *et al.*, 2009). This capacity to trim uridylyls is useful, to allow for proofreading, as U-insertion editing involves the addition of more residues than necessary by the corresponding TUTase (Byrne *et al.*, 1996).

The interactions between stabilising and catalytic proteins, inferred from yeast two-hybrid data and co-expression data using tagged recombinant proteins *in vitro,* have been elucidated by Schnaufer *et al* in 2010, giving a more detailed structure of the editosome, see Figure 1.6. It has also been suggested that these editosome components interact due to their complementary electrostatic properties (Shaneh and Salavati, 2010). Since the study by Schnaufer *et al.*, 2010, a KREPA3-KREPA6 hetero-dimer has been visualised by crystallography. This trans-tetramer model adds further weight to the notion that the editosome contains a core made up of OB folds, which is important in maintaining the overall structure and for binding double stranded (ds) RNA substrates of editing (Park, *et al.*, 2012a; Park *et al.*, 2012b).

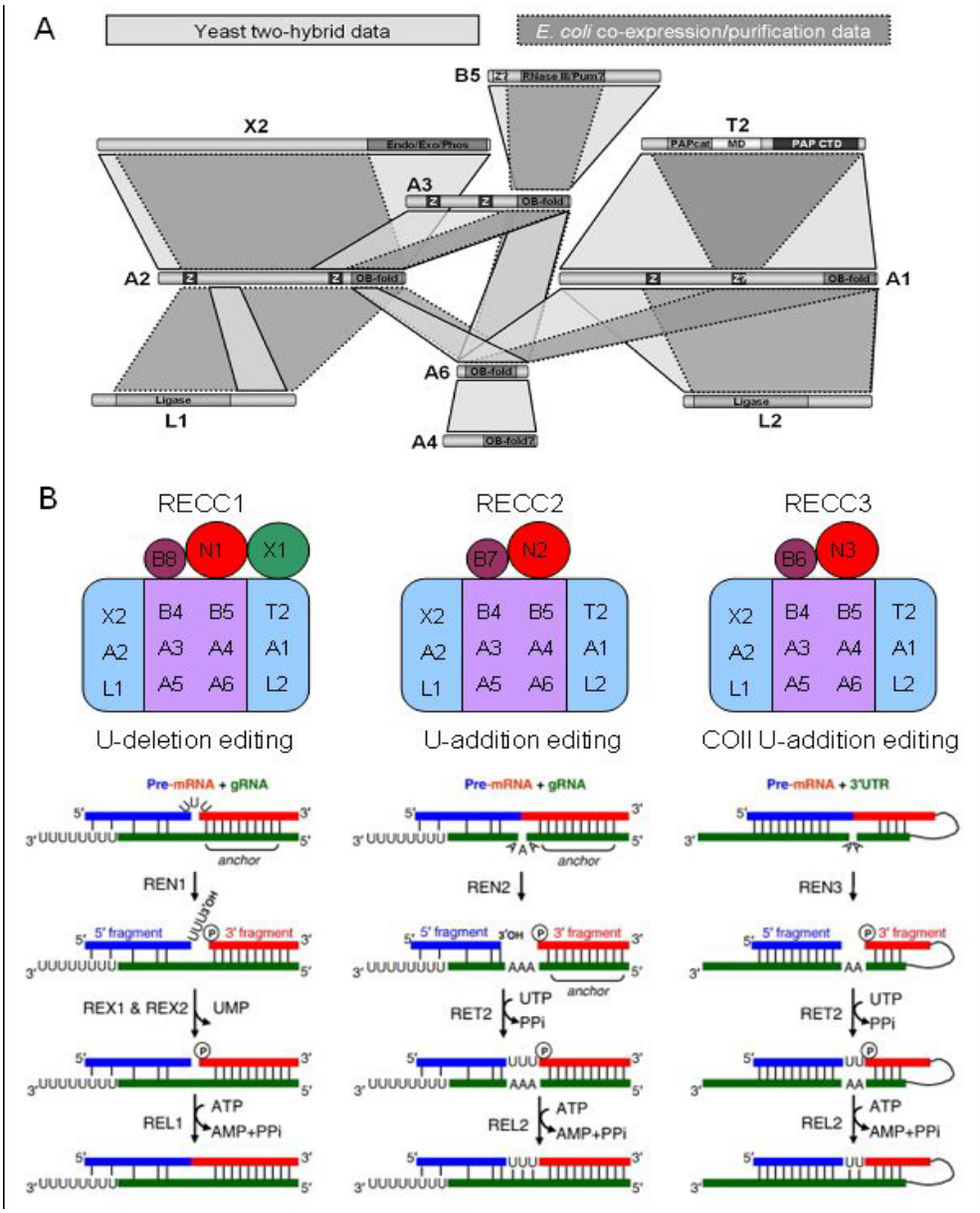


Figure 1.6. Protein Interactions of the 20S Editosome, and the role in editing for different types of editosome.

A) The 20S editosome contains an addition subcomplex (right), containing L2 ligase, A1 protein and T2 TUTase, and a deletion subcomplex (left), containing L1 ligase, A2 protein and X2 exoUase. Taken from Schnaufer *et al.*, 2010. B) All 20S editosomes contain the same core set of proteins, but differ in composition by the B proteins (B4 -8), endonucleases (N1 – 3) and exoUases (X1 and X2) that associate with them, giving rise to 3 different types of editosomes. Each is responsible for governance of three different kinds of RNA editing. RECC1 (left), RECC2 (middle) and RECC3 (right) are involved in deletion, insertion and COII insertion editing respectively. COII insertion (by RECC3) differs from regular insertion editing as the gRNA acts in *cis* and not *trans*. Adapted from Ernst *et al.*, 2009 and Ringpis *et al.*, 2010.

1.3.2 The endonucleases and KREPB accessory proteins

Kinetoplastid RNA editing endonucleases (KRENs) and kinetoplastid RNA editing B proteins (KREPBs) associate in distinct pairs with a common set of core proteins to create three different kinds of editosomes, or RNA editing core complexes (RECC 1 - 3). Each kind governs a different kind of editing event (see Figure 1.6 B). Repression of KREN1 (formerly known as KREPB1) caused a specific reduction in deletion editing, whilst KREN3 (formerly known as KREPB2) ablation caused a 40% reduction in edited COII transcripts, indicating a specific role in editing of this transcript (Trotter *et al.*, 2005; Carnes *et al.*, 2008b; Ernst *et al.*, 2009; Carnes *et al,* 2012a). KREN2 (formerly known as KREPB3) knockdown causes a general growth phenotype affects insertion editing events and causes an accumulation of unedited and a reduction in edited transcripts (Carnes *et al.*, 2005). KREPB6, KREPB7 and KREPB8 associate with KREN3, KREN2 and KREN1, respectively, and are thought of as accessory proteins since they associate with the core complex, to give different functional properties (Guo *et al.*, 2012). The KRENs associate physically with the insertion subcomplex, whereas the KREPB proteins associate directly, or indirectly, with the deletion subcomplex components, as indicated by their weak involvement in precleaved addition based editing assays. The specificity of these three different editosomes drives substrate recognition (Guo *et al.*, 2012). The KREPB6 - B10 proteins may drive this specificity by allowing the adaptation of their associated endonucleases to particular substrates (Lerch *et al.*, 2012).

1.3.3 The TUTases

There are two distinct RNA editing TUTases (RET1 and RET2) within the *T. brucei* editosome that are involved in editing (Aphasizhev *et al.*, 2002). Stability of the editosome is not affected by TUTase down regulation. RET2 is a 3' uridylyl transferase, responsible for a single U base addition and exists within the addition subcomplex (Aphasizhev *et al.*, 2003). RET1, on the other hand, is responsible for the addition of poly(U) tracts (or ladders) onto gRNA, and gives stability to the transcript. This was apparent as RET1 RNAi caused a decrease in steady state mRNAs without disrupting transcription (Aphasizhev *et al.*, 2002). RET1 and RET2 differ in their properties and essentiality. The down regulation of RET2 leads to the complete inhibition of addition editing *in vitro,* without affecting deletion editing, and also growth inhibition after 80 hours induction (Aphasizhev *et al.*, 2002; Aphasizhev *et al.*, 2003; Aphasizheva *et al.*, 2009). RET1 RNAi, on the other hand, has no effect on gRNA U tail addition with respect to deletion editing and has little effect on *in vitro* U-insertion editing. This suggests there is a division of labour between RET1 and 2 (Aphasizhev *et al.*, 2002; Ernst *et al.*, 2003; Aphasizhev *et al.*, 2003; Ringpis *et al.*, 2010). RET2 exists as a single copy in the editosome, and is bound by its middle region to KREPA1 (Fig. 1.6) (Schnaufer *et al.*, 2010; Ringpis *et al.*, 2010).

1.3.4 The exoUases

Uridylyl-specific editing exonucleases (3'– 5' ExoUases) catalyse the removal of a single non-base-paired uridylyl at a time and are inhibited by base paired uridylyls. Within the editosome there are two exoUases: KREX1 and KREX2, which have U-specific excising activities (Kang *et al.*, 2005; Rogers *et al.*, 2007; Ernst *et al.*, 2009). The knock-down of KREX1 and KREX2 has differential effects on the cell. Whilst KREX2 RNAi and ablation caused no discernible growth phenotype in BSFs and a slight growth impediment in PCFs, both deletion and addition editing activities were subtly reduced and there was a size decrease in functioning editosomes. KREX1 RNAi, however, caused a sizable decrease in deletion editing activities and specifically caused the loss of KREN1 editosomes (Kang *et al.*, 2005; Rogers *et al.*, 2007; Ernst *et al.*, 2009; Carnes *et al.*, 2012c). This may suggest that KREX1 and 2 exhibit division of labour, where KREX1 serves to remove uridylyls during deletion editing, whilst KREX2 may remove the excess uridylyls at insertion editing sites, and in doing so functions as a proofreading enzyme (Ernst *et al.*, 2009; Carnes *et al.*, 2012c).

1.3.5 REL1 vs. REL2: a tale of two ligases

The final process of editing in trypanosomes involves ligation of RNA substrates and is performed by two RNA editing ligases, REL1 and REL2 (Figures 1.6 and 1.7). Ligation itself is a multi-step process involving ATP hydrolysis with covalent binding of AMP to a lysine residue in the ligase active site, transfer of AMP to the 5' phosphate of the 3' substrate, and, finally, formation of a phosphodiester bond

between the 5' phosphate and 3' hydroxyl termini of the two RNA strands. RNA termini may be joined in the presence of a complementary gRNA template strand (Cruz-Reyes and Sollner-Webb, 1996; Blanc *et al.*, 1999; Odell *et al.*, 2000; Palazzo *et al.*, 2003; Ho *et al.*, 2004; Nandakumar *et al.*, 2004).

REL1 and REL2 are most closely related to the superfamily of covalent nucleotide transferases, and share their RNA joining properties (Ho *et al.*, 2004). The most closely related enzyme to REL1 is T4 RNA Ligase 2 (T4Rnl2), which was established as a nucleotide transferase as it contained the motifs I, III, IIIa, IV and V; the latter two of these are essential to the activity of the ligase (Ho and Shuman, 2002). T4Rnl2-like ligases are more widespread in other organisms than T4Rnl1-like ligases, and these seal dsRNA breaks *in vitro* (Ho and Shuman, 2002; Wang *et al.*, 2003b). The *in vivo* function of T4Rnl2 is unknown, but it has previously been suggested that it is involved in RNA editing, RNA repair and the capping of dephosphorylated RNA ends (Ho and Shuman, 2002; Yin *et al.*, 2003).

Ligases that seal dsRNA breaks, such as REL1, REL2 and T4Rnl2 most likely originated as general RNA repair enzymes in the presence of protein replicating machinery, before DNA existed, as RNA repair of this type is very uncommon in newly arising metabolic pathways (Ho *et al.*, 2004; Chan *et al.*, 2009). The proto-ligase in this scenario would be non-specific and would only require terminal phosphates for ligation. Selectivity for RNA substrates could have happened with a domain swap, or change to the C-terminal domain (Nandakumar and Shuman, 2004). The presence of small gRNA-like molecules in a proto-editosome would have also allowed the evolution of more complex dsRNA molecules (Cheng and Unrau, 2010).

In *T. brucei* REL1 and REL2 are embedded within the editosome, and occupy the same biological niche, which is unlike the scenario for T4Rnl1 and T4Rnl2 (Ho and Shuman, 2002; Schnaufer *et al.*, 2003). REL1 and REL2 are 52 and 48 kDa in size, respectively, localise to the mitochondrion and contain a KXXG active site motif (McManus *et al.*, 2001; Schnaufer *et al.*, 2001). REL2 is structurally similar to REL1, (41% sequence identity and 61% similarity) and both ligases are more similar to T4RNL2 than to any other RNA ligase (Shaneh and Salavati, 2010). Relevant sequence alignments and the crystal structures of REL1 and T4Rnl2, complete with their adenosine substrates are shown in Figure 1.7 overleaf. The crystal structure of REL1 shows that at the active site the α-phosphate of ATP is stabilised within the binding pocket and that its adenylylation is dependent on the presence of divalent magnesium (Deng *et al.*, 2004).

The catalytic N-terminal catalytic domain (CD) is required for the RELs autoadenylylation activity. However, if the N-terminal CD is expressed recombinantly without its interaction domain (ID), it appears to be less active *in vitro* than full length REL1 (Deng *et al.*, 2004). The C-terminal ID of REL1 is required for integration into native editing complexes, via direct interaction with KREPA2, and does not contain any catalytic sites (Schnaufer *et al.*, 2003; Gao *et al.*, 2005; Schnaufer *et al.*, 2010). The C-terminal ID of T4Rnl2 is essential for strand sealing specificity, because it is required for substrate binding (Ho *et al.*, 2004). Therefore, T4Rnl2 and the RELs consist of two domains: a C-terminal ID and an N-terminal CD. In the RELs these domains are responsible for protein-protein interaction and for strand sealing activities, respectively. Where T4Rnl2 and the RELs greatly differ is in their ID (Schnaufer *et al.*, 2003). REL1 and REL2 differ from most other DNA or

RNA ligases as they do not have a separate OB-fold domain, rather this is provided in *trans* by the REL's interaction partners within the editosome (Schnaufer *et al.*, 2003; Worthey *et al.*, 2003; Deng *et al.*, 2004).

It is clear from gene knockdown and knockout studies that REL1 is essential to the cell, whereas REL2 RNAi (despite an efficient knockdown) does not induce a growth phenotype (Huang *et al.*, 2001; Rusché *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz *et al.*, 2002; Gao and Simpson, 2003; O'Hearn *et al.*, 2003). However, it should be noted that a subtle cell deformation was reported for one RNAi study (O'Hearn *et al.*, 2003). Ablation of REL1 in BSFs causes loss of fully edited mt transcripts within 46 hours and cessation of cell division within 70 hours (Schnaufer *et al.*, 2001). REL1 was also shown to be essential in PCFs as no null mutant could be created (Rusché *et al.*, 2001).

The reason why REL1 is essential, but REL2 appears not to be, is not clear, but differences in their RNA substrate specificity and ATP affinity may lend some ideas, since the REL's appear to have distinct ligation activities. The adenylylation reaction governed by these ligases differs corresponding to their differences in affinity for phosphate and ATP, in that REL2 has a higher affinity for ATP than REL1 (Rusché *et al.*, 2001; Cruz-Reyes *et al.*, 2002; Palazzo *et al.*, 2003).

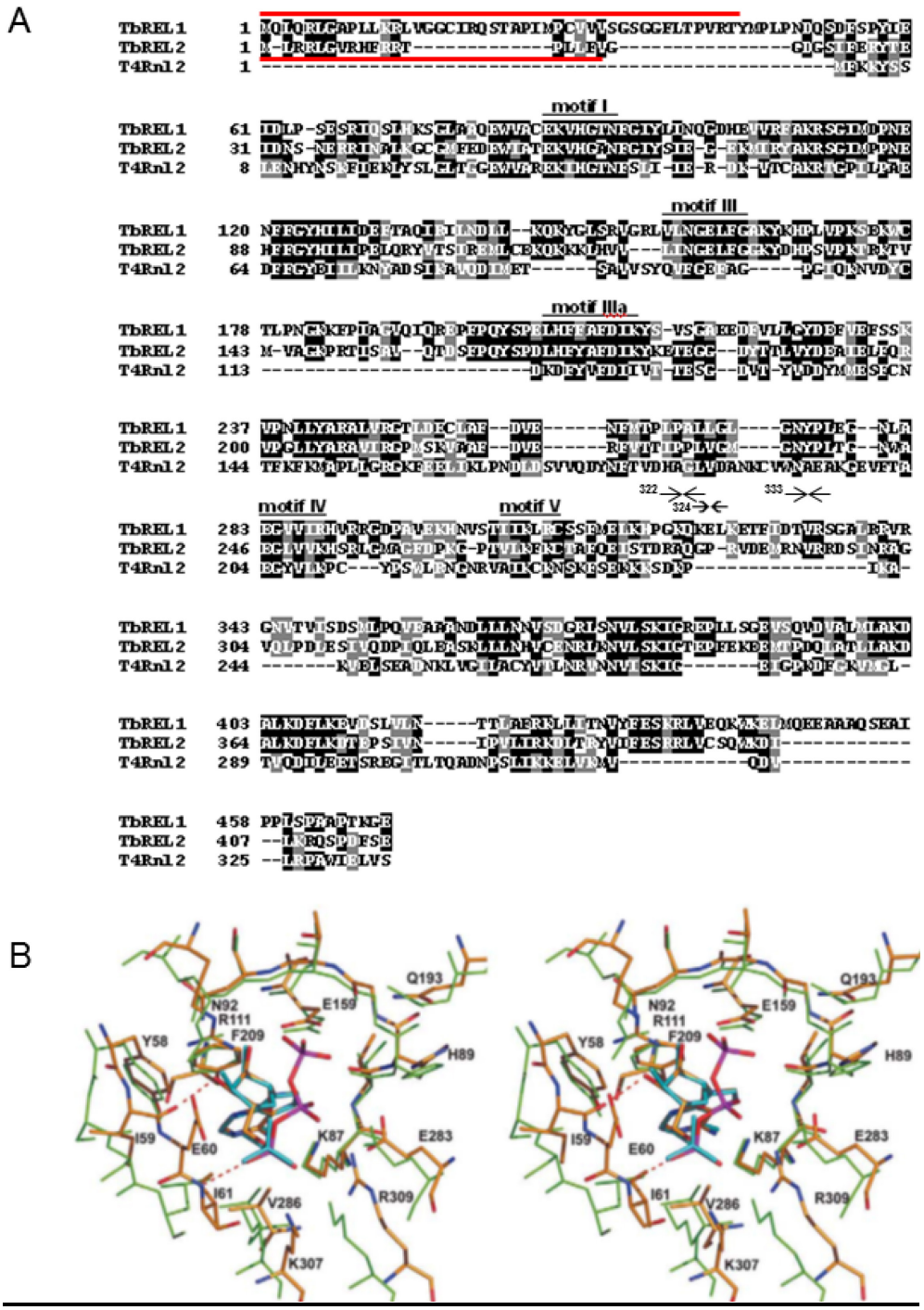


Figure 1.7. Sequences and crystal structure of the RNA Editing Ligases.

A) Sequence alignment of the *T. brucei* RELs, highlighting the important 5 motifs (lined) present in the nucleotidyl transferases. Arrows indicate the CD-ID fusions used in Chapter 2. Red lines indicate the mt targeting. B) The crystal structures of the ligases, highlighting the essential amino acid residues responsible for their catalytic properties. Left: REL1-ATP, right: T4Rnl2-AMP (Taken from Deng *et al.*, 2004).

REL1 is also less stringent about what substrates it will ligate, annealing RNA fragments with both overhangs and nicks. REL2, however, is more stringent in its activities and is restricted to the ligation of perfectly nicked duplexes, although it appears that both RELs, as with DNA ligases, have a preferential ligation for a perfect nicked duplex (Blanc *et al.*, 1999; Igo *et al.*, 2000; Rusché *et al.*, 2001; Schnaufer *et al.*, 2001; Cruz-Reyes *et al.*, 2002; Palazzo *et al.*, 2003).

In the aforementioned scenarios REL1 can compensate for the loss of REL2, but REL2 cannot compensate for the loss of REL1 (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Gao and Simpson, 2003). This questions the need for two ligases. Tandem Affinity Purification (TAP) tag purification and yeast-2-hybrid analysis of the two ligases put them into two separate subcomplexes of the editosome (as mentioned previously in Section 1.3), with REL1 in the deletion subcomplex and REL2 in the addition subcomplex (Schnaufer *et al.*, 2003). Due to this subcomplex positioning, it could be assumed that REL1 and REL2 are involved in ligation of deletion and addition editing substrates, respectively. Gao and Simpson addressed this question by monitoring the state of edited RNA after REL1 and REL2 RNAi knockdown. Down-regulation of REL2 had little effect on the abundance of any of the edited transcripts studied, including COII, which only contains 4 addition, and no deletion editing sites. In contrast, down regulation of REL1 greatly affected transcripts involved in addition (Cyb, ND7) deletion editing (ND7), but had little or no effect on COII editing. This led to the conclusion that REL2 may be less active, or inactive *in vivo,* and that the two ligases have different biological roles (Gao and Simpson, 2003).

Knockdown or knockout of REL1 to determine function and specificity is further complicated by the observation that in some cases the editosome becomes

less stable in this absence (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002). However, this is contradictory to some other findings, although residual REL1 from genetic manipulation may confound the true effects of its loss on editosome integrity (Stuart *et al.*, 2002). Expression of a catalytically dead REL1 enzyme suggests division of labour between the two RELs, and their separate involvement in deletion and insertion editing (Huang *et al.*, 2001). The most attractive hypothesis to the essentially of REL1 does not pertain to its roles in deletion or addition editing, but instead suggests that this ligase has a role in the repair of erroneous cleaved substrates (Huang *et al.*, 2001).

1.3.6 Accessory complexes in *T. brucei*

For RNA editing to occur, there must be simultaneous processing for gRNAs and pre-mRNAs (since maxicircles and minicircles are transcribed as monocistrons), suggesting a role for accessory factors in allowing efficient editing of the transcripts (Reifur *et al.*, 2010). Mitochondrial RNA precursor processing endonuclease (mRPN) is involved in the maturation of polycistronic pre-gRNAs to monocistronic gRNAs (Grams *et al.*, 2000; Madina *et al.*, 2011). The presence of RNA editing helicase 1 (REH1) increases editing efficiency in the presence of multiple gRNAs, by aiding 3' - 5' gRNA detachment (Li *et al.*, 2011).

There are a number of proteins involved in editing outside the core editosome complex, responsible for stabilising the gRNA-mRNA duplexes and shuttling gRNAs and mRNAs to the RECC to undergo editing. TbRGG1 (which is named as such because it contains the RGG RNA binding domain) is equally present in both

PCFs and BSFs, sedimenting at 35 - 40S. Its ablation causes disruption in the regulation of mRNA editing, but does not affect never edited transcripts (Vanhamme *et al.*, 1998; Hashimi *et al.*, 2008). TbRGG2 contains a C-terminal RNA recognition motif, and its knockdown affects pan edited transcripts only (Fisk *et al.*, 2008; Ammerman *et al.*, 2012; Foda *et al.*, 2012).

Both the Mitochondrial RNA binding protein (MRP1/2) heterodimer and Y box RNA binding protein of 16 KDa (RBP16) are essential to editing and have RNA-RNA annealing properties. Simultaneous knock down of MRP1/2 and RBP16 causes a growth phenotype in PCFs, but not BSFs, without affecting gRNA abundance. This may be reflected in the quantity of transcripts that are required to be edited at each lifecycle stage for mitochondrial biogenesis. It is thought that RBP16 helps reveal the mRNA anchor sequence, and at the same time the MRP1/2 heterodimer exposes the gRNA anchor sequence. Both are essential for editing specific mRNAs as they promote RNA-RNA annealing activities (Hayman and Read, 1999; Schumacher *et al.*, 2006; Ammerman *et al.*, 2008; Fisk *et al.*, 2009). RBP16 binds the poly U tails of gRNAs, acting to stabilise mRNA and promote gRNA-mRNA interactions (Pelletier and Read, 2003).

The mitochondrial RNA binding complex (MRB1) is not stably associated with the 20S editosome, rather it associates with the RECC *in vivo* via dynamic RNA interactions, and has an indirect effect on editing (Domingo *et al.*, 2003; Ammerman *et al.*, 2012). Down regulation of one of its sentential components, TbRGGm, leads to smaller editosome complexes (through loss of components), abnormal kDNA division and a skewing of mt RNA populations (Acestor *et al.*, 2009). This is an interesting observation, since Dk cells possess smaller MRB1 complexes, which is

caused by a loss of maxicircle DNA (Schnaufer *et al.*, 2002; Acestor *et al.*, 2009). This MRB1 complex binds gRNA through the stabilising actions of gRNA associated proteins 1 and 2 (GAP1/2) and is involved in extensive or pan editing (Fisk *et al.*, 2008; Acestor *et al.*, 2009; Ammerman *et al.*, 2010; Ammerman *et al.*, 2012). RNA stability is governed by a number of other factors like mitochondrial editing mRNA stability factor 1 (MERS1) and RNA editing helicase 2 (REH2) (Weng *et al.*, 2008; Hashimi *et al.*, 2009; Hernandez *et al.*, 2010).

There are many examples listed here and elsewhere of accessory complexes having an indirect effect on RNA editing on ablation, and sedimenting at 35 - 40S on glycerol gradients. Altogether (see Figure 1.8) this suggests the presence of a large and dynamic complex responsible for processing and shuttling gRNAs and pre-mRNAs to the 20S catalytic core complex for editing (Göringer, 2012).

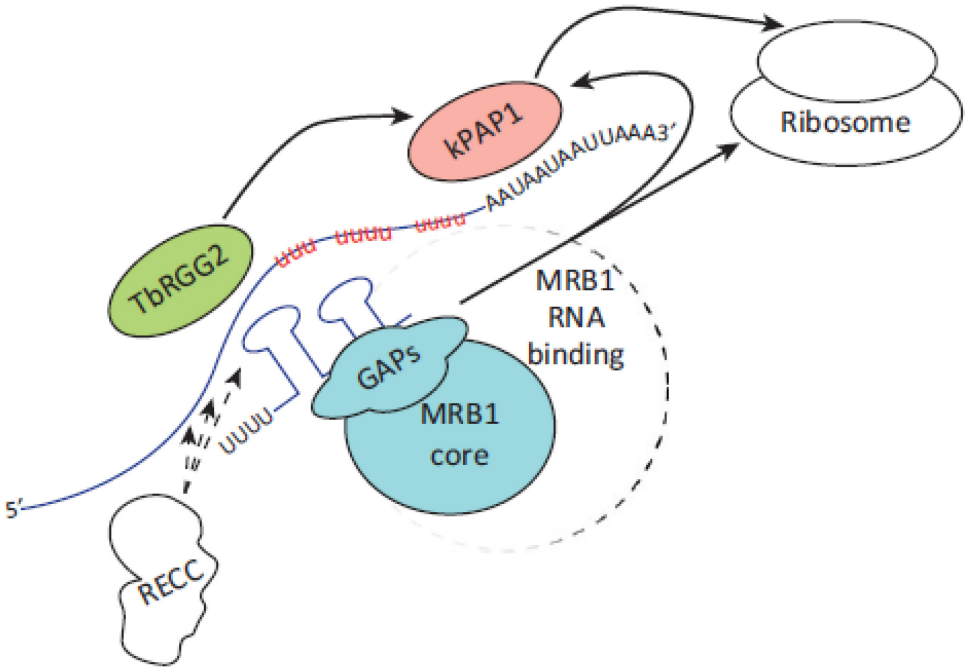


Figure 1.8. Proposed interactions between accessory complexes.

In this model, which focuses closely on the MRB1 complex, the gRNAs are brought in toward MRB1. At the same time TbRGG2 promotes gRNA-mRNA annealing. The RECC, or editosome, associates with these complexes via the RNA it edits. The finished transcripts are them deemed to be translational competent after the addition of a long A/U tail by the KPAP1 complex, and can be shuttled to the mitoribosome.

Abbreviations: RECC – RNA editing core complex (20S editosome), TbRGG2 – protein 2 with RNA RGG binding motif, KPAP1 - kinetoplast poly(A) polymerase 1, GAPs – gRNA accessory proteins, MRB1 - mitochondrial RNA binding complex 1

Reproduced from Hashimi *et al.*, 2013.

1.4 RNA maturation and translation

There is a strong association between RNA editing, stability, maturation and translation (Figure 1.9). Poly(A) tails have been shown to be developmentally regulated *cis*-elements that stabilise mRNA or promote its decay (Bhat *et al.*, 1992; Read *et al.*, 1992b; Read *et al.*, 1994a; Militello and Read, 1999). Further investigation of RNA populations revealed that transcripts with short poly(A) tails were unedited transcripts, whereas populations with a mixed A tail length were editing intermediates. It was also suggested from the same study that long A tails were required for stability (Militello and Read, 1999).

Short poly(A) tails are added onto mt transcripts through the actions of a kinetoplast ploy(A) polymerase (KPAP1), which allows the transcripts to maintain *cis*-stability in all stages of editing. KPAP1 is localised to the two antipodal regions of the kDNA disc and is essential to both BSF and PCF parasites (Etheridge *et al.*, 2008). In addition, the kinetoplast polyadenylation and uridylylation factor hetero-dimer (containing KPAF1 and 2) promotes the addition of long poly(A/U) tails through the actions of KPAP1 and RET1. KPAF proteins contain pentatricopeptide (PPR) repeat sequence motifs, which are also responsible for stabilising 12S and 9S rRNAs. These long A/U tracts were not found flanking COI or Cyb transcripts in BSFs, which are not edited in this lifecycle stage. Furthermore, transcripts with long A/U tails and proteins with PPR repeats localise to the mitoribosomes, indicating that such transcripts are translation competent (Pusnik *et al.*, 2007; Aphasizheva *et al.*, 2011).

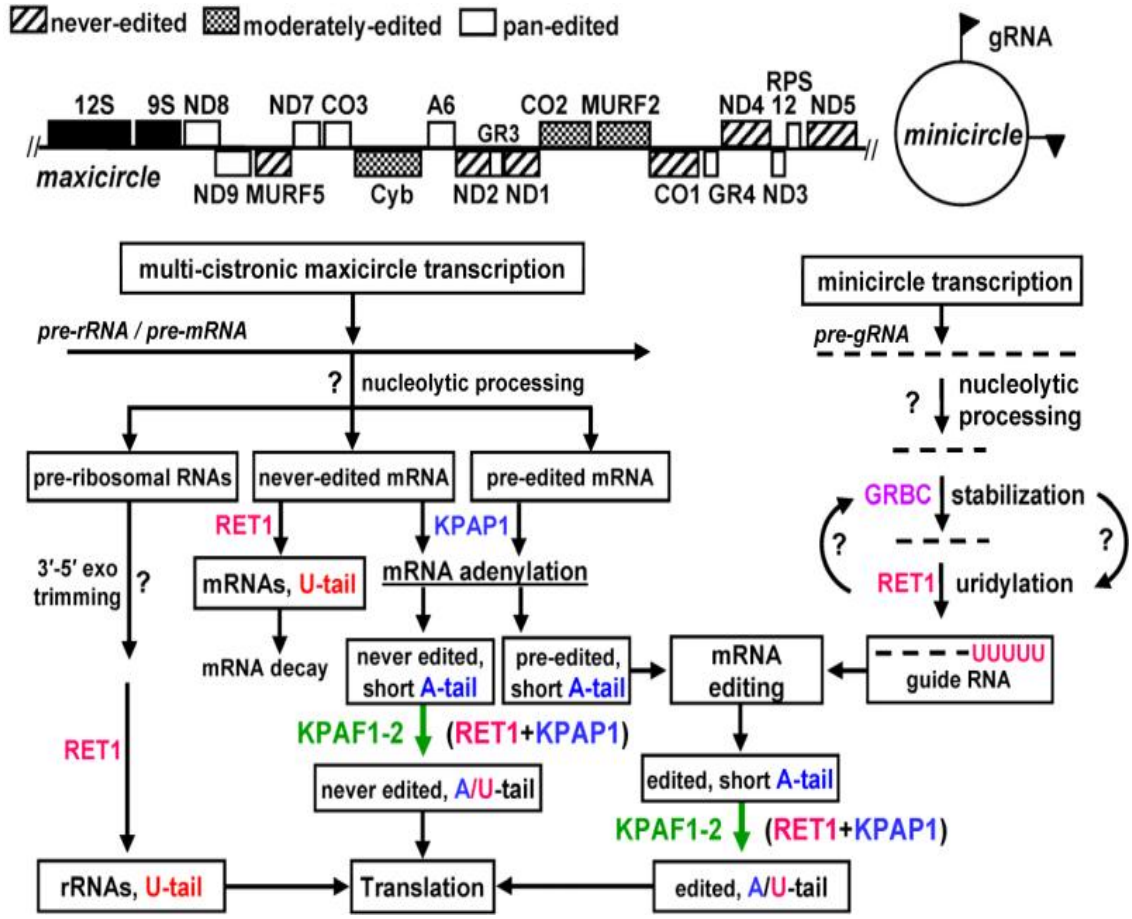


Figure 1.9. Schematic of mitochondrial RNA processing in trypanosomes.

Guide RNA and mt RNA maturation and processing are closely linked. This diagram highlights the complexes associated with RNA maturation and indicates the fate of never edited and edited transcripts.

Abbreviations: RET1 - RNA editing TUTase, KPAP1 - kinetoplast poly(A) polymerase 1, KPAF1/2 - kinetoplast polyadenylation and uridylylation factor 1/2, GRBC - gRNA binding complex (referred to in the text as MRB1 - mitochondrial RNA binding complex).

Reproduced from Aphasizhev and Aphasizheva, 2011.

It is not currently known how these PPRs recruit tRNAs or mitoribosomes, or how they aid AU tail formation, or any other aspect of translation. Degradation of transcripts is independent of this long A/U tract (Ryan *et al.*, 2003). However, little is known about the degradation products in trypanosomes. It is thought that this process is also inhibited by the secondary and tertiary structures of the transcripts, as well as their stage specific regulation (Ryan *et al.*, 2003).

The mitochondrial genomes of kinetoplastids lack tRNA genes, so all are imported from the cytosol after they are nuclearly expressed (Hancock and Hajduk, 1990). Trypanosomatids lack a bacterial $tRNA^{met}$ initiator of translation, and instead utilise a tRNA formyltransferase after import to formylate $tRNA^{met}$ allowing the recognition of initiation factor 2. A single RNA editing event (CCA – UCA) then allows the imported tRNA to decode mitochondrial transcripts. The large number of proteins in mitochondrial ribosomes in kinetoplastids is thought to compensate for the relatively short rRNAs they have in comparison to other eukaryotes (as reviewed by Schneider, 2011; Niemann *et al.*, 2011).

1.5 Context of the PhD objectives

We wished to discern why REL1 is essential to cell growth and viability, but REL2 appears not to be. We hypothesised that in addition to its function in resealing fully edited mRNA after U deletion, and perhaps U addition, REL1 is required for repair of erroneously cleaved mRNAs. Due to its more constrained substrate requirements, REL2 would not be able to compensate for loss of this activity, pinning any RNA repair function solely on REL1.

To help determine this, we wanted to determine whether it was the catalytic properties, or its position within the deletion subcomplex that made REL1 essential. Either REL1 is essential because of a particular catalytic functionality, i.e. in repairing miscleaved RNAs during editing, or REL1 is essential because of its physical location in the editosome. These hypotheses are not mutually exclusive, and REL1 may be essential for both of these reasons.

Preliminary studies by Matthew Spencer had indicated that only ectopic expression of an additional copy of REL1, but not of REL2 or of chimeric REL proteins, can fully rescue the growth phenotype produced upon REL1 ablation in conditional knockout (cKO) lines.

Sequencing the 5’ ends of RNA editing intermediates after genetic ablation of REL1 would give information on whether the RNA fragments produced would be products of addition, deletion or misediting.

To address the conundrum of REL2’s apparent redundancy, even though it is catalytically active, we carried out an evolutionary analysis. If REL2 has no essential role in editing, or to the cell in general, we would expect the ligase to be neutrally

evolving. Due to the known essentiality of REL1 we would expect this enzyme to be under strong purifying selection, which would reduce the number of deleterious mutations acquired over evolutionary time.

1.5.1 Research Objectives

Firstly, we used a molecular biology approach to generate and purify editing complexes via tagged chimeric ligases in order to dissect the respective contributions of position within the editosome *versus* substrate specificity to the essential role of REL1 in editing. This experiment was also expected to shed light on how REL1 compensates for the loss of REL2 in insertion editing. REL1 and REL2 have distinct domains for catalysis and interaction with their associated partner proteins in the editosome respectively. The ectopically expressed copies of REL1, REL2 and chimeric REL proteins were TAP-tagged, allowing purification and analysis of the complexes these proteins associate with. These experiments also allowed investigation into whether the chimeric proteins can associate with the predicted subcomplexes and, if so, how position and catalytic properties affected function (see Figure 2.2 for schematics of the TAP tagged "rescue" ligases).

Secondly, another related aim was to undertake a comprehensive identification of *in vivo* REL1 substrates by determining 5' ends of mitochondrially derived transcripts before and after REL1 inactivation, taking advantage of the existing REL1 conditional knock-out cell line. Through the development of a novel 5' end mapping and RNA sequencing approach, we hoped to reveal those transcripts that remain unligated after endonucleolytic cleavage, to identify the precise cleavage

sites, and to determine downstream editing events quantitatively.

Thirdly, RNAi studies have shown that knockdown of REL2 does not cause any growth effects in BSF *T. brucei* (Drozdz *et al.*, 2002; Gao and Simpson, 2003; O'Hearn *et al.*, 2003). However, given the uncertainties associated with the incomplete gene inactivation in RNAi, one of the aims of this study was to attempt to create a null mutant of REL2. In addition to clarifying whether REL2 function is indeed completely redundant, generation of such a cell line will also allow the complementation study outlined above to be conducted in a more definitive manner.

The final aim was to investigate essentiality of REL1 and REL2 by means of an evolutionary analysis. This involved separate comparison of REL1, REL2 and an interactive KREPA protein from different species of closely related trypanosomatids, with a means to determine whether the proteins are under positive or purifying selection, or whether they are neutrally evolving. The relative type and strength of selective pressures were expected to indicate whether REL2 may still have a function within – or independent of - the editosome that RNAi has not revealed.

1.5.2 Research Questions

This research project aimed to answer the following separate questions pertaining to RNA editing, focusing on the specific role of REL1 ligase, within trypanosome mitochondrial biology:

1 Why is REL1 essential in editing, and REL2 is not? In particular, is it the catalytic properties and/or physical positioning within the deletion subcomplex of the editosome that makes it essential to editing?

2 What are the substrates of REL1? Can identification of those RNA substrates that remain unligated when REL1 is ablated help determine its precise function?

3 Can the essentiality of REL1 and suggested redundancy of REL2 be confirmed using an evolutionary approach?

Chapter 2

A genetic complementation approach to understand why REL1 is essential

2.1 Introduction to project

It is not fully understood why REL1 is essential and REL2 is not, since both insertion and deletion of uridylyls are required for accurate editing of pan edited transcripts, and both enzymes are closely related to each other and the RNA repair enzyme, T4Rnl2 (Ho and Shuman, 2002; Ho *et al.*, 2004; Deng *et al.*, 2004). Unpublished studies by Achim Schnaufer have indicated through the overexpression of catalytically inactive ligases that REL1 is essential to the cell, but REL2 does not appear to be. This adds weight to the published literature (as discussed in Chapter 1) that REL1 is essential, but REL2 appears to be dispensable to the cell (Huang *et al.*, 2001; Rusché *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz, *et al.*, 2002; O'Hearn *et al.*, 2003).

The apparent non-essentiality of REL2 in the cell, has been taken as evidence that either the ligase is not active *in vivo*, or that REL1 may also function in addition editing reactions, collectively suggesting that REL2 is non-essential to the RNA editing process (Drozdz *et al.*, 2002; Gao and Simpson, 2003; O'Hearn *et al.*, 2003).

Other studies have suggested that both REL1 and REL2 have their distinct roles in editing. The subcomplex division of the ligases may suggest that they have division of labour within editing, and indeed REL1 and REL2 display distinct catalytic properties *in vitro,* pertaining to deletion or insertion editing, respectively (Cruz-Reyes *et al.*, 2002; Schnaufer *et al.*, 2003). Even though both ligases prefer perfectly nicked duplexes, biochemically, REL1 and REL2 have distinct properties, and perform differential roles in deletion and addition editing ligation, respectively (Blanc *et al.*, 1999; Igo *et al.*, 2000; Cruz-Reyes *et al.*, 2002; Palazzo *et al.*, 2003;

Schnaufer *et al.*, 2003). It is known that the substrate requirements of REL1 are less specific and more relaxed than those of REL2, which has strong preference for fully base paired RNA duplexes (Cruz-Reyes *et al*, 2002; Palazzo *et al*, 2003; Rusché *et al*., 2001). Indeed, REL1 is not required for insertion editing to occur *in vitro* (Huang *et al.*, 2001). However, it has also been suggested that REL1 may be able to compensate for REL2, in its absence, by also functioning in addition, as well as deletion, editing and this would certainly explain why REL1 is an essential ligase (Cruz-Reyes *et al.*, 2002). However, the study by Huang *et al,* which inferred that REL2 is still required for addition editing, may suggest that REL1 has another very important cellular function in RNA repair (Huang *et al.*, 2001).

Miscleavage and misediting are known to occur *in vivo.* For example, truncated cDNA sequences of pan edited A6 and ND7, which are products of cleavage during editing, have been reported (Koslowsky *et al.*, 1991). Such events would lead to the loss of transcripts, unless there was a method to rescue them via re-ligation or cleavage-re-ligation, respectively.

Therefore, one objective of this thesis is, through genetic complementation methods, to determine why REL1 is essential in editing when REL2 is not, as this matter requires resolution. This current study looks to discern whether it is REL1's position in the deletion subcomplex, or its specific catalytic properties, or both, that makes it indispensable for RNA editing.

To ascertain the catalytic roles of the two ligases in editing, chimeric ligases were constructed and expressed in a REL1 conditional knock-out (cKO) cell line. Rescue copies of full length REL1 and REL2 were used as positive and negative

controls, respectively, to determine if the approach worked correctly. Endogenous REL2 still remains in this cKO REL1 cell line. This manner of approach allowed the involvement of the REL1 catalytic domain in addition editing and a REL2 catalytic domain in deletion editing to test if this structural dichotomy reflects a biological one. Figure 2.1, overleaf, shows the four possible outcomes from each cell line constructed in the absence of tetracycline inducible REL1. It also outlines the two possible scenarios tested in explanation of REL1's essentiality through expression of chimeric ligases.

Firstly (1), if REL1's catalytic properties are the reason that it is essential, then placing a REL1CD into the insertion subcomplex should compensate for the knock-out of the regulatable REL1 in this system. Secondly (2), if a ligase is required in the deletion subcomplex, but not necessarily REL1, then placing a REL2CD into this subcomplex should compensate for the loss of REL1.

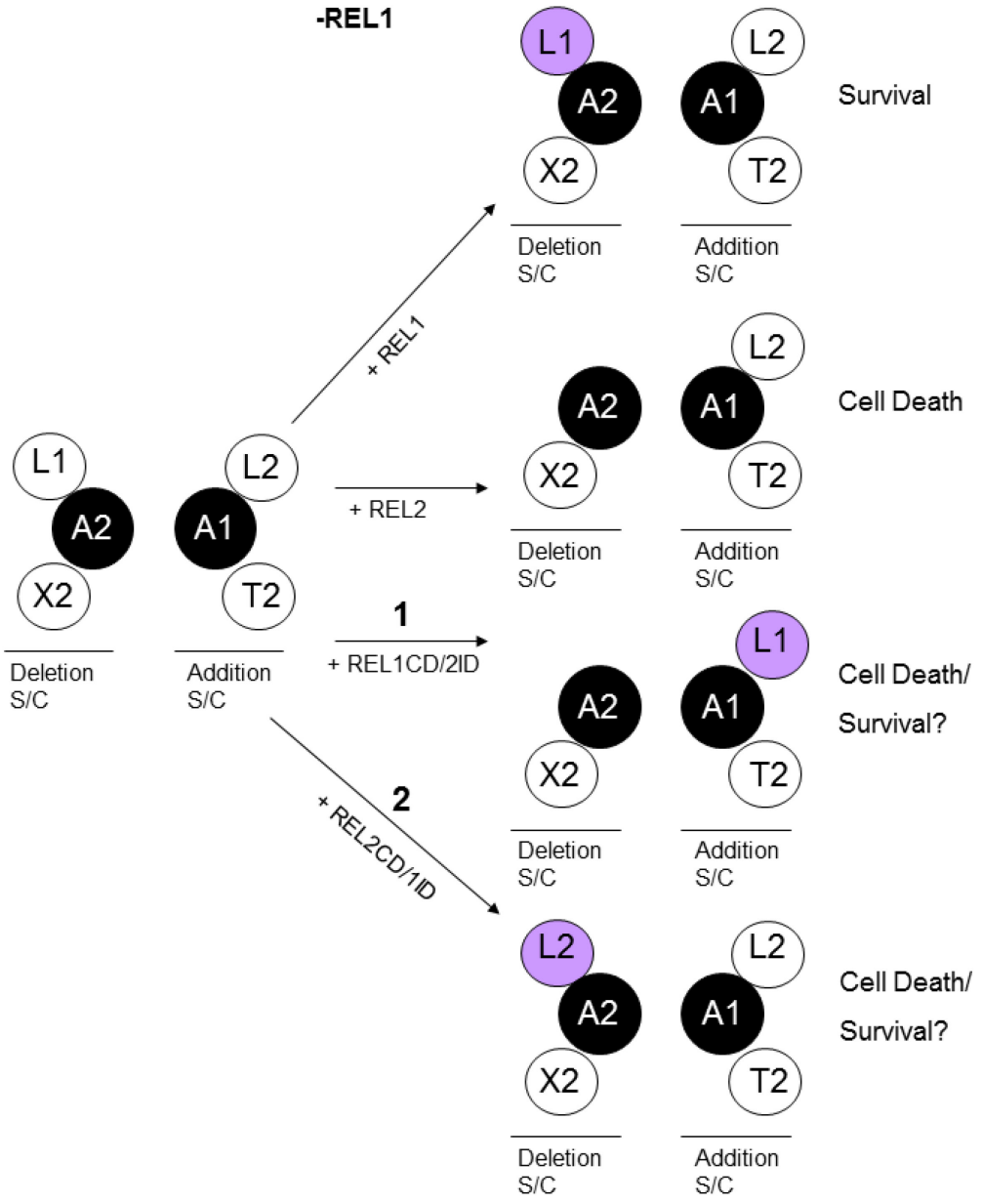


Figure 2.1. Outcomes of genetic complementation approach.

REL1-TAP and REL2-TAP, as controls, are expected to caused cell survival and death respectively. The outcome of tagged chimeric ligase integration is unknown, but will test whether REL1 can function from the addition subcomplex (1), or if REL2 can replace REL1 in the deletion subcomplex (2).

Abreviations: L2- REL2 ligase, A1 – KREPA1 protein, T2 - TUTase, L1 - REL1 ligase, A2 – KREPA2 protein and X2 - exoUase.

2.2 Methods

2.2.1 Cell lines used: cKO REL1 – TAP

REL1 cKO BSF cell lines, constitutively expressing TAP-tagged ectopic versions of either REL1, REL2, or chimeric proteins with recombined catalytic (CD) and interaction domains (ID); i.e. REL2CD:REL1ID and REL1CD: REL2ID, were constructed prior to this study by Matthew Spencer using pHD1344-derived vector, pHD1344t-TAP (Alibu *et al.*, 2004; Carnes, *et al.*, 2012a). The interaction domains of REL1 and REL2 mediate binding to the deletion and insertion subcomplex, respectively (See figure 1.6). Relevant schematics are illustrated in Figure 2.2 overleaf. Cells were passaged in 5-ml cultures with HMI9 medium containing 10% foetal calf serum (FCS) and selective drugs (2.5 μg/ml G418, 5 μg/ml hygromycin, 2.5 μg/ml phleomycin, 2.5 μg/ml puromycin and 1 μg/ml tetracycline to induce the ectopic REL1 allele).

To ensure cells were correctly expressing constitutive TAP tagged and endogenous REL1 (in the presence of tetracycline in the media) proteins at a comparable level, crude lysates were made for Western blot analysis. Cells were pelleted, and resuspended in 2 × SDS sample buffer, (4% SDS, 20% glycerol, 120 mM Tris-HCL, pH 6.8, 0.2% bromophenol blue, 200 mM DTT) to give 1×10^6 cells in 10 μl.

For subsequent miniTAP, glycerol gradient and auto-adenylylation experiments cell lines were grown for 48 hours in the absence of REL1 before pelleting, so allow for good ablation of ectopically expressed REL1.

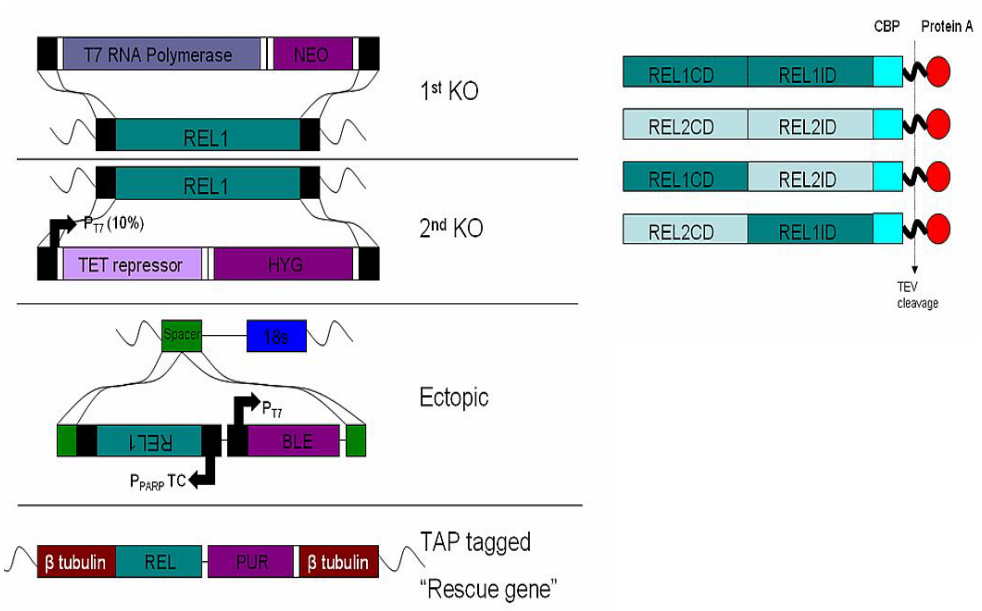


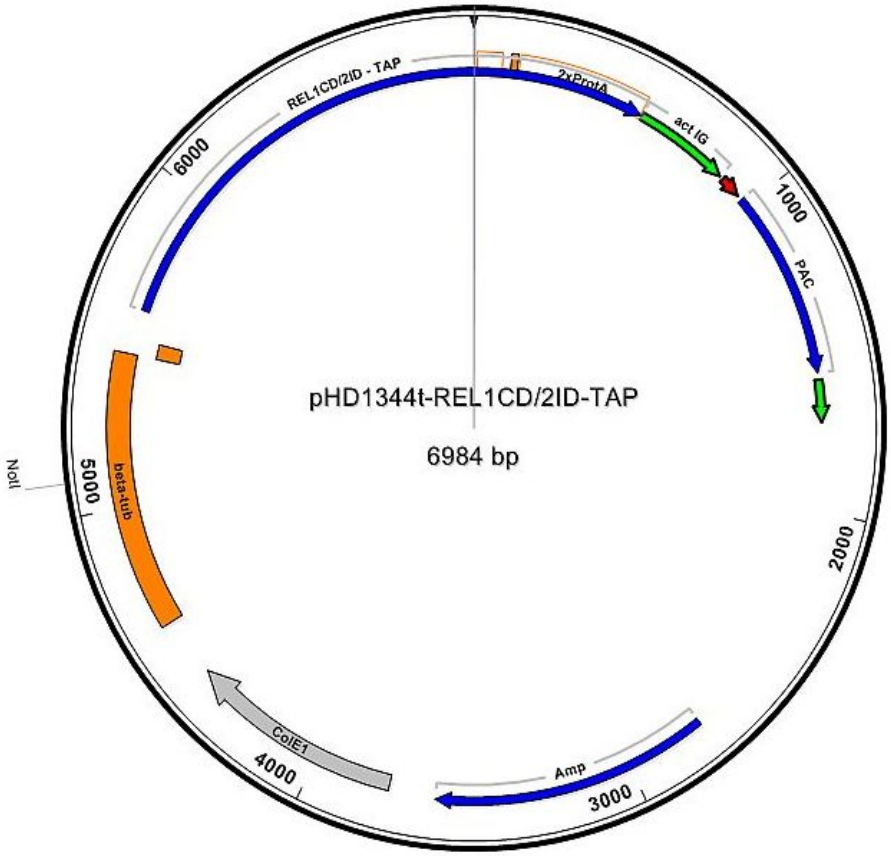
Figure 2.2. cKO constructs and strategy used.

Left: In the established cKO cell line (Schnaufer *et al.*, 2001) the first and second endogenous REL1 alleles have been replaced by T7RNAP/NEO and TETR/HYG cassettes, respectively, and the ectopic rescue copy is regulated by tetracycline. The cell lines used in this study also expressed constitutively expressed TAP tagged proteins (selection of transfectants with puromycin). Right: A schematic representation of recombinant and chimeric proteins expressed within the four cKO REL1 cell lines used. The TAP tag is located on the C terminus and CD and ID refers to the catalytic (adenylylation) domains and protein-interaction domains of these ligases, respectively.

The primer sequences and cloning strategy used here can be found in Appendix 1.

2.2.2 Creation of new REL1CD/2ID-TAP fusion constructs

To make chimeric sequences with 2 different fusion points (referred to as fusion 333 and 322), two different REL1 fragments corresponding to slightly different versions of the CD were amplified from genomic DNA of 427 strain *T. b. brucei* (Wirtz *et al.*, 1999) using the primer combinations A and B or A and C (tabulated in Figure 2.3). Next, to allow the subsequent insertion of the chimera into the available restriction sites in the pHD1344t-TAP plasmid (Alibu *et al.*, 2004; Carnes *et al.*, 2012a), the protein-protein interaction domain (ID) of REL2 had been mutagenised by Matthew Spencer to remove the *Hind*III site, using site-directed mutagenesis and primer sets (see Appendix 1 for primers and strategy) . This product was then subject to PCR reaction with primer sets D and F or E and F, respectively, to create the two different fusion points within the REL2 protein-protein ID. A standard Phusion 50 µl PCR reaction was set up (10 ng 427 gDNA, 0.5 µl Phusion high fidelity DNA Polymerase (Finnzymes) and 10 µl corresponding buffer, 200 µM dNTPs and 0.5 µM primers) with the following temperature programme: 98°C 2 minutes [98°C 10 s, 70°C 30 s, 72°C 30s] for 35 cycles, followed a five-minute incubation at 72°C. PCR reactions were analysed on a 1% TBE agarose gel, and amplicons were gel excised and cleaned up as per manufacturers' instruction using Nucleospin® Extract II (Macherey-Nagel). The corresponding REL1 CDs and REL2 IDs amplified were mixed in an equimolar ratio and subjected to a PCR reaction under the aforementioned conditions with primer set A and F. These primers simultaneously removed the STOP codon from the REL2 ID and added a *Hind*III and a *Bam*HI site to the 5' and 3' end of the chimera, respectively. PCR products were run on a 1% agarose gel, excised and cleaned using Nucleospin® Extract II.



Name	Sequence	Details
A	5' ATAAAGCTTATGCAACTCCAAAGGTTGGG	F" 5' REL1CD
B	5' GATTGATGGAGTCTCTTCGTACCGTGTCGATAAATGTC	R' fusion 333 REL1CD
C	5' CATCTACGCGAGGCCCTTGCTTACCGGGGTGCTTCAAC	R' fusion 322 REL1CD
D	5' GACATTTATCGACACGGTACGAAGAGACTCCATCAATC	F' fusion 333 REL2ID
E	5' GTTGAAGCACCCCGGTAAGCAAGGGCCTCGCGTAGATG	F' fusion 322 REL2ID
F	5' ATAGGATCCTTCGCTAAAGTCAGGAGACT	R' REL2 ID

Figure 2.3. Plasmid constructs for expression of TAP-tagged chimeric ligases.

The *in silico* map was constructed using Lasergene software (DNASTAR). The *Not*1 linearization of this plasmid allowed integration and constitutive expression of the TAP tagged protein from a β tubulin locus.

The fusion points used to create chimeras $REL1CD_{333}REL2ID$ and $REL1CD_{322}REL2ID$ and the fusion point of the chimera that did not integrate, ($REL1CD_{324}REL2ID$) are shown in Figure 1.7 (Chapter 1).

PCR products were ligated into Zero BLUNT® TOPO® vector (Invitrogen), as per manufacturers' instruction. Ligation products were used to transform 50 µl of competent XL1-Blue cells (Stratagene), which were spread on plates containing 50 µg/ml kanamycin. Three random colonies were picked and grown up in 2 ml of LB medium and 50 µg/ml kanamycin. Plasmid DNA was isolated using the peqGOLD kit (PEQLAB). Diagnostic restriction digests using *Eco*RI were set up using 100 ng of miniprep DNA and analysed on a 1% agarose gel to check for the presence of inserts, before sending DNA for big dye reaction sequencing (Genepool, Edinburgh). Correct inserts and pHD1344t-TAP (Alibu *et al.*, 2004; Carnes, *et al.*, 2012a) backbone were gel purified from plasmids doubly digested *Hind*III and *Bgl*II and *Hind*III and *Bam*HI, respectively. The two fragments were ligated in an equimolar ratio using T4 DNA ligase (NEB). XL1-blue cells wre transformed with ligation product and colonies were picked to grow midiprep cultures. DNA was isolated as per manufacturer's instruction (Machery-Nagel). 10 µg of *Not*I linearised plasmid was used to nucleofect 4×10^7 cKO REL1 cells using methods detailed by Burkard *et al.*, 2007. Briefly, cells were pelleted at 1,300 rpm, resuspended in 100 µl of transfection buffer (90 mM sodium phosphate, 5 mM potassium chloride, 0.15 M calcium chloride, 50 mM HEPES, pH 7.3), and mixed with 10 µg of linearised plasmid before nucleofection (Amaxa program Z-001). Cells were resuspended in media with G4.15, hygromycin and phleomycin (as detailed in Section 2.2.1), subjected to 10-fold dilutions, and allowed to recover for 6 hours in 24 well plates before drug selection with 0.1 µg/ml puromycin antibiotic. Clones were selected after 7 days. Clones A4 (fusion 333) and B4 (fusion 322) were used in subsequent analysis.

2.2.3 Growth analysis of cKO REL1 – TAP cell lines

The aforementioned cell lines were subjected them to growth analysis by Matthew Spencer (with and without tetracycline) over a 5 to 6 day period to ascertain which RELs could rescue the cKO REL1 growth phenotype. Briefly, cells were grown in 5 ml cultures containing selective drugs (2.5 μg/ml G418, 5 μg/ml hygromycin, 2.5 μg/ml phleomycin and 2.5 μg/ml puromycin) without the presence of tetracycline and were diluted to 100,000 cells/ml each day after counting. Fresh tetracycline (1 μg/ml) was added daily.

2.2.4 TAP purification of tagged proteins

All proteins were TAP-tagged to allow their purification (and that of associated protein complexes) via a tandem affinity purification protocol (TAP) (Rigaut *et al.*, 1999). For small scale purifications, a protocol modified from the full trypanosome TAP protocol (Schnaufer *et al.*, 2003), here called miniTAP was used (Figure 2.4).

Before the miniTAP could be established magnetic beads (Dynabeads® M-270 epoxy - Invitrogen) were covalently linked to IgG, as described by Oeffinger *et al.*, 2007, and modified by Achim Schnaufer. Briefly, 4×10^9 beads were resuspended in 4 ml of 0.1 M $NaPO_4$, pH 7.4, through vortexing and were aliquotted into four separate 1.5-ml Eppendorf tubes. Each tube of beads was treated separately as follows. The bead suspension was gently shaken for 10 minutes before tubes were placed into a magnetic rack and buffer was aspirated. Beads were washed once with 1 ml $NaPO_4$ and incubated with 1 ml of antibody mix (2.5 mg rabbit IgG (Sigma), 50 mM $NaPO_4$, 1 M ammonium sulphate) for ~20 hours at 30°C, with gentle agitation.

Beads were subsequently washed, quickly, with 600 µl 100 mM Glycine-HCL, pH 2.5, once with 600 µl 10 mM Tris, pH 8.8 and once, quickly, with 600 µl opf fresh 100 mM triethylamine. Coated beads were then subjected to four 5-minute washes with 1 ml PBS, one wash with 1 ml PBS with 0.5% Triton-X 100 for 5 minutes and a final 15-minute wash with PBS with Triton-X 100. Beads were finally resuspended in 1 ml PBS with 0.02% sodium azide.

For each miniTAP procedure, 1×10^8 cells were, washed once with ice cold 1 M phosphate buffered saline with 6 mM glucose (PBS-G), pelleted again, and either stored at -80°C or directly processed as follows. In short, 100 µl of IPP150 (10 mM Tris-HCL pH8, 50 mM NaCl, 0.5 mM EDTA, 0.1% NP40, 1 mM DTT), containing mini EDTA – free protease inhibitor (Roche), 1 tablet/10 ml) was added to a fresh or frozen pellet of 1×10^8 cells. Cells were lysed by adding Triton X-100 to a final concentration of 1% (where frozen cells were used, these were allowed to thaw on ice first). The lysate was incubated on ice for 20 minutes before centrifugation at 10,000 g for 15 minutes at 4°C (Fresco 21, Thermoscientific). The supernatant (*i.e.* cleared cell lysate) was added to 10 µl of 2 × BSA-preincubated magnetic beads and were left rotating at 4°C for two hours. The supernatant was removed and the beads washed three times with 250 µl of IPP150 and once with 100 µl of TEV cleavage buffer (10 mM Tris-HCL pH8.0, 150 mM NaCl, 0.1% NP40 before the addition of 10 U AcTEV protease (Invitrogen) in 30 µl of TEV cleavage buffer and two hours incubation, with gentle agitation, at 16°C.

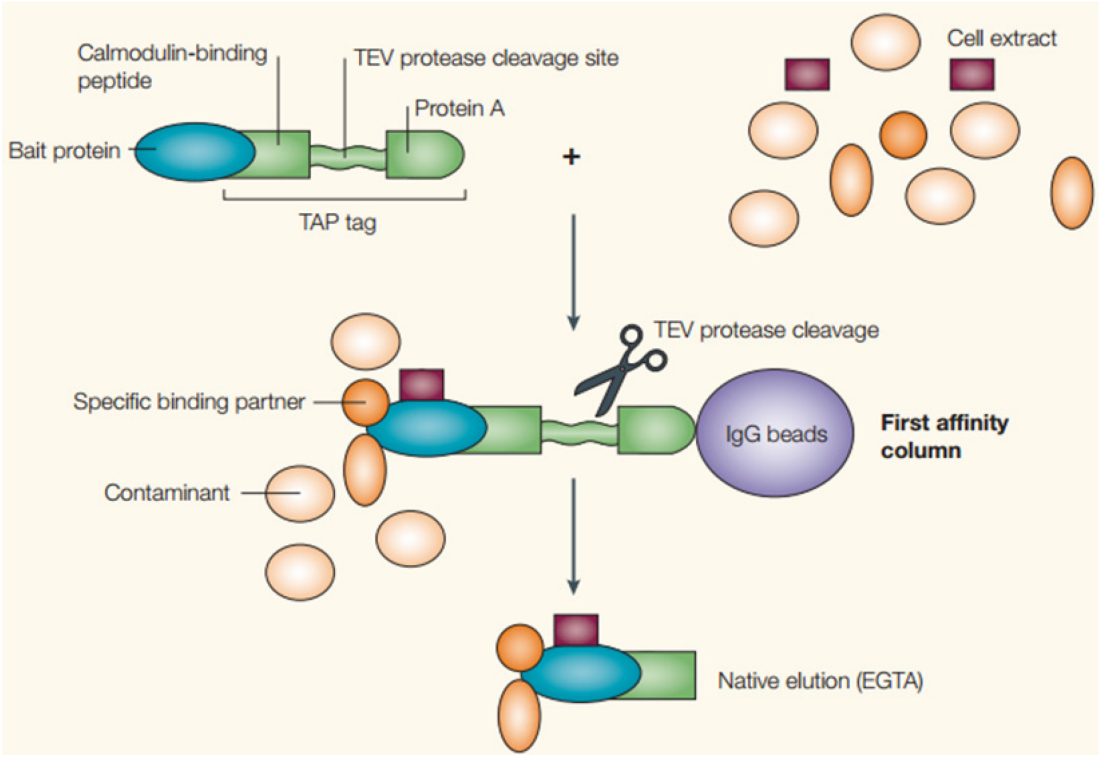


Figure 2.4. The MiniTAP procedure.

This method was used throughout all experiments in the thesis. TAP tagged ligases were constitutively expressed in the conditional REL1 KO environment and affinity purification used to isolate the ligases and their associated proteins. Since eluates were not sent for mass spectrometry, a single step of purification was sufficient for analysis.

TAP tagged proteins were purified using magnetic beads covalently linked to IgG, and were eluted through cleavage with AcTEV protease. After elution, TAP tagged ligases contained the CBP (Calmodulin Binding Peptide) part of the tag and so could be detected using an antibody raised to this part of the tag.

Image modified from Huber, 2003.

Eluates and samples from each intermediate step of the protocol were collected for Western blot analysis to optimise the procedure.

2.2.5 Glycerol gradient sedimentation of native editosome subcomplexes

Cells were grown to a density of approximately 1-2 × 10^6 cells /ml with 1L HMI9 containing selective drugs, washed in PBS-G, and pelleted. Flashfrozen pellets were stored at -80ºC prior to glycerol gradient sedimentation. The total number of cells present in each pellet is tabulated below in Table 2.1.

Table 2.1 Number of cells used in glycerol gradients and subsequent TAP.

TAP tagged construct	# cells in pellet
rREL1	7.15×10^8
rREL2	1.05×10^9
REL2CD/1ID	9.96×10^8
REL1CD/2ID 1	1.27×10^9
REL1CD/2ID Q	1.31×10^9

Pellets were thawed and cellular matter was lysed in 500 µl of IPP150 and 1% Triton X-100. After a 20-minute incubation on ice, debris was pelleted by centrifugation at 10,000g for 15 minutes at 4ºC. Cleared cell lysate was then directly loaded onto 10-ml glycerol gradients and subjected to a 9-hour centrifugation at 38,000 rpm using a Beckman SW40Ti rotor and a Beckman L-60 Ultracentrifuge. 10

– 30% glycerol gradients were poured using a Hoefman SG15 gradient mixer, and were stored for a maximum of 1 hour on ice before use. Briefly, 5 ml of buffer A (10 mM Tris HCL pH 7.2, 10 mM MgCl2, 100 mM KCl, 1 mM DTT, 10% glycerol) and buffer B (10 mM Tris HCL pH 7.2, 10 mM MgCl2, 100 mM KCl, 1 mM DTT, 30% glycerol) was added to compartments 1 and 2, respectively. Both buffers contained freshly added protease inhibitors (1 mM Pefabloc, 2 µg/ml Leucopeptin and 1 µg/ml Pepstatin A in a total 11 ml). The two buffers were mixed gradually as per manufacturers' instruction and poured into Beckman thick-walled centrifuge tubes (331374). One to two gradients were run at a time, and 500 µl fractions were collected from the top and kept on ice prior to miniTAP analysis.

Freshly collected fractions were pooled ready for analysis into the following 4 pools: 1-3, 6-8, 9-11 and 12-14, based on published information (Schnaufer *et al.*, 2003) and from optimised 10-ml, 9- hour glycerol gradients on wild type (wt) 427 cells. These fractions represent non-incorporated ligases, ligase integration into subcomplexes, ligase integration into the 20S editosome and ligase integration into larger complexes, respectively. The composition of these fractions was confirmed in pilot experiments using wt 427 cells and antibodies available for editosome components KREPA1, A2, A3 and REL1 (Panigrahi *et al.*, 2001). 40 µl aliquots were taken from each of the pools for western analysis and the remainder was subjected to a TAP analysis, using 50 µl IgG-coupled, BSA-blocked magnetic beads. TAP was performed as in Section 2.2.4, and scaled up accordingly to accommodate for starting cell number. TAP tagged proteins were eluted in 150 µl of TEVCB and were concentrated using StrataClean resin (Stratagene). Briefly 5 µl of resuspended beads were added to the TEV eluate in 1.5-ml Eppendorf tubes, vortexed thoroughly

and subjected to centrifugation at 2,000 g for 1 minute. The supernatant was removed and the resin was resuspended in 10 µl of distilled water and 5 µl of 4 × SDS sample buffer. Samples were boiled for 5 minutes, fractionated by SDS-PAGE, and transferred to a PVDF membrane by Western blotting (see Section 2.2.7).

2.2.6 Ascertaining the activity of TAP tagged ligases through radioactive adenylylation

200 ml of culture of each cell line was grown and pelleted, then subjected to miniTAP purification. Aliquots of whole cell lysates and eluate samples were kept for Western blot analysis. TEV eluates were subjected to adenylylation and deadenylylation reactions as described by Cruz-Reyes *et al.,* 2002 and Sollner-Webb *et al*, 2001.

Half the TEV eluates, containing 5% glycerol, (6 µl) were first fully deadenylylated with 16 mM freshly made tetra pyrophosphate solution (pH 8.0) through pre-incubation for two minutes on ice. Excess phosphate was then removed through the addition of one unit of pyrophosphatase (Sigma) and reactions were incubated on ice for a further five minutes before adenylylation. To all reactions 12 µl adenylylation master mix was added, creating reaction conditions with 25 mM KCl, 12.5 mM HEPES, pH 7.9, 5mM Mg acetate, 0.25 mM DTT and 10 µCi/µl (3.3 µM) [α-^{32}P] ATP. After a five-minute incubation at room temperature, reactions were stopped through addition of 20 µl 2 × SDS sample buffer (with 1% BSA) and denatured at 70°C for 10 minutes. Samples were run on a precast 10 % Bis-Tris Midi gel (Invitrogen) in 1 × MOPS, with 'NuPage antioxidant (Invitrogen) in the upper

chamber, for 45 minutes at 200 V.

After the removal of gel from its casing, the bottom of the gel, below the 40 kDa marker, contianing free [α-^{32}P] ATP, was excised and disposed of. The remaining gel was transferred to a Perspex box and incubated in 200 ml fixing solution (50% methanol/10% acetic acid) with gentle agitation for 30 minutes. The gel was then incubated with 200 ml equilibration buffer (7% methanol/7% acetic acid/1% glycerol) for another 30 minutes with agitation. The gel was then removed, placed onto two pieces of 3MM Whatman paper, covered with parafilm and dried for 1 hour at 80 °C in a 583 vacuum gel dryer (Biorad). Once completely dried, the gel was exposed for 20 hours to a phosphor-imaging screen (Molecular Dynamics), before scanning on a Typhoon Trio variable mode imager (GE Healthcare). A 50 μm resolution image was taken, using the phosphor setting, and was analysed using ImageQuantTL. Densitometry results were visualised using Graphpad prism.

2.2.7 Western Blotting

Equivalent amounts of protein (pertaining to 1 × 10^6 cells starting material) were assessed by Western blotting. All samples were boiled in 2 × SDS sample buffer, (4% SDS, 20% glycerol, 120 mM Tris-HCL, pH 6.8, 0.2% bromophenol blue, 200 mM DTT) for 10 minutes before 10µl of sample was run per gel lane. SDS-PAGE was carried out for 90 minutes at 150 V, using the Nupage® (Invitrogen) system (Novex® 10% BisTris gels with 1 × MOPS running buffer). Gels were blotted onto pre-equilibrated Immobilon-P (Millipore) PVDF membranes, using a Biorad® apparatus at 90 V for 45 minutes, before blocking overnight in TBST buffer (10 mM Tris-HCl, 150 mM NaCl and 0.05% Tween-20) with 10-20% dry milk as blocking agent. All washes were carried out with TBST and all antibody blocking steps involve a one hour incubation with antibodies in TBST / 5% dry milk.

Blots were stripped after exposure using stripping solution (0.1 M DTT, 0.05 M Tris HCL, 2% SDS, pH 7) and incubation at 50 ºC for 1 hour. Blots were then washed for 30 minutes in TBST before repetition of the Western procedure from the initial blocking step.

Bands were detected using ECL or ECL plus (Amersham), which visualised signals given from horseradish peroxidase (HRP)-coupled goat secondary antibodies (Bio-Rad).and were developed using Kodak MBX films and an SRX-101A X ray developer (Konica Minolta). The antibody concentrations are tabulated overleaf in Table 2.2.

Table 2.2 Antibodies and respective concentrations used in Chapter 2.

Name	Protein	Concentration	Secondary Ab (Biorad)	Reference
α-REL1 (P3C1)	REL1	1/1000 or 1/100	1/2000 or 1/1000 α-mouse	Gift Stuart Lab (Panigrahi *et al.*, 2001)
α-KREPA1 (P4D8)	KREPA1	1/50	1/1000 α-mouse	
α-KREPA2 (P1H3)	KREPA2	1/100	1/1000 α-mouse	
α-KREPA3 (P3C12)	KREPA3	1/50	1/1000 α-mouse	
α-CBP	CBP (TAP-tag)	1/1000	1/2000 α-rabbit	Millipore
PAP	Protein A (TAP-tag)	1/5000	N/A	Sigma
TAT α-tubulin	B-tubulin	1/5000	1/5000 α-mouse	Gift Matthews lab

2.3 Results

2.3.1 Growth analysis of REL1 cKO -TAP cell lines

Matthew Spencer's results (Figure 2.5) indicated that only ectopically expressed REL1 (dashed purple line), but not REL2 or either of the two chimeric forms (other dashed lines), can rescue the growth phenotype caused by shutting down expression of REL1 in the cKO cell line. The parental REL1 cKO cell line (black) was used for comparison. All REL1 expressing cells grew exponentially. All other cells exhibited growth arrest by 70 hours, followed by death. All tagged chimeras are expressed, comparably (Figure 2.6).

2.3.2 TAP analysis of tagged proteins and their integration into the editosome

In order to test whether these TAP tagged proteins associated with the expected subcomplexes within the editosome, a suitable purification procedure had to be established. Optimisation of the miniTAP using IgG covalently linked to magnetic beads (single stage of purification) allowed efficient purification of TAP-tagged complexes from trypanosome lysates (results not shown). Relatively small numbers of cells were sufficient to reveal tagged and associated proteins by Western blotting from whole cells lysates (Figure 2.7). The four cell lines (expressing each of REL1, REL2, REL1CD/2ID and REL2CD/1ID-TAP) were then subjected to the miniTAP protocol, followed by Western analysis using antibodies against CBP, REL1 and KREPA2 to test whether the TAP-tagged REL proteins integrated into the editosome and could be pulled down efficiently.

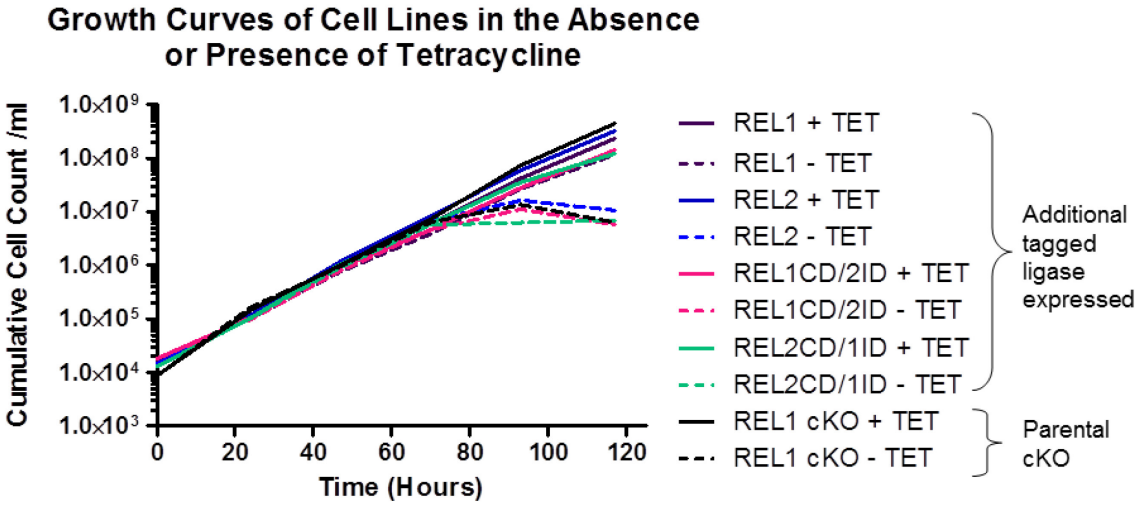


Figure 2.5. Growth Curve of cKO Cell Lines.

Matthew Spencer's growth curve clearly shows that only an additional copy of REL1 (dashed purple line) can rescue the growth phenotype caused by REL1 cKO conditions (*i.e.* in the absence of the tetracycline inducer).

Solid and dashed lines correspond to cells grown in the presence and absence tetracycline, respectively.

The expression and integration data before and after TAP analysis is shown in Figure 2.7. These results indicate that all proteins could be purified by miniTAP, as indicated by the presence of a CBP band of the correct size in whole cell lysates (WCL) at ~60 kDa and in eluates (E(10) at ~50 kDa (Figure 2.7A). The size shift from whole cell lysates to eluates is indicative of the loss of protein-A from the TAP tag upon TEV cleavage. REL1 and REL2CD/1ID tagged ligases could be visualised by Western using α-REL1, as this antibody is indicative of the ID only. Figure 2.7B shows the presence of inducible REL1 at ~50 kDa in all cell lines grown in the presence of tetracycline, as expected. Tagged REL1 can be detected in -tet samples only in REL1 and REL2CD/1ID whole cell lysate and eluate samples. The inducible copy of REL1 is also visible in eluates (+tet) from cell lines constitutively expressing REL1, REL2 and REL2CD/1ID, but not REL1CD/2ID. To add to this KREPA2 (which is consistently masked in whole cell lysates) cannot be detected in eluates from REL1CD/2ID (Figure 2.7C). This shows all save the REL1CD/2ID chimeric protein could integrated properly into the editosome, since KREPA2 was present in the eluates of the other three cell lines. This indicates that 20S editosomes can successfully be pulled down by miniTAP via tagged REL1, REL2 and REL2CD/1ID proteins. This finding further indicated that the REL1CD/2ID chimeric protein created, perhaps because it was not folded correctly, was not able to integrate into the editosome efficiently. This motivated us to construct new REL1CD/2ID-TAP chimeric protein for expression within the cKO REL1 environment.

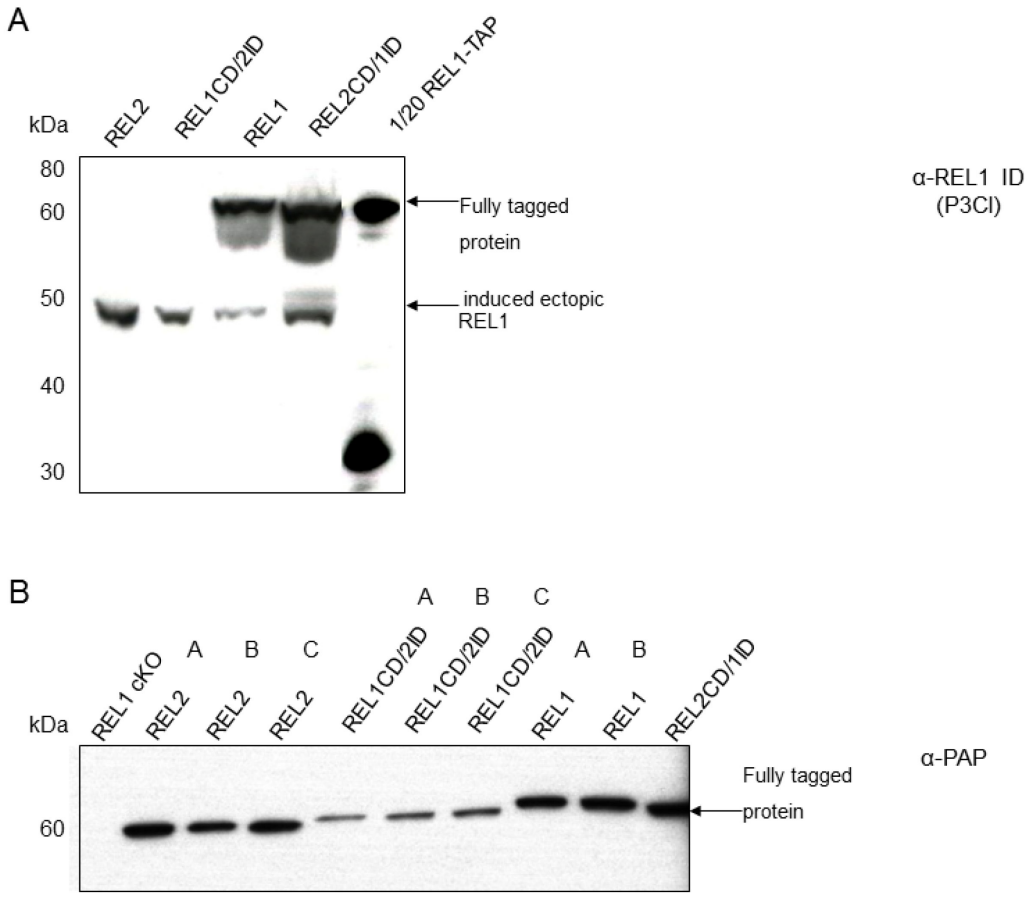


Figure 2.6. Initial confirmation of expression of TAP-tagged ligase proteins.

Western analyses were conducted with whole cell lysates (1×10^6 cells per lane) of REL1 cKO cells constitutively expressing TAP-tagged REL1, REL2, or chimeric proteins. Uninduced REL1 cKO cells and purified REL1-TAP fractions were used as negative and positive controls, respectively.

A) Detection with 1/1000 P3Cl (REL1), which is specific for the C-terminal protein-protein ID, clearly indicating the presence of endogenous REL1 across the cell lines as well as the REL1 ID present in the TAP-tagged REL1 and REL1ID/2CD constructs. The size increase from induced ectopic REL1 is consistent with the size of the TAP tag. B) Detection of the protein-A part of the TAP tag with a 1/5,000 dilution of PAP indicates comparable expression of tagged proteins across all cell lines, and gives no signal for the control REL1 cKO cells, as expected. Image courtesy of Matthew Spencer.

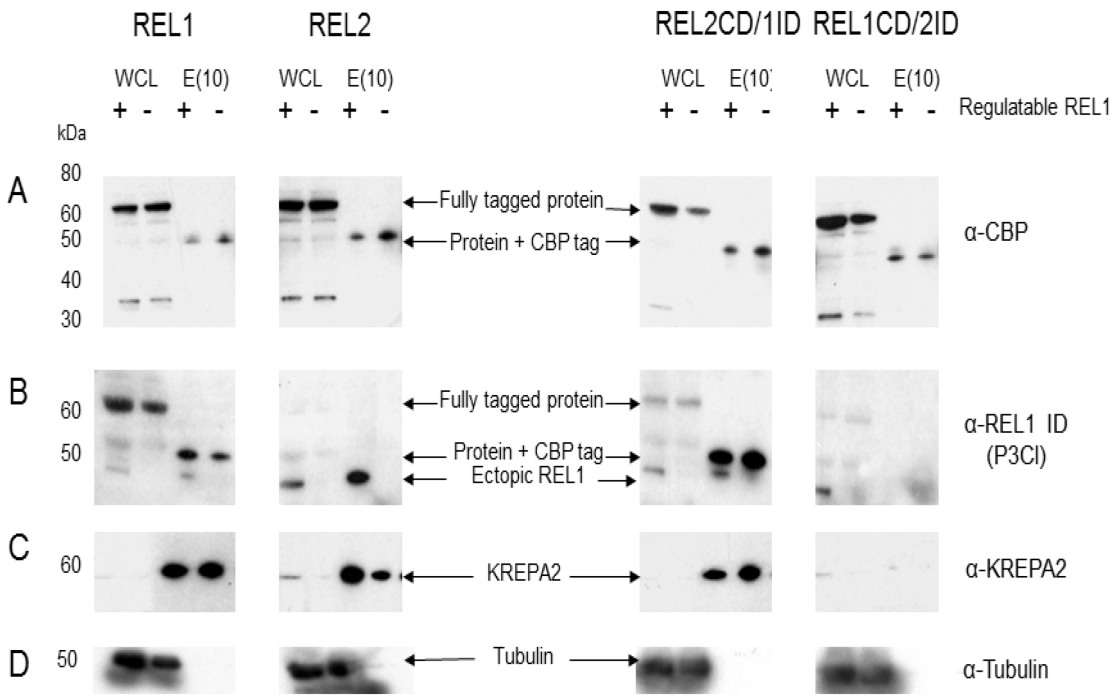


Figure 2.7. TAP purification of tagged REL1, REL2 and chimeric proteins.

Tagged proteins and associated complexes were purified from whole cell lysates using IgG-coated magnetic beads (which bind the protein A part of the TAP tag) followed by TEV protease cleavage. TEV eluates were analysed by immunoblot using antibodies against A) calmodulin binding protein (CBP part of the tag), B) REL1ID, C) RECC component KREPA2, or D) Tubulin. The data suggest all tagged proteins except for the REL1CD/REL2ID chimera successfully integrated into editosomes.

Abbreviations: WCL – whole cell lysate, E(10) – eluate.

2.3.3 Creation of a new REL1CD/2ID tagged chimeric ligase

The initial integration analysis in Figure 2.7 indicated that the tagged REL1CD$_{324}$/2ID can not integrate into the editosome, even though the TAP tagged proteins, are comparable to the tubulin control. The newly created REL1CD/2ID constructs were based on two different fusion points (see Figure 1.7 and 2.3) designated as REL1CD$_{333}$/REL2ID and REL1CD$_{322}$/REL2ID. Figure 2.8A (overleaf) clearly shows that the four clones analysed express the new TAP tagged REL1CD/2ID fusions at a level comparable to REL1-TAP (as the CBP blot indicated a comparable amount of tagged protein throughout each miniTAP purification), but indicates a less efficient integration into native editosomes, as indicated by a weaker signal on the KREPA2 immunoblot (Figure 2.8B). For subsequent experiments clones A4 (REL1CD$_{333}$/REL2ID) and QB5 (REL1CD$_{322}$/REL2ID) were used, since integration into editosomes was more efficient in these cell lines than in the others analysed.

It is clear that in the absence of tetracycline in the media (and hence down-regulated REL1) these cell lines exhibit the same growth arrest (typically after ~60 hours) as the parental cKO REL1 cell line (Figure 2.8B). To dissect this finding further, the integration of the tagged ligases into the correct RECC subcomplex had to be ascertained, through the isolation of specific deletion and addition subcomplexes on glycerol gradients.

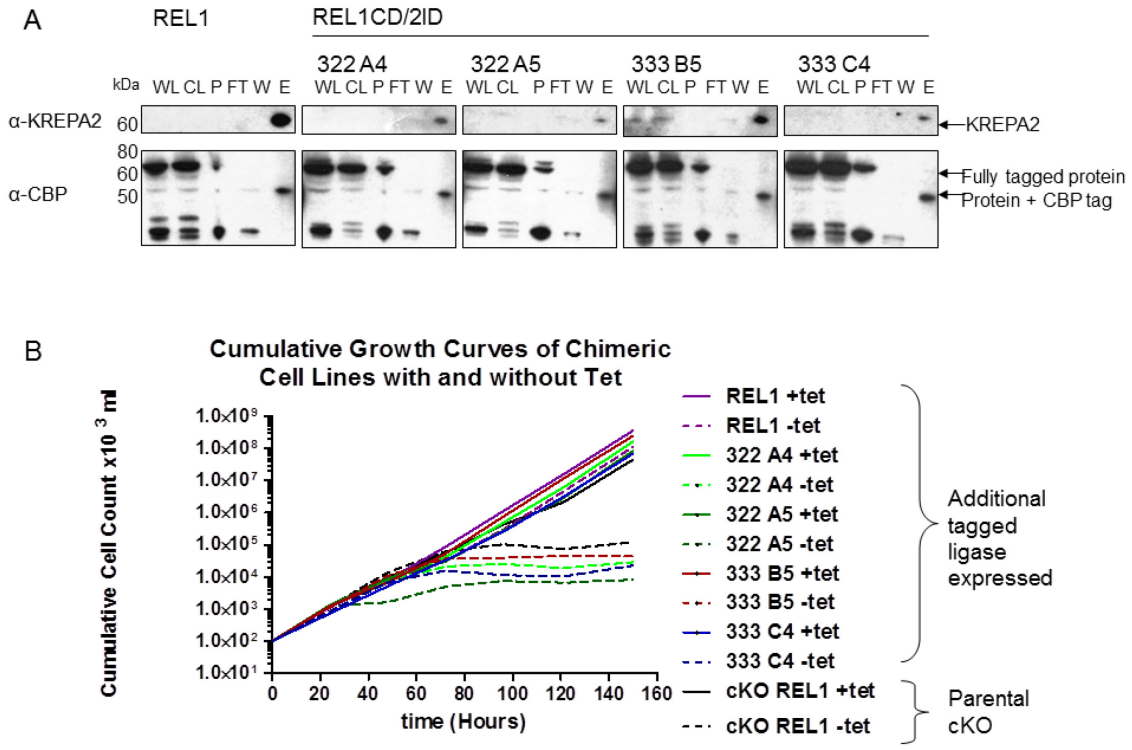


Figure 2.8. REL1CD/2ID growth curve and integration into the editosome.

A) TEV eluates were analysed by immunoblot using α-CBP and α-KREPA2 to ascertain expression levels and integration into 20S editosomes. The data suggest both clones of the two new REL1CD/REL2ID chimera fusions (322 and 333, Figures 1.7 and 2.3) are expressed comparably to the REL1-TAP protein and successfully integrate into editosomes, albeit not as efficiently. B) Growth of REL1 cKO lines shows only an additional copy of REL1 (dashed purple line) can rescue the growth phenotype caused by REL1 cKO conditions (*i.e.* in the absence of the tetracycline inducer), as can also be seen in Figure 2.5.

Abbreviations: WL – whole cell lysate, CL – cleared lysate, P – pellet, FT – flow through, E– eluate.

2.3.4 Isolation of subcomplexes using glycerol gradients

To establish conditions for subsequent analysis, cleared *T. brucei wt* 427 lysates were fractionated on 10-ml glycerol gradients. Two separate gradients gave similar results when concentrated samples were immunoblotted with monoclonal antibodies (MAbs) for REL1, KREPA1 KREPA2 and KREPA3 (Figure 2.9 A). From this, it was decided to pool fractions 9, 10 and 11 for the 20S editosome, as these fractions showed co-integration of all four editosome components. This was consistent with data from the literature (Schnaufer *et al.*, 2003).

TAP-expressing ligase cell lines were grown in the absence of tetracycline (in the absence of untagged REL1) before glycerol gradient fractionation and subsequent miniTAP. It is clear from the concentrated eluate samples subjected to Western blotting analysis (Figure 2.9 B) that TAP tagged ligases were able to integrate into the subcomplexes, for which they were intended. REL2CD/1ID-TAP and REL1-TAP ligases integrate into the deletion subcomplex, since immunoblotting detects KREPA2 in the TEV eluates of the subcomplex and 20S fractions. KREPA1 is detected in the 20S eluates, from which REL1-TAP indicates that the whole editosome is pulled down. Likewise, REL2-TAP, REL1CD_{333}/2ID-TAP and REL1CD_{322}/2ID-TAP fusions integrate into the addition subcomplex as KREPA1 can be detected in the subcomplex and 20S editosome fractions.

CBP antibodies indicate the expression of a TAP tagged ligase in fractions pertaining to unincorporated, subcomplex, RECC and post editosome fractions across all 5 cell lines. A schematic diagram of the editosome components detected in TAP eluates, and inferred integration position of tagged ligases within the editosome

is shown in Figure 2.9 C. Nonetheless, detection of editosome proteins in 20S fractions was inconsistent from REL2-TAP, REL1CD$_{333}$/2ID-TAP and REL1CD$_{322}$/2ID-TAP eluates, in that only KREPA1 could be detected, and not KREPA2 as would be expected. KREPA3 could not be detected in any samples, after glycerol gradient and miniTAP even though it can be clearly detected in mitochondrial extract (mitoprep) controls (Figure 2.9B).

2.3.5 Discerning the activity of TAP tagged ligases

Prior to undertaking the adenylylation assays with the TAP tagged ligases, a control experiment was set up to determine if the TEVCB (containing DTT) would interfere with activity. Here, rREL1 expressed in *E. coli* and T4Rnl2 were subjected to adenylylation reaction, in respective storage and adenylylation buffers, with and without the presence of TEVCB. The presence of TEVCB did not affect the efficiency of adenylylation (results not shown).

The activity of the isolated TAP tagged ligases was determined using adenylylation and deadenylylation assays, with radiolabelled ATP in conjunction with phosphorimaging analysis, so that this could be related to the growth curves. Deadenylylation through addition of free phosphate was attempted, because REL2 has a high affinity for ATP and so is already adenylylated within the cell, and we wanted to measure this adenylylation activity. Auto-adenylylation was used as a proxy for TAP-tagged ligases, and had advantages over a full ligase activity assay, since different ligases in a sample can be distinguished (Cruz-Reyes *et al.*, 2002).

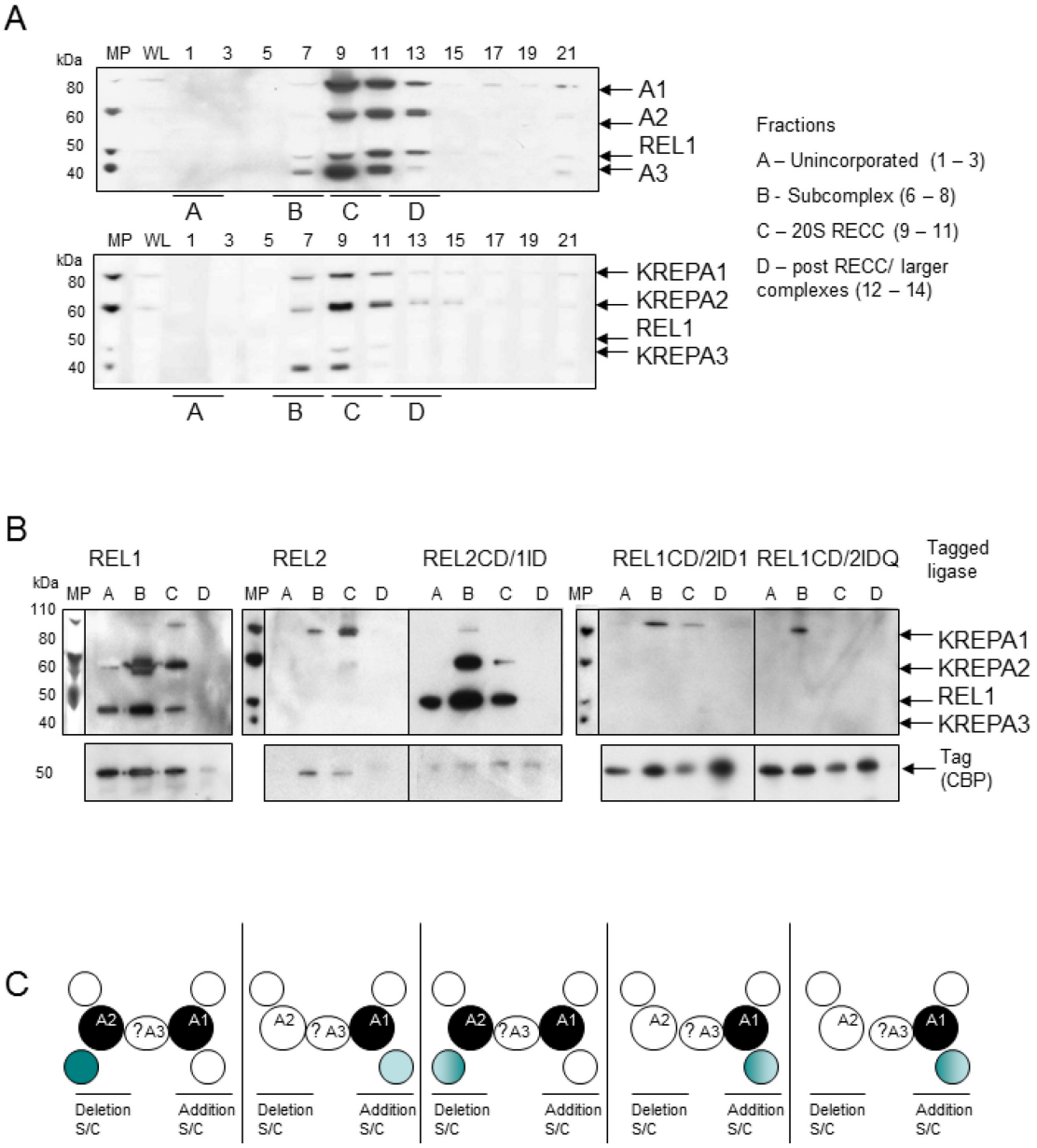


Figure 2.9. Integration of TAP tagged ligases into editosome.

(A) Glycerol gradients and subsequent Western blots were performed first with wt 427 BSF cells to determine reproducibility and the fractions to be pooled for subsequent TAP. (B) Western blots of TAP purified pooled glycerol gradient fractions after fractionation of lysates from REL1-ablated cells. Eluates were probed using antibodies against CBP, REL1, or RECC components KREPA1, A2 and A3. (C) A schematic of TAP tagged ligases into the editosome. Sites of tagged ligase integration are indicated in blue and protein components detected by Western blotting are indicated in black. KREP A3 could not be detected in (B), even though mitochondrial extract controls (mitoprep) indicated all antibodies were working correctly. Abbreviations: MP - mitoprep control, S/C – subcomplex, WL – wild type whole cell lysate.

The most active ligase, by far, was REL1-TAP, present as a band at ~60 kDa. A small band at ~48 kDa indicates the presence of a small amount of regulatable REL1 after ablation. Endogenous REL2 could also be detected in these lanes (Figure 2.10A, lanes 1 and 6) as well as in the lanes containing REL2CD/1ID-TAP ligase indicative of integration into the deletion subcomplex (Figure 2.10A, lanes 3 and 8), at 47 kDa. The strength of the endogenous REL2 adenylylation signal is comparable in lanes 1, 3, 6 and 8. Phosphate treatment did not significantly improve the detectable activities of ligases possessing the REL2 catalytic N-terminal domain, in the absence of endogenous REL1 suggested that the attempted deadenylylation had not been successful (Figure 2.10A, lanes 2, 4, 5, 7, 9 and 10). These reactions did not reveal the presence of endogenous REL2, suggesting that REL2-TAP, REL1CD$_{333}$/2ID-TAP and REL1CD$_{322}$/2ID-TAP integrated correctly at REL2's site in the addition subcomplex. However, since the deadenylylation reaction did not work, even after multiple attempts, the activities of the REL1CD and REL2CD cannot be reliably compared.

Western blotting was carried out on whole cell lysate and eluate fractions collected from each TAP purification, using CBP antibody to determine amount of TAP tagged protein present, and KREPA2 to determine the integration efficiency into the editosome (Figure 2.10B). Results indicate that neither the amount of isolated ligase used in each radioactive assay nor the integration efficiency into native editosomes was equal across the five cell lines.

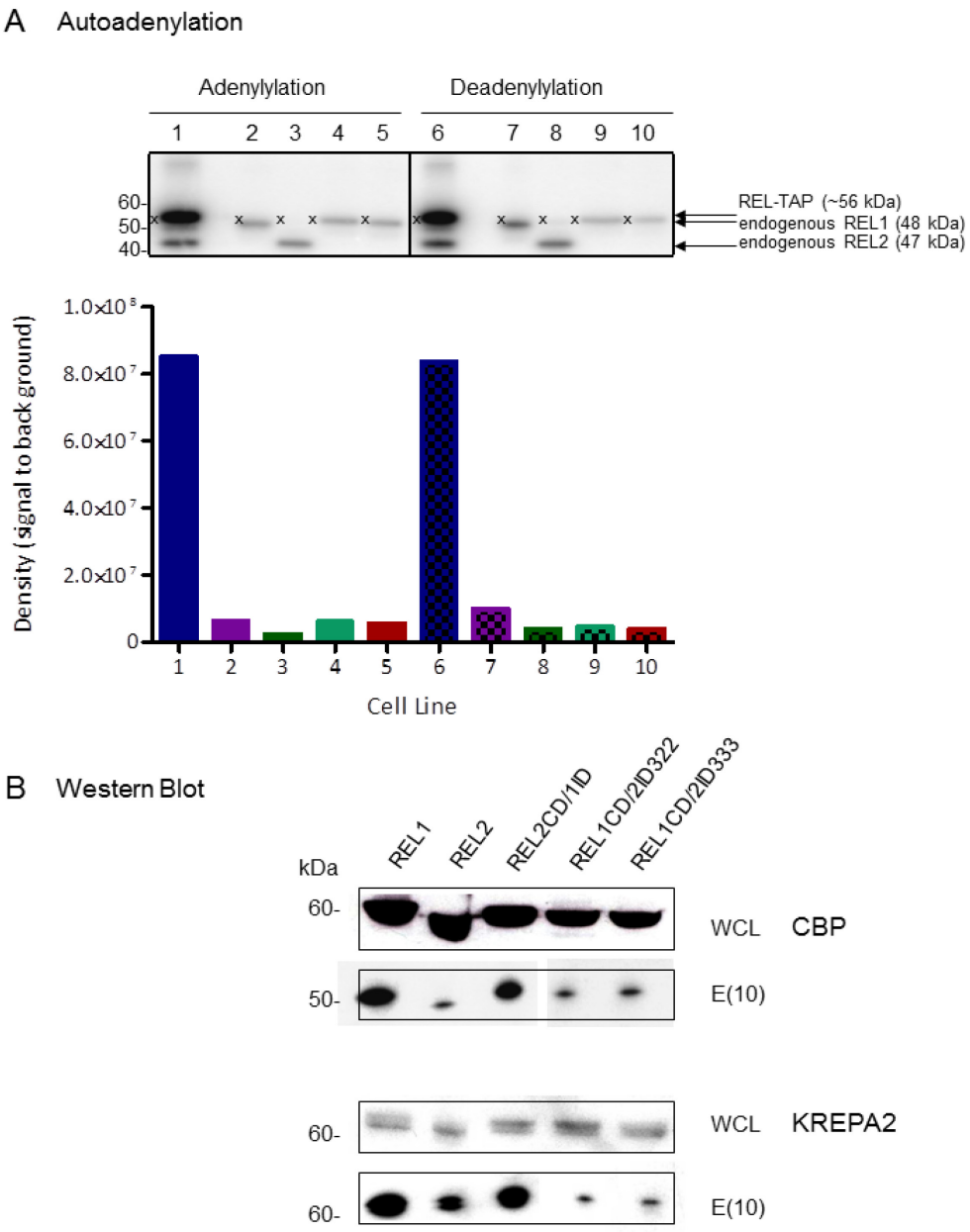


Figure 2.10. Activity of TAP tagged ligase determined through radioactive auto-adenylylation.

A) Auto-adenylylation gel and quantification of gel using phospho-imaging. Assay conditions were as follows: 1-5 without prior adenylylation; 6-10 deadenylylation prior to adenylylation. 1,6 REL1-TAP, 2,7 REL2-TAP, 3,8 REL2CD/1ID-TAP, 4,9 $REL1CD_{333}/2ID$-TAP, 5,10 $REL1CD_{322}/2ID$-TAP. B) Western blot of cell equivalent whole cell lysates and eluates used in each assay. Each lane was loaded with approximately 0.5×10^6 cells and probed with 1/500 α-CBP and 1/2000 α-rabbit antibodies. Blots were then stripped and reprobed with 1/50 A2 and 1/1000 α- mouse antibodies. Abbreviations: WCL – whole cell lysate, E(10)– eluate.

Indeed, full length REL1-TAP and REL2CD/1ID-TAP were the most abundant in the TEV eluate and all whole cell lysates contained a comparable amount of tagged ligases (CBP) (Figure 2.10B). Figure 2.10, in sum, clearly indicates that the REL2CD/1ID-TAP is not as active as REL1-TAP. Although the amounts of REL2-TAP, $REL1CD_{333}$/2ID-TAP and $REL1CD_{322}$/2ID-TAP ligases (CBP) and of KREPA2 were comparable in the TEV eluates, suggesting integration into editosomes with similar efficiency, it is clear that there was both lower recovery and integration of the tagged ligases, in comparison to the remaining two cell lines. This complicates the interpretation of activity of the REL1CD/2ID chimeric ligases, since the amount of these recovered ligases used in each activity assay is less in comparison to REL1-TAP or REL2CD/1ID-TAP.

2.4 General Discussion

Results gained from this Chapter suggest a more complicated scenario than originally anticipated (see Figure 2.1). REL1-TAP was the only ligase expressed that restored the growth phenotype, caused by loss of regulatable REL1. This was as hoped, and provided a positive control to the study, demonstrating that this approach could be pursued. As expected the REL2-TAP ligase could not rescue the growth phenotype caused by REL1 ablation, and provided a negative control to this study. The chimeric ligases provided a means to involve REL1 in the addition editing subcomplex and REL2 in the deletion editing subcomplex

Once two new REL1CD:2ID chimeras were constructed and were shown to integrate correctly in RECCs, the repeated growth curve for four separate clones also revealed no rescue of the growth phenotype on ablation of regulatable REL1. All growth curves were constructed using the parental cKO as a comparison and to validate the growth curves. This indicated that all growth phenotypes, typically appearing around 60 hours, were caused by the loss of REL1. Although a constitutively expressed epitope tag may in theory cause cytotoxicity, the highly expressed REL1-TAP protein fully rescued the growth defect of the parental cKO cell line, demonstrating that it did not affect REL1 function significantly (Medina *et al.*, 2000).

REL2-TAP, REL1CD$_{333}$/2ID-TAP and REL1CD$_{322}$/2ID-TAP did not integrate as efficiently as the TAP ligases containing the REL1 interactive domain. This is most likely due to the presence of endogenous REL2. However, knowing that REL2CD/1ID-TAP can integrate as efficiently as REL1-TAP provides further

evidence that REL1 is essential at its deletion subcomplex position. In this scenario the REL2 CD, when present in the deletion subcomplex, cannot rescue the growth phenotype caused by REL1 ablation. Taken at face-value (from growth curves and miniTAP Western blots) a REL1CD is required at the deletion subcomplex for normal functioning of RNA editing and growth.

Glycerol gradient sedimentation of cleared cell lysate, and subsequent miniTAP and Western analysis confirmed TAP tagged ligases were correctly integrated into the subcomplexes for which they were intended. Although the KREPA3 interacting protein could not be detected in each of these pooled glycerol gradient fractions, it could be detected in the mitoprep positive control run alongside these samples (see MP in Figure 2.9). Although its absence could reflect the TAP tag masking the core of the 20S editosome, KREPA3 could be detected in REL1-TAP expressing cell fractions in previous studies (Schnaufer *et al.*, 2003). The components of the adjacent deletion subcomplex, however, could be detected from pooled TAP eluates from glycerol gradients (Figure 2.7 and 2.8) using REL1 and chimeric REL2CD/1ID cell lines. In the remaining cell lines expressing a ligase with a REL2 ID, only the subcomplex where integration occurred could be detected in the affinity purified fractions. This may indicate instability of the 20S editosome in the absence of a ligase integrating into REL1s position in the deletion subcomplex. This has also been reported by some groups (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002), but not others (Stuart *et al.*, 2002; Gao and Simpson, 2003). This discrepancy may be explained in terms of editosome stability (which may be due to the absence of REL1) and the length of experimental procedure, comprising of 9-hour glycerol gradient sedimentation, followed by fractionation and subsequent miniTAP. If there

were any destabilisation effects caused by REL1 loss then they would be more apparent after glycerol gradient and subsequent TAP than after TAP only.

Next, the activities of the ligases needed to be taken into consideration. Unusually the pyrophosphate treatment did not increase the auto-adenylylation activity of REL2-TAP, or the REL2CD:1ID-TAP, to levels comparable with REL1-TAP, even though in the literature auto-adenylylation activity of REL2 exceeds that of REL1 after deadenylylation (Cruz-Reyes *et al.*, 2002). The conditions of the deadenylylation assay used were as described by (Cruz-Reyes and Sollner-Webb, 1996; Cruz-Reyes *et al.*, 2002) and fresh pyrophosphate was used for each experiment (as suggested by Jorge Cruz-Reyes, personal communication). Together, this indicates that deadenylylation did not work as expected.

The activity measured for the isolated tagged ligases in the TEV eluates needs to be normalised for the amount of ligase present in these samples. Unfortunately, due to technical constraints, these blots could not be quantified. However, it was obvious that more TAP tagged ligase was present in the eluates from the cell lines expressing the REL1 ID. $REL1CD_{333}$/2ID-TAP and $REL1CD_{322}$/2ID-TAP ligases are also recovered more efficiently in the presence of endogenous REL1, suggesting that its presence improves integration into the editosome and pull down, possibly through a stabilising effect on the whole of the editosome, at least under purification conditions (results not shown). All chimeric ligases were much less adept at auto-adenylylation in comparison to REL1, so strong conclusions cannot be made, since the growth phenotype, at least in part, may reflect their relative catalytic activites.

Most studies to date have focused more on the essential motifs of the RNA ligases, and have not addressed the role of the interactive domain in governing catalytic activity, through a stabilisation effect for example. It is conceivable that the recombination of catalytic and protein-protein interaction domains (N- and C-terminal parts of the protein, respectively) in the chimeric ligases interfered with activity. This would in itself explain the lower activity of these ligases in auto-adenylylation reactions. A study of REL1, directly assessing the activity of this catalytic domain, revealed that the N-terminal domain was not as catalytically active as the full length ligase in auto-adenylylation and complete ligation assays (Deng *et al*, 2004). However, it appears from mutational studies that the closest relative of REL1, T4Rnl2, has a C-terminal domain that is dispensable for catalytic activity when the adenylylation step is bypassed, but has an N-terminal domain which retains activity after its isolation, albeit with a different pH optimum (Ho *et al*, 2004). The primary structure of this C-terminal domain is partially conserved between T4Rnl2 and the RELs, which do not possess the OB fold domain, present in DNA ligases (Ho *et al.*, 2004; Nandakumar *et al.*, 2004). Further investigation into T4Rnl2 has revealed that this C-terminal domain is required for RNA substrate specific activities, whereas the N-terminal domain has been implicated in the first (adenylylation) and last (strand sealing) steps of ligation (Nandakumar and Shuman, 2004; Nandakumar and Shuman, 2005; Nandakumar *et al.*, 2006). This may explain why the chimeric ligases were not as active as their full length counterparts, since the domain swap may have interfered with their catalytic function. It is not assumed that the tag itself reduced ligase activity, since REL1-TAP was able to restore growth phenotype in the absence of ectopically expressed REL1.

2.5 Conclusions and Outlook

The results for this part of the project are not as clear cut as hoped. Because of the reduced catalytic activity of the chimeric ligases, it can only be concluded that REL1 is essential at its position in the deletion subcomplex. However, the negative data produced from the chimeric ligases constructed has provided interesting information regarding the apparent cross talk between the CD and IDs of the RELs.

These cell lines may still be used in the future. To alleviate the potential fragility of the editosome, when no ligase is expressed in place of ectopic REL1 in the deletion subcomplex (as with REL2, REL1CD$_{333}$/2ID-TAP and REL1CD$_{322}$/2ID-TAP expressing cell lines), an inducible and catalytically dead copy of REL1 could be expressed in its place. With added stabilisation, the glycerol gradients and subsequent TAP and Westerns may reveal more components of the editosome. It may also improve integration and recovery for the auto-adenylylation assays, but it would not address the activity of the ligases. The Schnaufer laboratory also has a FRET-based fluorescent assay for measuring REL1 activity, which could be employed as a measure of full round ligase activity. The equivalent amount of REL1-TAP ligase eluate used in the adenylylation assays could be detected (results not shown) and not only would this approach be more sensitive, but it would also measure more than the first step of the ligation activity. However, the limitation of this approach is that the activity ascertained would be additive of multiple ligases within a complex. Although auto-adenylylation involves the first part of the editing reaction only, it is useful in assigning activity to separate ligases. First, however, deadenylylation and adenylylation assays must be properly established. It would also be important to isolate subcomplexes and subject them to *in vitro* addition and

deletion editing specific assays. This approach would use pre cleaved RNA substrates to discern chimeric ligase involvement in the restoration of editing, which may be very slight due to low activity. Another approach would be to discern the *in vitro* roles of chimeric ligases in editing at their respective subcomplex position through the titration of ATP and PPi (as used by Cruz-Reyes *et al.*, 2002). This method would comprehensively determine whether the catalytic function or subcomplex positioning is key to REL1's essentiality.

This study also provided indirect evidence for the cross-talking of REL CDs and protein-protein ID, which has not been extensively studied, and illustrates an interesting observation that may be pursued further through mutational and deletional study.

Chapter 3

Identifying the substrates of REL1 using limited and deep sequencing

3.1 Introduction to project

The identification of REL1 substrates may provide clues to its precise role, and therefore the basis for its essentiality to the process of RNA editing. In this Chapter we used cKO REL1 cell lines (the parental cell line, Chapter 2) to grow cultures of trypanosomes with normal and substantially reduced levels of REL1, with the aim to isolate and analyse RNA. Specific mitochondrial transcript substrates were sequenced, by the use of a 5' RNA linker ligated post isolation, which took advantage of the 5' monophosphate produced as a result of endonucleolytic cleavage. These substrates are readily ligatable to such a linker by an enzyme that can join single stranded RNA, like T4Rnl1 (Romaniuk and Uhlenbeck, 1983; Tessier *et al.*, 1986). By mapping the 5' ends of these ligation substrates it is possible to see what cleavage products remain unligated, and from this deduce the specificity of REL1. A similar 5' trapping approach was successful in other studies (Bruderer *et al*, 2003; Granneman *et al*, 2009).

One of the aims of this part of the study was to infer whether REL1 could be functioning as a general RNA repair enzyme (5' PO_4 to 3' OH) for erroneously cleaved RNA substrates, in addition to its role in sealing correctly edited sites. Such an idea is not a new one; indeed there have been several comparisons of REL1 to T4Rnl2, which was proposed to have general RNA repair activity (Ho and Shuman, 2002; Palazzo *et al.*, 2003; Yin *et al.*, 2003; Ho *et al.*, 2004; Nandakumar *et al.*, 2004; Deng *et al.*, 2004). Central to the experimental design for the capture of editing intermediates was also the removal of abundant ribosomal RNA, degradation products of which may also have 5' monophosphates (Ryan *et al.*, 2003). The approach described here, however, selectively amplified mitochondrial transcripts by

use of specific reverse primers, and in doing so, eliminated the chances of sequencing undesirable RT-PCR products. Whilst a global approach was not achieved in the time frame given with this PhD study, this approach still allowed specific questions pertaining to REL1 substrates, such as: what particular editing events does REL1 govern? And how frequently do these occur?

The outline of the sequencing strategy is shown schematically overleaf in Figure 3.1.

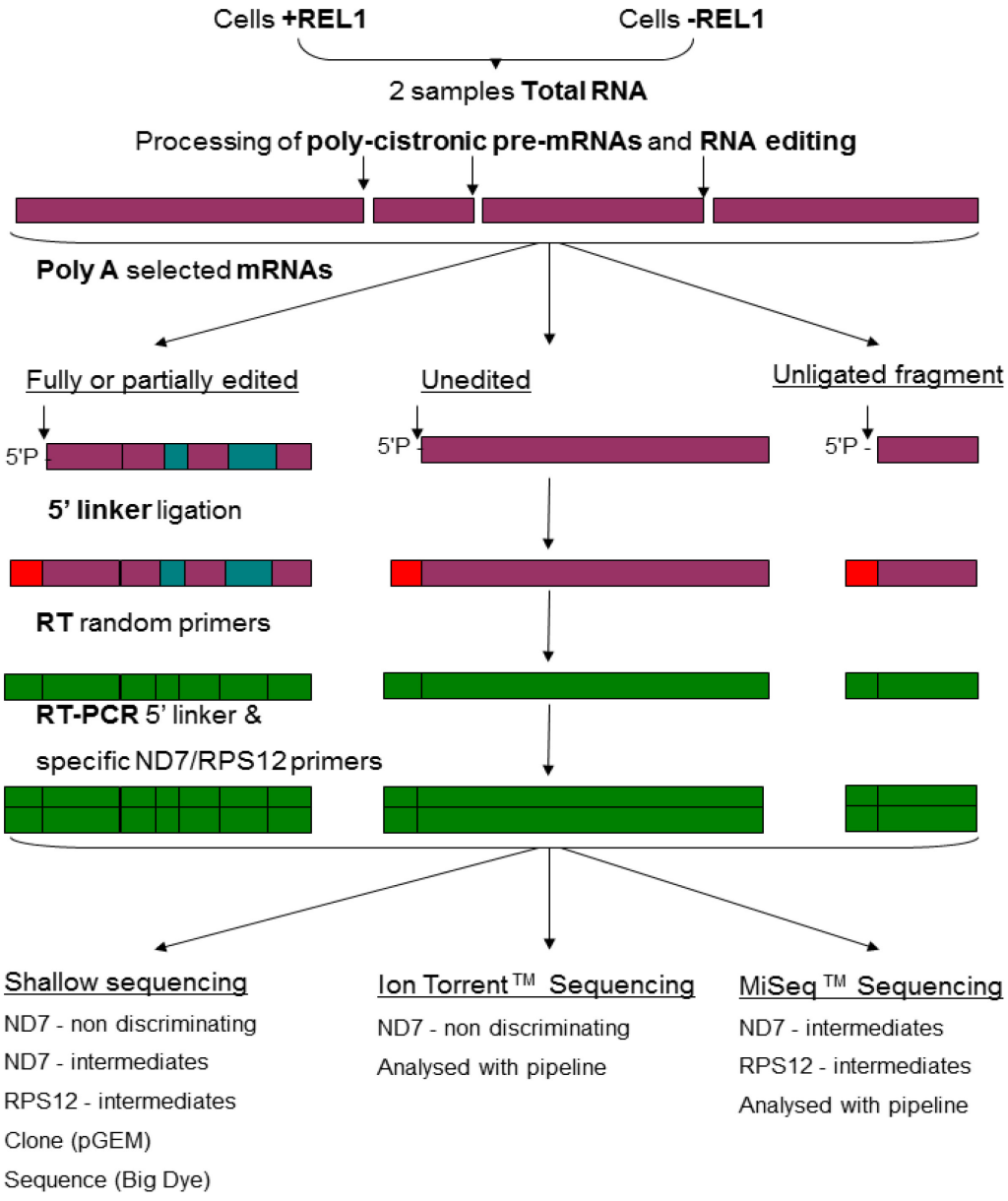


Figure 3.1. Sequencing workflow

Approximately 5 × 10 ^{8} cells (equating to 500 ml cultures in logarithmic growth) were cultured for each of the RNA preps. RNA was isolated using TRI reagent ® solution instead of a column based RNA isolation kit, since columns incur a size bias to the RNA population extracted. 5' ends were mapped with the use of an RNA linker, which was ligated to the 5' monophosphate product of endonucleolytic cleavage during the processing of polycistons, or editing substrates.

Colour scheme: red – 5' RNA linker, purple – unedited region of transcript, blue – edited region of transcript, green – DNA generated from RT and PCR.

The reaction condition for 5' linking and subsequent RT using tagged hexamer primers were adapted from Granneman *et al.*, 2009 and Matsumoto *et al.*, 2010.

3.2 Methods

3.2.1 Cell culture

CKO REL1 cell lines (used as a parental cell line in Chapter 2, Figure 2.2, Schnaufer *et al.*, 2001) were maintained in HMI-9 media containing 10% (v/v) FCS and selective drugs (2.5 μg/ml G418, 5 μg/ml hygromycin, 2.5 μg/ml phleomycin and 1 μg/ml tetracycline). Log phase cells were cultured to 500 ml, both in the presence and absence of tetracycline, and counted and harvested after 48 hours. Pellets were kept on ice prior to the addition of 1 ml of TRI reagent ® (Ambion) per 1×10^8 cells grown in the presence and absence of tetracycline (+tet and –tet). TRI lysates were left at -80°C until required for RNA extraction. Aliquots of whole cell lysates at 48 hours, for cells grown in the presence and absence of tetracycline, were kept for Western blot analysis.

3.2.2 RNA extraction

For all RNA work, benches were cleaned thoroughly with 70% ethanol and RNase*Zap*® (Ambion), and filter tips were used throughout to reduce chances of RNase contamination. Phenol-chloroform clean up and TRI Reagent extractions were performed in the fume hood, and a double layer of gloves were worn as a precaution. Autoclaved DEPC water was used instead of distilled or Milli Q water with all of the RNA work up to the final PCR reactions involved in generating sequencing material.

RNA from approximately 6×10^8 pelleted trypanosome cells was isolated using TRI reagent ® solution and its associated standard protocol (Applied

Biosiences). Briefly; cells suspended in TRI reagent were left to thaw and incubate at room temperature for 5 minutes before the addition of 130 µl and 630 µl of Bromo-chloro-3-propanol (BCP) for +tet and –tet samples respectively. Lysates were left to incubate at room temperature for 15 minutes before centrifugation at 12,000 g for 15 minutes at 4ºC. The aqueous phase was transferred to a fresh tube and 650 µl and 3.15 ml of isopropanol was added for + tet and –tet samples, respectively. The mixtures were briefly vortexed and left to incubate for 10 minutes at room temperature before the RNA was pelleted by centrifugation at 12,000 g for 10 minutes at 4°C. The RNA pellets were washed in 10 ml 75% ethanol by centrifugation at 12,000 g for 10 minutes. The pellets were allowed to air dry before their resuspension in 100 µl DEPC water.

3.2.3 DNA clean up of total RNA

RNA was treated with DNA-free DNase (Ambion), as specified by the kit before the lysate was subjected to two rounds of poly(A)+ RNA selection using the standard protocol of MicroPoly(A)Purist™ kit (Ambion). Two parallel reactions were set up for Ambion® DNA-free™ (Life) DNase Treatment with 40 µl total RNA, 4 µl 10 x DNase I Buffer and 0.5 µl rDNase I. After gentle mixing, the reactions were incubated at 37°C for 30 minutes, before the addition of a further 0.5 µl rDNase I and subsequent incubation at 37°C for 30 minutes. After the DNase Inactivation Reagent was resuspended by vortexing, 4 µl was added to the RNA reactions to stop the reaction and this was subjected to continual mixing by pipetting for 2 minutes at room temperature. To finish, reactions were centrifuged at 10,000 × g for 1.5 min

and transferred to a clean tube to await poly(A)+ RNA purification on oligo(dT) beads.

3.2.4 Purification of poly(A)+ RNA using Poly(A) Purist™ kit

To minimise loss of RNA throughout this process the DNA free cleaned RNA was added directly to the oligo(dT) column without prior ethanol precipitation. Aside from this, the protocol was carried out as per manufacturer's instruction. Briefly, an equal volume 2 × Binding Solution was added and mixed thoroughly and each RNA sample was added to 1 tube oligo(dT) cellulose and mixed well by inversion and pipetting. The tubes were incubated for 5 min at 65 °C to denature secondary RNA structures, before leaving them for 1 hour at room temperature rotating.

The oligo(dT) cellulose was pelleted by centrifugation at 3,000 × g for 3 min at room temperature, and the flow through was collected and kept on ice until the end of the procedure to ensure good poly(A)+ RNA recovery. The cellulose was initially resuspended by vortexing in 500 µl Wash Solution 1, before transferring to a spin column to aid removal of non-specifically bound material by centrifugation at 3,000 × g for 3 min at room temperature. The flow through was discarded. The beads were washed once more with 500 µl Wash Solution 1 and three times with Wash Solution 2 in the same manner. Spin Columns containing the bound poly(A)+ RNA were placed into new microfuge tubes and 200 µl of warm (70°C) RNA Storage Solution was added to the Oligo(dT) Cellulose. Tubes were briefly vortexed and immediately centrifuged at 5000 × g for 2 min to elute the RNA. This process of elution was repeated. This was immediately followed by a second round of oligo(dT) selection

using the cellulose present in the spin column. The 350 µl of the 400 µl eluted RNA was added to 350 µl of 2 × Binding Solution and the denaturation, rotating incubation, washes and elution were carried out as before. All aforementioned solutions were provided by the kit. RNA was then ethanol precipitated by the addition of 1 µl glycogen, 0.1 volumes 3 M sodium acetate pH 5 and 2.5 volumes 100 % ethanol. Reactions were vortexed briefly before incubation in a –80°C freezer for 30 min. After this time, poly(A)+ RNA was washed in 200 µl of 75% ethanol and pellets were left to air dry before being resuspended in 20 µl DEPC water. Both total and poly(A)+ RNA were quantified and assessed for purity using the Agilent Bioanalyzer system. Simultaneous removal of ribosomal RNA and enrichment of mRNA transcripts was assessed by Northern blotting, using a riboprobe for actin mRNA (a kind gift from the Matthews lab). The Northern blotting protocol described below was modified from one described in Chapter 7 of Maniatis *et al.*, 1982 and used more recently in (Mayho *et al.*, 2006).

3.2.5 Northern Blotting

For this procedure all containers were washed thoroughly with detergent, rinsed with distilled water, sprayed with 70% ethanol and left to air dry before use. All base solutions used in the Northern blotting procedure, which were not provided with the corresponding kit, may be found in Appendix 2.

A 1.2% 3-(N-morpholino)propanesulfonic acid (MOPS)-based agarose gel containing 1% formaldehyde was poured and allowed to set in a fume hood. The gel was loaded onto the gel with 1 µg and 300 ng of total and poly(A)+ RNA,

respectively, in 5 µl DEPC along with the following: 9 µl formaldehyde, 3 µl 37% formamide, 2 µl MOPS and 2 µl RNA loading buffer . Samples were denatured at 65°C for 5 minutes, before running for 90 minutes on the RNA Agarose gel at 150 V in 1 × MOPS running buffer. The gel was post-stained with 0.5 µg/ml EtBr for 15 minutes, destained twice for 30 minutes with distilled water, all at room temperature, and visualised using a UV transilluminator (Syngene) to determine loss of ribosomal bands in the poly(A)+ RNA when loaded in an equivalent concentration alongside total RNA.

RNA was transferred onto a positively charged membrane (Roche) through capillary action overnight. The blot was assembled in the following manner and time was taken to ensure bubbles were removed at each stage by rolling a 10 ml pipette over the layer in question. A large tray was filled with 10 × saline-sodium citrate (SSC), and a wide sheet of Perspex was placed over the tray, leaving a gap either side for the pre-wetted Whatman chromatography paper (Fischer Brand) wick which was subsequently placed on top of the Perspex. The gel was next laid down, with parafilm surrounding it, to prevent drying overnight, followed by the membrane (pre-wetted in 10 × SSC) and 2 gel sized Whatman filter papers (pre-wetted in 2 × SSC). Finally a 6 inch layer of tissues were added on top, followed by a hard backed text book and the RNA was allowed to transfer overnight. The following day the nylon membrane was allowed to dry and then UV cross linked in the Stratalinker for 1200 counts or 0.12 joules.

After cross linking the RNA to the membrane, the blot was transferred to a hybridisation tube and subjected to pre-hybridisation incubation with 10 ml of DIG hybridization buffer (Roche). After this time, the 10 ml of hybridization buffer was

poured away and replaced with 7 ml fresh buffer and 1 µl actin riboprobe, prepared by DIG RNA Labelling Kit (SP6/T7) (Roche) (kind gift from the Matthews lab) and was left to incubate overnight at 68 ˚C. The blot was then subjected to two 30 minute washes with 2 × SSC/0.1% SDS and one 30 minute wash with 0.5 × SSC/0.1% SDS, all at hybridization temperature. The blot was transferred from a hybridisation tube to a tub and washed for 1 minute at room temperature with wash buffer (Maleic acid buffer + 0.03% Tween 20), before a 1 hour incubation at room temperature with Maleic acid buffer with 1% DIG Block in a new tub. After blocking the membrane, to prevent any non-specific binding, it was transferred to a new tub and incubated for 30 minutes with 50 ml of Maleic Acid buffer with 1% DIG Block and Anti-DIG*. Subsequently, the blot was washed 6 times for 5 minutes with wash buffer on a rocker, before soaking for 2 min in detection buffer. After excess detection buffer was removed, the membrane was placed in a heat sealable bag (Jencons) and 1 ml of CDP-star detection agent (10 µl of substrate in 1 ml Detection Buffer) was added. After a two-minute incubation, excess liquid was removed and the bag was heat sealed and left at 37˚C for 15 minutes. The blots were visualised using X-ray film.

3.2.6 Nano Agilent Chip analysis of total and poly(A)+ RNA

To prepare for running a nano chip one of the wells of an electrode cleaner chip was slowly filled with 350 µl RNaseZAP (Ambion), which was then subsequently placed into the Agilent 2100 bioanalyzer, and left with the lid closed for 1 minute. This action was repeated with another electrode cleaner containing 350 µl of RNase-free DEPC water, which was left for 10 seconds with the lid closed and another 10

seconds with the lid open for the water on the electrodes to evaporate before closing the lid.

The gel matrix was prepared as follows. All reagents in the Agilent RNA nano kit were left to equilibrate to room temperature in the dark for 30 minutes before use. The RNA 6000 Nano dye concentrate was vortexed briefly and spun down, and 1 µl of this was added to a 65 µl aliquot of pre-filtered Agilent RNA 6000 Nano gel matrix and the mixture was centrifuged for 10 minutes at room temperature at 13000 × g. The Nano chip was then inserted into the chip priming station, with the base of the plate of the station set to position (C), in preparation of sample loading. Once the matrix-dye mixture had equilibrated at room temperature 9 µl of it was added at the bottom of the well marked (G). Making sure that the plunger was positioned at 1 ml, the chip priming station was closed until the latch clicked and the plunger of the syringe was pressed down until it was securely held by the clip. After 30 seconds the plunger was released with the clip release mechanism. After 5 seconds, the plunger was slowly pulled back to the 1 ml position. On opening the priming station 9 µl of the gel-dye mix was pipetted slowly into each of the wells marked G, and 5 µl of the RNA 6000 Nano marker was pipetted into the well marked with the ladder symbol and into each of the 12 sample wells. Wells due to be empty were filled with 5 µl of the RNA 6000 Nano marker plus 1 µl of DEPC water. Ladders were thawed and kept on ice prior to analysis. To minimize secondary structure, samples were heat denatured (70°C, 2 minutes) before loading on the chip. The chip containing 1 µl of each sample in each of four sample wells was horizontally placed in the adapter of the IKA vortex mixer and vortexed for 1 minute at 2000 rpm. The chip was then immediately inserted in the Agilent 2100

Bioanalyzer, which had the chip selector in position (1), the lid was closed and the analysis run using the Expert software.

3.2.7 5' linkage of poly(A)+ transcripts

100 pmol of a 5' RNA linker (5-/InvddT/GTTC*AGAGUUCUACAGUCCGACGAUC*-3, a kind gift from the Tollervey lab - Granneman *et al,* 2009) was ligated over night at 16°C to 100 ng poly(A)+ RNA in an 80 µl reaction with ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, pH 7.0), 80 u RNasin (Promega) and 40 u T4 RNA ligase 1 (NEB). Italicised nucleotides refer to RNA bases in the 5' linker. RNA was precipitated as follows: 20 µg glycogen, 1/10 volume 3 M Na acetate and 2 volumes 100% EtOH were added and the mixture incubated at -20°C for 30 minutes, before centrifugation for 15 minutes at 11,000 × g in a bench top centrifuge (Technico Maxi). Supernatant was discarded and the pellet washed in 200 µl 70% EtOH before resuspension in 20 µl of DEPC-treated water.

3.2.8 First strand synthesis

RNA was subject to incubation at 65°C for five minutes with 125 ng of random hexamer primer (5'-CCTCTGAAGGTTCACGGATCCACATCTAGANNNNNN), to maximise hybridisation. The tag had been introduced to facilitate global mtRNA analysis, which was unsuccessful. This was used in a 20 µl reaction mixture containing 1st strand buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 20 U RNase OUT, 0.2 mM DTT, 200 U Superscript II (Invitrogen) and 10 mM

dNTPs was subjected to the following temperature programme: 25°C 15 minutes, 42°C 50 minutes and 70°C for 15 minutes. 1 µl of RNaseH was added to this, and the reaction was subjected to two further incubations; 37°C for 20 minutes and 90°C for 10 minutes. RT reactions were cleaned using phenol-chloroform extraction. Briefly, RT reactions were made up to 100 µl with DEPC and an equal volume of phenol-chloroform was added in Eppendorf tubes. Tubes were vortexed briefly and spun at 10, 000 g for 1 minute and the aqueous layers were transferred to new tubes. A further 100 µl of DEPC was added to each phenol-chloroform mixture and the aqueous layer extracted as previously described. Extracts in DEPC were precipitated in the presence of glycogen, as described previously, and cDNA pellets were resuspended in 20 µl DEPC-treated water.

3.2.9 Second strand synthesis

Second strand cDNA synthesis was performed in a 25 µl reaction with 200 pmol of 5' linker-specific primer (5'-GTTCAGAGTTCTACAGTCCGACGATC), 200 pmol of either a non-discriminating ND7 primer (ND71) (5'-CGGAAGACATTGTTCTACAC), a primer designed to amplify ND7 transcripts that had already entered the editing cycle (ND72) (5'-GTACCACGATGCAAATAAC) or an RPS12 specific primer (5'-AAAAACATATCTTATATCTAAA). PCR reactions also included 1 U GoTaq® DNA polymerase (Promega), 200nM dNTPs, 1.5 mM MgCl2 and 1 × flexi buffer (pH 8.5). PCR reactions were subjected to the following temperature program: 94°C 2 minutes, 35 cycles of 94°C 30 seconds, 50°C 1 minute, 72°C 2 minutes, and 72°C

7 minutes. The regions of ND7 where primer sets anneal are illustrated in Figure 3.5.

3.2.10 Cloning of sequences for limited sequence analysis

PCR reactions were run out on a 2% agarose gel at 100 V for 2 hours. Discrete bands or smears were excised (see Figure 3.6) using Gene Catcher disposable gel excision tips. DNA was purified using Nucleospin II kit (Macherey-Nagel). These sequences were cloned into pGEM®-T Easy (Promega), as per manufacturer's instructions and sent for Sanger sequencing (big dye reaction) at the GenePool faculty of Edinburgh University.

Sequences were aligned manually in Microsoft Word.

3.2.11 Preparation of 5' linked RT-PCR products for Ion Torrent™ and MiSeq™ Illumina based sequencing

RT-PCR products were generated as described in Sections 3.2.7 to 3.2.9. Two lots of 25 µl RT-PCR reactions were set up for each of the three primer sets described above and 1/5th of the reactions were analysed on a 2% agarose gel and purified using the Nucleospin II kit before being sent to the Western General Hospital for analysis using Ion Torrent (ND71- non discriminating) or the GenePool (University of Edinburgh) for MiSeq™ sequencing (RPS12 and ND72 - intermediates). Before Ion Torrent™ sequencing, the RT-PCR library prep was cleaned using an AMPure® XP kit (Beckman Coulter. Inc). This kit purified sequencing material with a cut-off of 100 bp, as an upper cut-off of 300 bp would not have been desirable. The 100 bp cut-

off allowed the removal of sequences lacking ND7 transcripts. The Ion Torrent P1 and A adapter sequences were 5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3' and 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3', respectively. Since +tet and –tet samples were processed on the same chip, they were differentially barcoded with IonXpress11 (TCCTCGAATC) and IonXpress12 (TAGGTGGTTC), respectively.

The process of Ion Torrent is shown in Figure 3.2. Ion torrent sequencing provided 300 bp single (non-paired) end reads, which had either forward (category 1) or reverse orientation (Category 2). MiSeq produced 250 bp paired-end reads, that either spanned the length of the transcript (Category A), or partially overlapped (Category B).

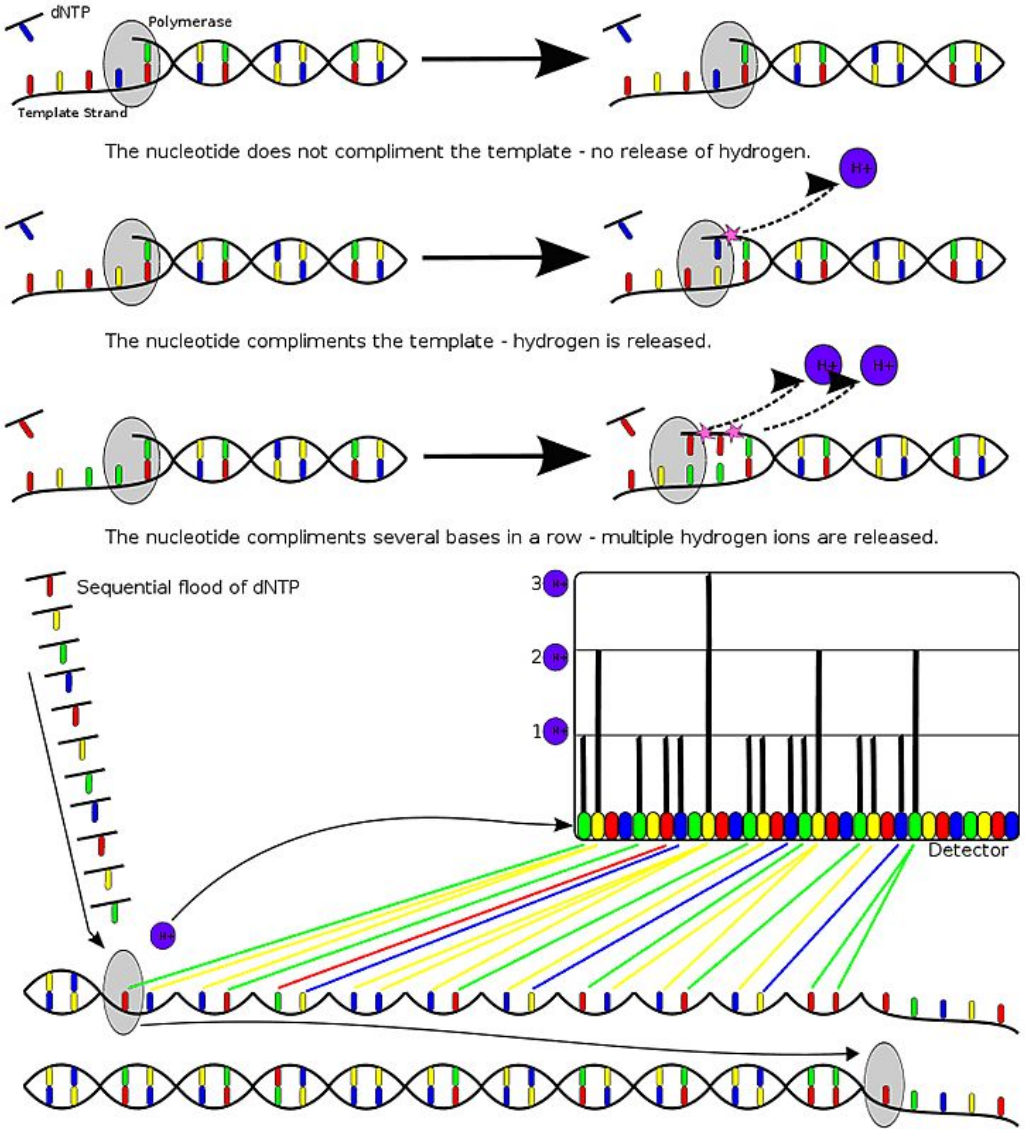


Figure 3.2. The process of Ion Torrent sequencing

This method works on the principle that a nucleotide match against a stretch of complementary sequence causes a release of a hydrogen atom and a subsequent drop in pH, which can be measured. From the peaks measured the sequence can be deciphered. The 5' linked RT-PCR products used in this study were subjected to a 100 bp cut off clean up and barcoding. Taken from Wikimedia Commons.

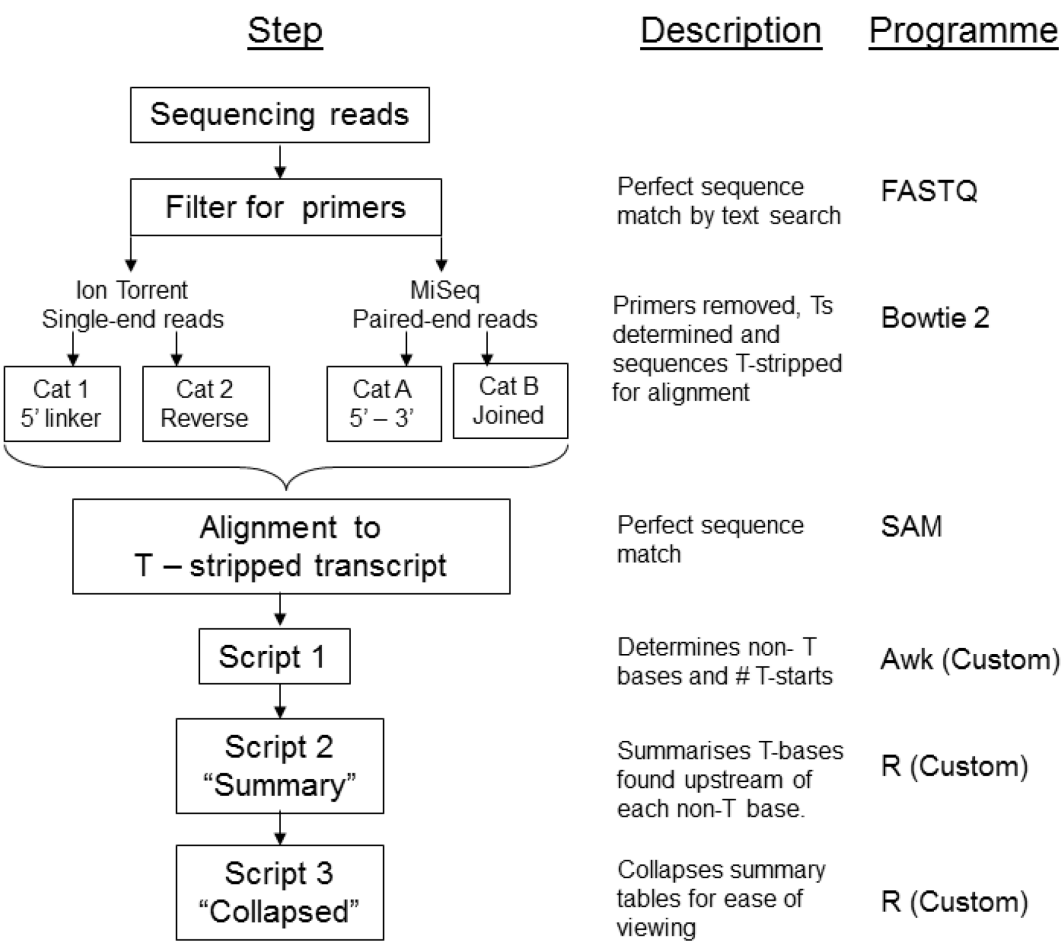
3.2.12 Analysis of editing intermediates from Ion Torrent™ and MiSeq™ sequencing

Sequences were subject to bioinformatics sorting using an R-based pipeline devised by Al Ivens (CIIE). The workflow for this is shown in Figure 3.3. Tables that were finally generated consisted of the number of T-stripped sequences. Since we wanted to determine the number of uridylyls added or deleted during the editing process as well as define the type of editing sites captured by 5' linker at a particular position in the sequence, we first had to align reads from RT-PCR to a sequence without T reads. Read data was separated into category 1 and 2 (Ion Torrent analysis) or A and B (MiSeq), depending on how they were processed (see Figure 3.3 for more details). Ion Torrent category 1 and 2 data were analysed separately for analysis of percentage of editing type at each nucleotide site, and the former was used to map the true 5' ends of the transcripts captured by RNA linker. Since paired end data was retrieved from MiSeq analysis both categories A and B were combined where applicable for mapping the 5' ends of transcripts and for determining the percentage of editing at each nucleotide site.

Firstly, the number of sequences pertaining to the position of the 5' linker (i.e. the most 5' end) for category 1 were mapped. Secondly, the average number of uridylyls added before the position on the T-stripped ND7 sequence was plotted. These graphs were constructed using Graphpad Prism (version 6). Finally, the percentages of unedited, partially edited and correctly edited sequences were calculated for each position of the stretch of ND7 sequence in excel. From this data graphs were plotted using R.

3.2.13 Quantitative real-time PCR (Q-PCR) to ascertain RT efficiency of amplification in +REL1 and -REL1 samples

RT cDNA products (made as described in Sections 3.2.2 - 3.2.5) were subjected to a 1/10 dilution, before setting up 25 µl PCR reaction with 15 µl SYBR® green (Applied Biosciences) and 2.5 µl cDNA mix and 12.5 µl of a 1.5 mM β-tubulin primer mix (F'-TTCCGCACCCTGAAACTG, R'-TGACGCCGGACACAACAG). Reactions +REL1/+RT, +REL1/-RT, -REL1/+RT and –REL1/-RT were set up in triplicate alongside a cDNA control amplified using an 18S primer mix (F'-CGGAATGGCACCACAAGAC, R'- TGGTAAAGTTCCCCGTGTTGA) to determine is the Q-PCR reaction worked as efficiently. Reactions were subjected to the following thermal conditions: 50ºC 120 s followed by 95ºC 10 minutes, followed by 40 cycles of 95ºC 15 s and 60ºC 60 s on an ABI prism PCR machine. ΔΔCT was calculated for +REL1 and -REL1 RT from samples run in triplicate, using StepOnePlus Real-time PCR system software.



3.3. Bioinformatics workflow to sort sequences

Ion Torrent sequencing reads were filtered for the presence of the 5' linker (Category 1, Cat 1) or reverse primer only (Cat 2), which were analysed separately. Category 1 reads were used for 5' end mapping and both Category 1 and 2 were used to visualise percentage editing type at each nucleotide site. MiSeq produced paired end reads that either spanned the entirety of transcripts individually (Cat A) or could be joined to produce full-length sequences (Cat B). Category A and B reads were either analysed together, or category A only reads were analysed and were used to devise graphs for 5' end mapping and to visualise percentage editing type at each nucleotide site.

Here is also indicated the programs used to write corresponding part of the pipeline, and a brief description of what each pipeline step entailed.

The work flow was devised by Al Ivens (CIIE)

3.3 Results

3.3.1 Analysis of RNA for sequencing

Briefly, cells were grown to logarithmic 50 ml cultures in the presence of tetracycline, before washing tetracycline from the media and using 1×10^4 cells to inoculate 500 ml HMI-9 media containing 10% (v/v) FCS and selective drugs (2.5 μg/ml G418, 5 μg/ml hygromycin, 2.5 μg/ml phleomycin and 1 μg/ml tetracycline). 500 ml cultures were grown in the absence or presence of tetracycline for 48 hours before cells were harvested and used for RNA extraction. A small culture flask of cells was kept to ensure cells died off as expected in the absence of REL1. No live cells were seen microscopically after 6 days.

The Northern blot (Figure 3.4, A) shows the enrichment of mRNA transcripts exemplified by actin after poly(A)+ selection. The Western blot (Figure 3.4, B) reveals strong knock-down of REL1 (in the absence of tetracycline) after 48 hours, due to the absence of the ectopically expressed protein at 50 kDa.

The quality of the poly(A)+ RNA was assessed by Agilent chip. The traces (shown overleaf in Figure 3.4, C-F) show great reduction in the amount of ribosomal RNA in the poly(A)+ samples. This can be seen from the loss of the three peaks pertaining to rRNA (Figure 3.4 C and D) the Agilent chip trace in the range between 200 and 4000 nucleotides. Agilent chip analysis suggested total RNA was not badly degraded from the smooth baselines to the left of the ribosomal peaks, indicated on the traces. A peak around 100 bp perhaps indicates a corresponding increase in SSU RNAs and tRNAs.

The same biological replicate of RNA was used for all sequencing reactions.

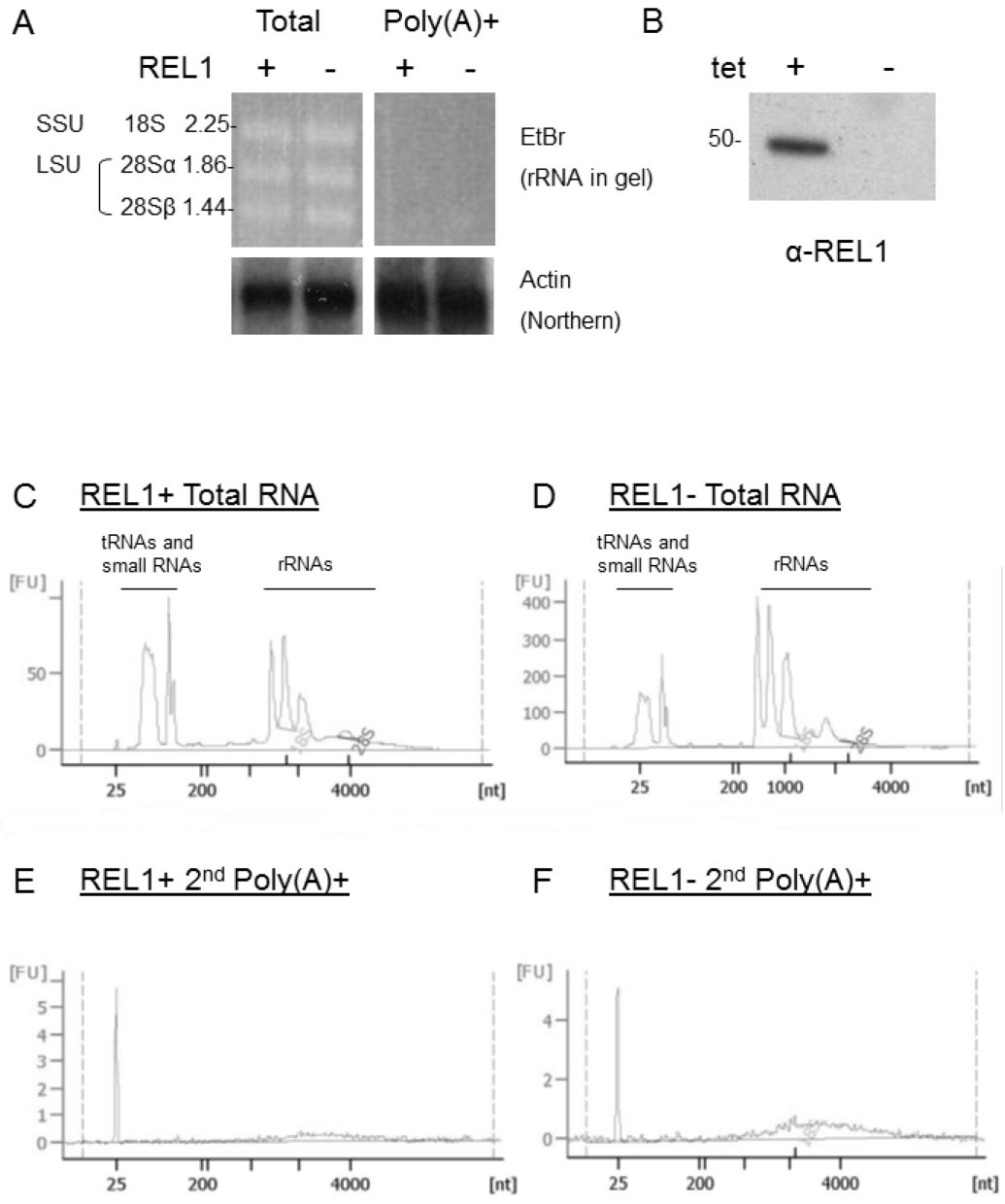


Figure 3.4. Northern blot and Agilent chip analysis for the RNA used in limited and Ion Torrent sequencing

A) 1 µg total RNA was run alongside an equivalent 300 ng poly (A)+ selected RNA and analysed by Northern blot. The increased signal from the actin probe confirms enrichment of mRNA transcripts. B) Western blot using 1000 α-REL1 and 1/2000 α-mouse antibodies indicates that REL1 protein is absent after 48 hours. C-F) Agilent electropherographs clearly show the diminution of rRNA specific peaks from total RNA (C and D) after two rounds of poly (A)+ selection (E and F).

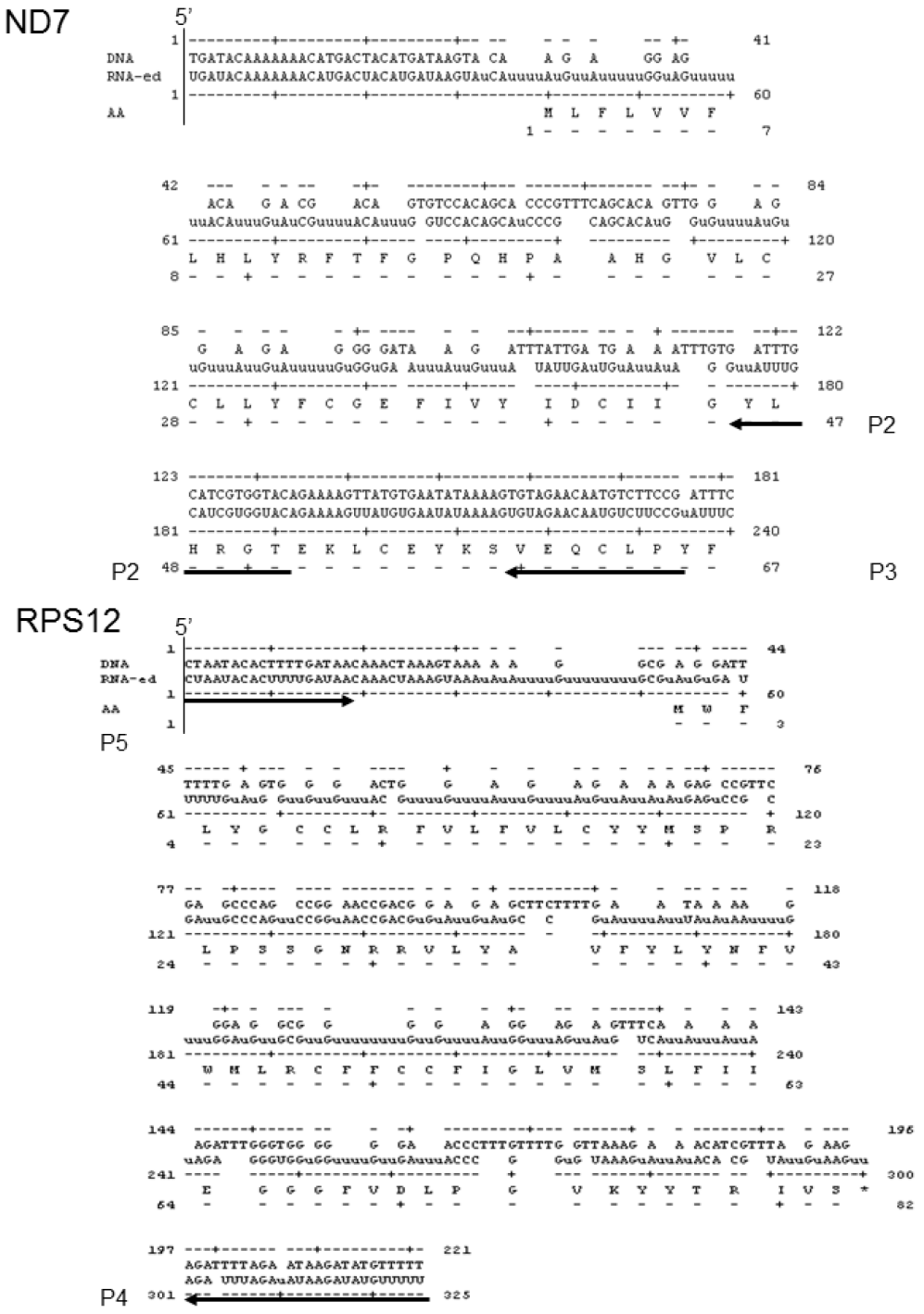


Figure 3.5. ND7 and RPS12 primers used in sequencing strategy

Above is indicated where primers P2 (5'-GTACCACGATGCAAATAAC), P3 (5'-CGGAAGACATTGTTCTACAC), P4 (5'- AAAAACATATCTTATATCTAAA) and P5 (5' CTAATACACTTTTGATAACAAAC) anneal on ND7 and RPS12 transcripts.

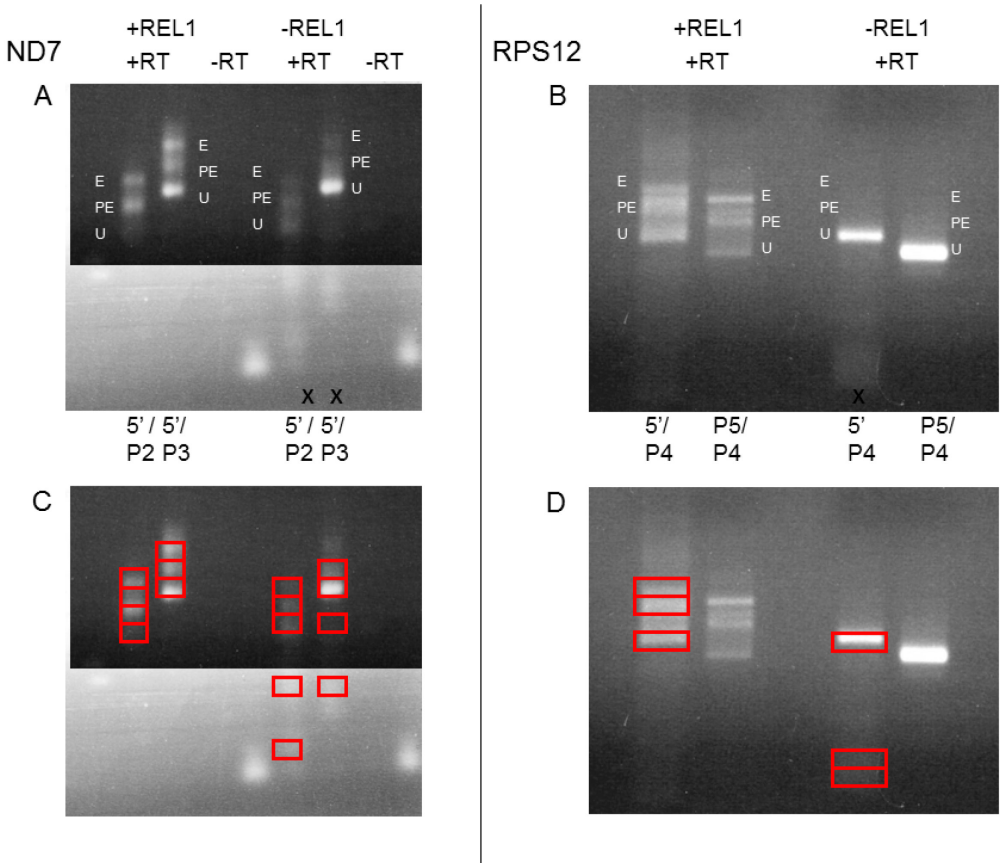


Figure 3.6. ND7 and RPS12 RT-PCRs visualised on a 2% agarose gel

A and B) E, PE and U indicate fully edited, partially edited and unedited (bar 1st editing site) transcripts of RPS12 respectively. Areas of ND7 and RPS12 picked up by specific primer are indicated in Figure 3.5. %' refers to a forward primer based on the 5' RNA linker. In the case of ND7, the reverse primer either selected for editing at the first two sites (P2) or there was no bias in the stage of editing in the transcript (P3). For RPS12 the reverse primer used (P4) contains the first editing site and hence selected for transcripts undergoing editing. Smaller fragments of interest are indicated with an X and may represent products of endonucleolytic cleavage which have not been relegated by REL1.

C and D) Red rectangles indicate bands which were excised, cloned and sent for Sanger big dye reaction sequencing. All manually aligned transcripts can be found in Appendix 3 All small fragments of interest are listed in Table 3.1.

3.3.2 Shallow sequencing: effect of REL1 ablation on ND7 and RPS12 mRNAs

PCR products obtained from Section 3.2.5 were resolved on a 0.8% agarose gel (see Figure 3.6). Bands produced a similar pattern to those in Schnaufer *et al.*, 2001, indicating the presence of partially, fully and unedited transcripts (as indicated by PE, FE, UE). Bands enclosed with a rectangle (right panel) were excised, gel purified, cloned into pGEM®-T Easy and sequenced. The full length sequence data, for sequences cloned from +REL1 and –REL1 RT-PCR products are shown in Appendix 3. Fragments of interest in the -REL1 samples are collated alongside those of RPS12 in Table 3.1.

From these sequences it can be seen that editing occurs generally from a 3' to 5' direction, although this was not exact (as demonstrated previously Decker and Sollner-Webb, 1990; Koslowsky *et al.*, 1991). Regardless of the presence or absence of REL1, sequences revealed extraneous uridylyls along the length of the sequence, even at non-editing sites. Within the ND7 edited region, linker ligation at deletion, addition and non-canonical editing sites was detected. The same substrates were detected up to three times. Within the RPS12 edited region only one example of a deletion editing event was detected, in comparison to eight misediting events. No examples of addition editing substrates were found on ablation of REL1, however, the 5' linker was found at insertion sites, where addition editing had already occurred. Of the fragments (see Table 3.1), there were three examples of sequence mutation within the ND7 and one example for RPS12. In the presence of REL1 (Appendix 3) the 18 ND7 cloned transcripts revealed two examples of correctly, fully edited sequences and 16 examples of partially, or misedited sequences.

Table 3.1. 5’ linked RNA fragments from ND7 and RPS12 specific RT-PCRs on RNA from cells grown in the absence of REL1

ND7	DNA	TA A G ATTTA TTG A TG A A ATTTGTG A	
	RNA	AuuuA uuG uuuA UA UUG A uUG uA uuA uA G G uuA	
Editing type	Deletion	-UA UUG A uUG uA uuA uA G G uuA	×1
		-G G uuA	×3
	Mis	-A UUG uA G uA G A G G uuA	×1
		-UG uuA .. AuuuAuuA UAG G uuA	×1
		-A uUG uA uuA uA G G uuA	×1
	Insertion	-uuA uA G G uuA	×2
		-uAuuA UG GGuuA	×1
		-uA uA G G uuA	×1
	DNA	C A C C C G TTTC A G C A C A G TTG G	
	RNA	C A uC C C G C A G C A C A uG G uG	
	Mis	-G C A uG C A C A uG G uG	×1
		-G C A C A uG G uG	×1
		-UG uG	×1
RPS12	DNA	TGTTTTG GTT A A A G A A A CA TCGTTTA G AAG AGA	
	RNA	G G uG U A A A GuAuuA uA CA CG UAuuGuAAGuuAGA	
Editing type	Deletion	-G UUG GUU uAuA uA GuA uA uuA CA CG UAuuGuAAGuuAGA	×1
	Mis	-ACUC uAuA uA GuAuuA uA CA CG UAuuGuAAGuuAGA	×1
		-CA CG UAuuGuAAGuuAGA	×1
		-uG U A A A GuA uA uA CA CG UAuuGuAAGuuAGA	×1
		-uG U A A A GuAuuA uA CA CG UAuuGuAAGuuAGA	×1
	Insertion	-uAuuuA CA CG UAuuGuAAGuuAGA	×1
		-uA CA CG UAuuGuAAGuuAGA	×2
		-uGuuuA CA CG UAuuGuAAGuuAGA	×1

Sequences were grouped into endonucleolytic products at deletion, addition and non-canonical editing sites and aligned against DNA and edited RNA sequences. No addition editing events were observed. Dashes indicate 5’ linker position. Correctly and incorrectly inserted uridylyl bases (Us) are highlighted in blue and red respectively. Underlined bases indicate areas where deletion editing should occur and bases highlighted in purple indicate unexpected non-U residues. The frequencies that the transcripts were encountered at are shown on the right.
Sequences with 5’ linkers attached addition sites had unexpected Us between linker and mRNA. Full length sequences can be seen in Appendix 3.

Interestingly, extra uridylyls (indicated in red) were clustered more tightly around the 5th and 6th deletion editing site (from the 3' end). For RPS12, no examples of correctly and fully edited sequences were found, in the six cloned examples, even though all had entered the editing process.

The obvious differences between the sequences obtained from +REL1 and –REL1 samples is the lengths of the transcripts obtained (See Appendix 3) and the state of editing that they were in. All +REL1 transcripts cloned and sequenced contained the most 5' and 3' end of ND7 and RPS12 at various states of editing. Full length –REL1 derived sequences remained unedited. The fragments presented in Table 3.1 were mostly at editing sites, although did not reflect cases of expected insertion editing substrates at addition sites and it was apparent that uridylyls were added and the transcript was cleaved subsequently in a misediting event.

3.3.3 Deep sequencing analysis of REL1 substrates through Ion Torrent ™ sequencing

Shallow sequencing revealed an interesting accumulation of unligated editing intermediates in the absence of REL1 produced by cleavage at non-canonical sites (misediting) and deletion sites. However, this approach provided a limited number of examples of editing events governed by REL1, so deep sequencing approaches were pursued to obtain a clearer picture. Cleaned RT-PCR products were visualised on a 0.8% agarose gel (Figure 3.7 A) revealing fully, partially and unedited transcripts (labelled E, PE, U and X respectively). RT-PCR products for +REL1 and –REL1 samples were cleaned up further after library prep to remove unincorporated adapters

(see Agilent electropherographs, Figure 3.7 B). Agilent traces here reveal a population of smaller transcripts on both samples under 100 bp. These peaks may correspond to the adapter and 5' linker. Transcripts totalling more than 100 bp were more abundant in –REL1 samples in comparison to +REL1 samples, indicative of smaller fragmented transcripts. Transcripts above 150 bp were diminished in the –REL1 sample, indicating a reduction of partially and fully edited ND7. Two separate runs of Ion Torrent were performed on each sample. The sequencing read outs for ND7 specific samples after two separate runs are tabulated in (Appendix 4).

The adapter trimmed reads were analysed from FASTQ files using the pipeline in Figure 3.3. Data tables produced from this are collated in Appendix 5. The positions of the most 5' ends of these substrates were graphically represented (Figure 3.9) and as was the percentage of unedited, correctly edited and misedited sequence at each ND7 position , separately for Category 1 and 2 reads (Figures 3.10 and 3.11).

Q-PCR results (Figure 3.8) indicated that the -REL1 sample was more efficiently reverse transcribed than the +REL1 sample, which was reflected in the $\Delta\Delta C_t$ of 3.338, which was calculated using the difference of C_t between these two samples. Samples were run alongside an 18S positive control to check the overall Q-PCR performance, since this RT sample (from RNA of wt 427 cells with γ replacement, amplified with random hexamers) and primer set (5'-CGGAATGGCACCACAAGAC, 5'- TGGTAAAGTTCCCCGTGTTGA) gave good results previously (provided by Caroline Dewar). Results also indicate that β-tubulin was specifically amplified, and so the $\Delta\Delta C_t$ was used to adjust the number of read starts in Figure 3.9 and later on in Figure 3.13.

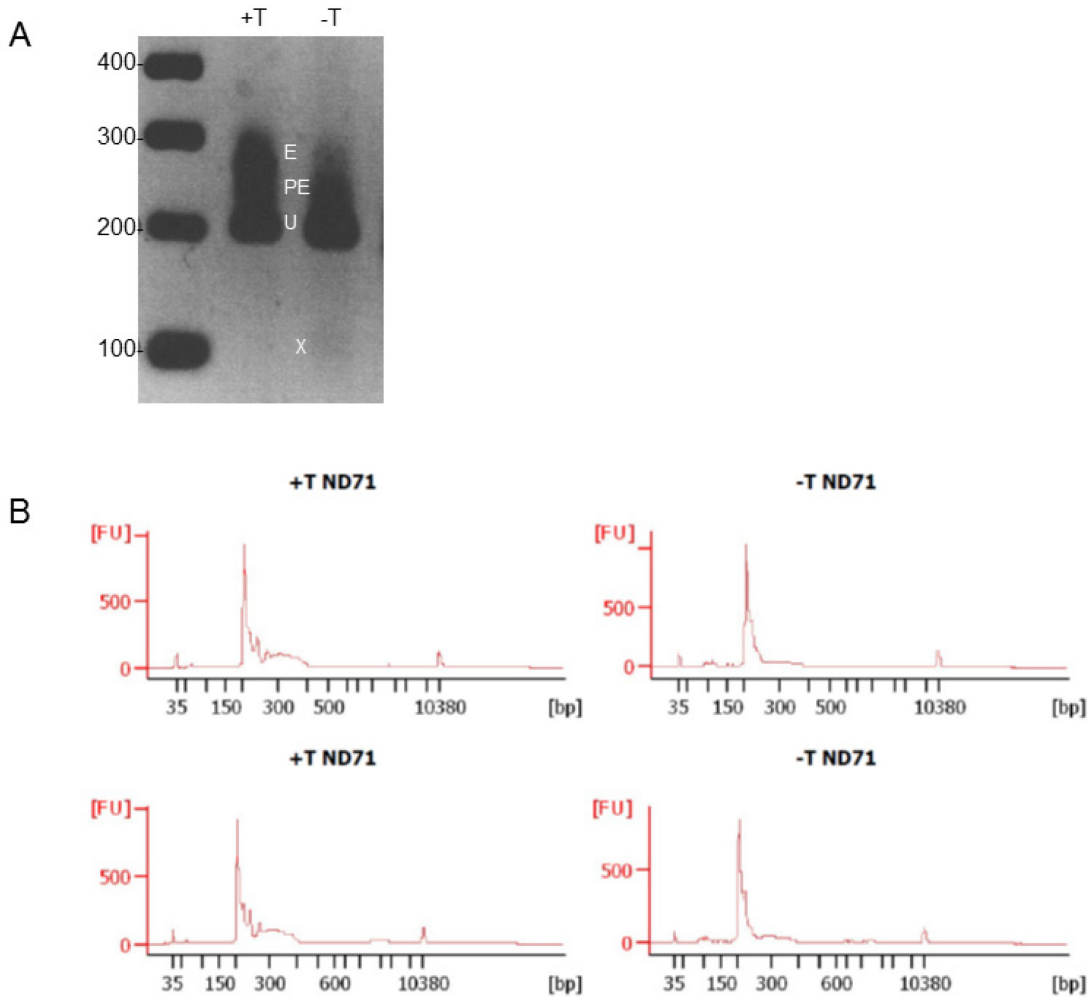


Figure 3.7. ND71 RT-PCR products for Ion Torrent sequencing

A) Gel visualisation of cleaned ND71 RT-PCR products. +T and –T correspond to cells growth in the presence and absence of REL1 respectively. E, PE and U indicate fully edited, partially edited and unedited transcripts, respectively. X corresponds to unligated fragments in the –T sample.

D) Agilent electropherographs for the corresponding gel in B. L - Ladder, +TND71 - +tet ND71 RT-PCR, -T ND71 - -tet ND71 RT-PCR. Repeats were run in duplicate (above and below) [bp] corresponds to transcript size in base pairs.

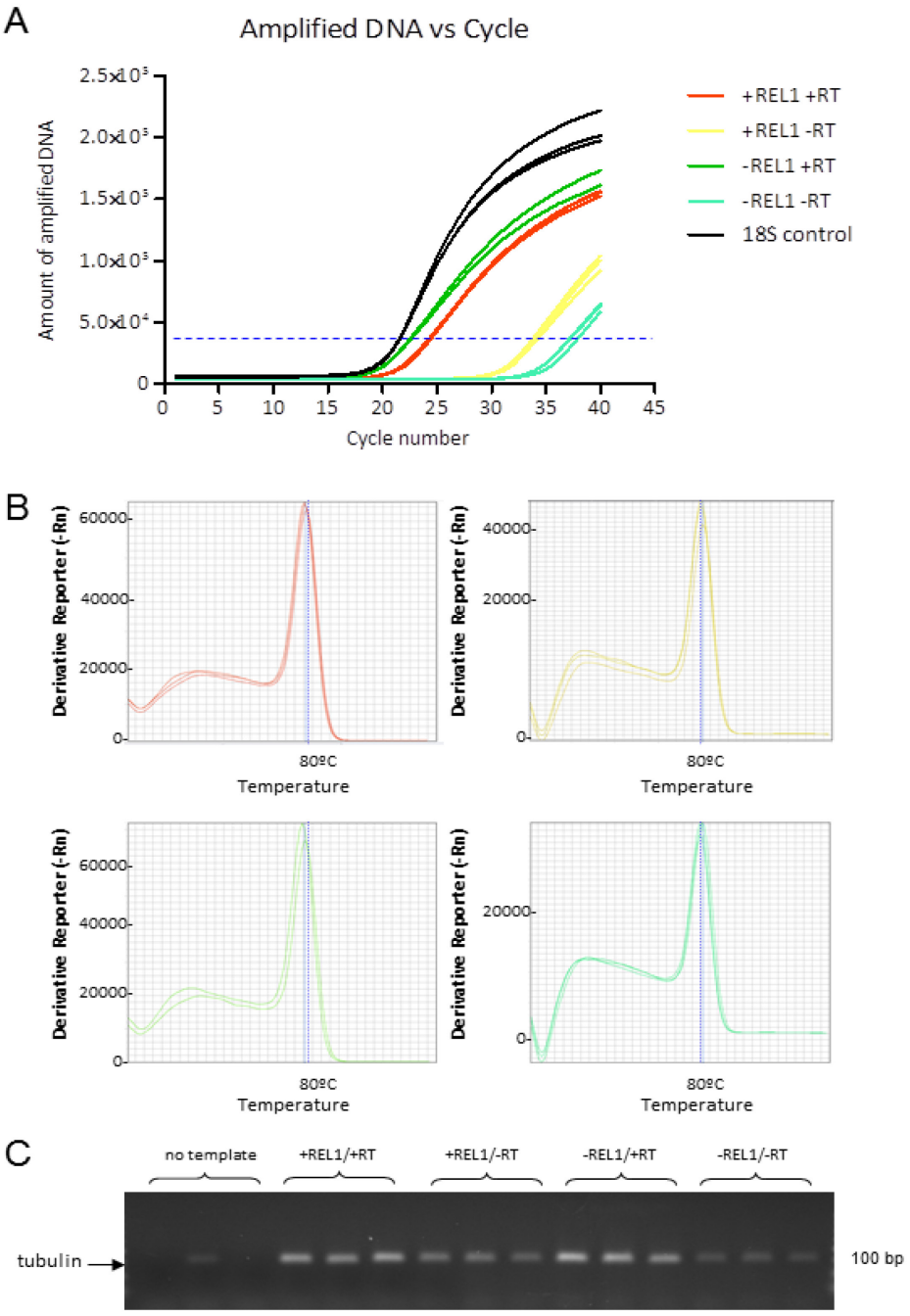


Figure 3.8. Q-PCR of +REL1 and -REL1 RT samples

A) Amplification curve of β-tubulin, complete with the 18S cDNA control (RNA from single marker cells, with wild type replacement gamma, reverse transcribed with random hexamers) for Q-PCR efficiency. The ΔΔCt for the +RT samples was calculated using the threshold line, intended to capture the reaction in logarithmic phase, indicated by a blue dashed line B) Melt curves for β-tubulin primers, which overlap as expected when a single product is amplified. C) Gel visualisation of Q-PCR after 40 cycles also indicates the presence of a specific band for β-tubulin. Q-PCR data was generated by Caroline Dewar.

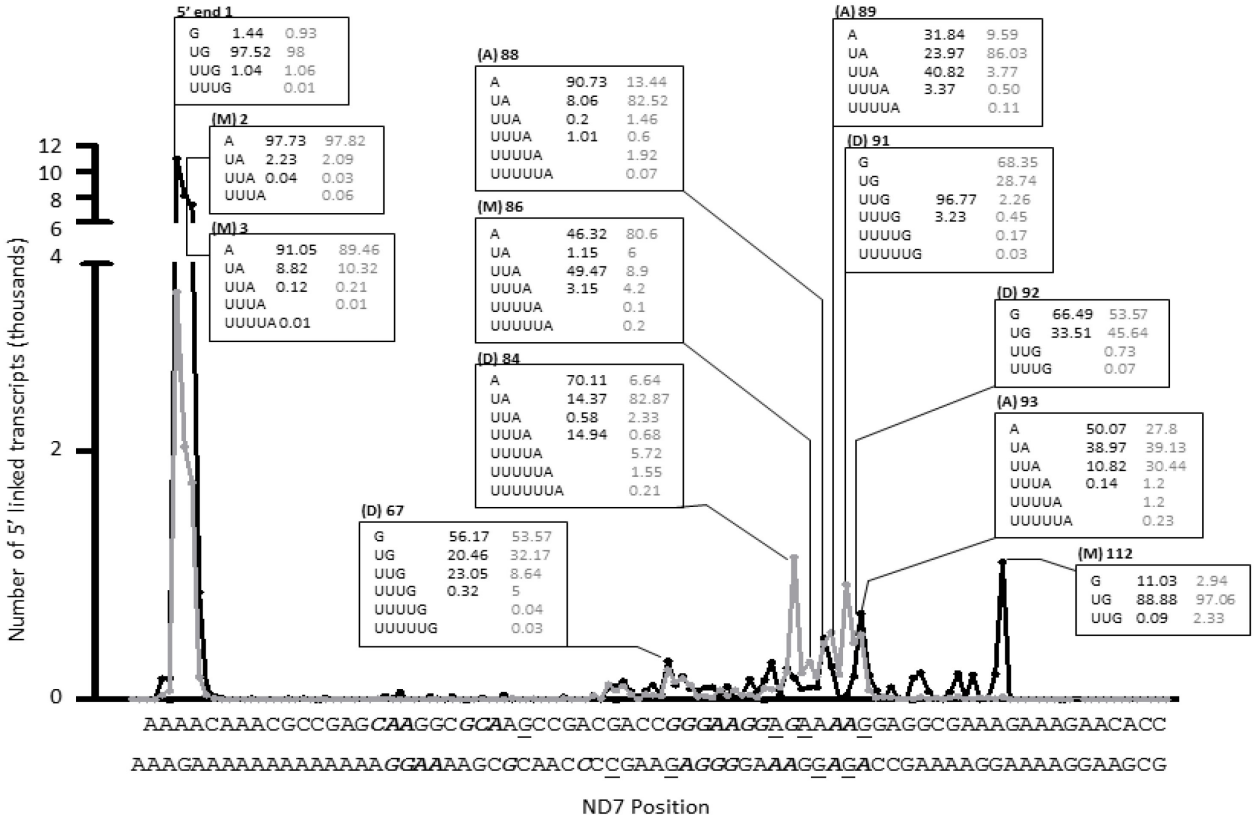


Figure 3.9. Marking the 5' ends of transcripts (5' linker/ND7P3 RT-PCR) from cells with REL1 induced/uninduced

The position of the 5' linker indicates the start of the sequenced transcripts in Category 1 reads. The number of starts (adjusted to the $\Delta\Delta C_t$ derived from Q-PCR results) was plotted against a T-stripped ND7 sequence. Starts had a heterogenous number of Us, expressed as percentages of the total within call-outs, for the peaks indicated. Black Line and numbering: +REL1 (+tet growth), grey line and numbering: -REL1 (-tet growth). The T- stripped sequence is given below. Underlined and bold/italic nucleotides refer to nucleotides immediately downstream of deletion and addition editing sites respectively.

Sites are marked as addition (A), deletion (D) or misediting (i.e. non canonical-editing sites) (M) editing sites and the relative position along the sequence is indicated. For the complete sequence, pre- and post-editing, see Figure 3.5.

The graph illustrating the 5' end of the ND7 transcripts retrieved from Ion Torrent sequencing (Figure 3.9) highlights the most prominent peaks (corresponding to read start position) in +REL1 and –REL1 samples. These peaks relate to the 5' ends after processing of the polycistronic pre-cursor and a total of four deletion sites, three addition sites and two misediting sites. The highest peaks coincide with the previously reported 5' end (Koslowsky *et al.*, 1990) and with the next two nucleotide positions downstream on the T-stripped sequence after for both the +REL1 and –REL1 samples. Although, it was clear that the 5' ends of ND7 were heterogenous from the linking at the first three nucleotide positions, the fact that we were able to map the most 5' site validates the sequencing approach and suggests that published results suggesting 5' homogeneity. The addition editing sites that have been flagged for both +REL1 and -REL1 samples contain adjacent deletion editing sites. The most common attachment sites for –REL1 are the deletion sites at positions 84, 91 and 92 (157, 172 and 173 on edited sequence, Figure 3.5) and the addition sites at positions 88, 89 and 93 (positions 165, 168 and 175 on edited sequence, Figure 3.5). For the +REL1 sample, the most common 5' attachment sites are the misediting site at position 112 and the addition sites at 88 and 93. In the absence of REL1 the percentage of expected addition editing sites linked markedly decreased. At positions 88 and 93 the greatest percentage of 5' linked sites were directly at the A bases, the expected site of attachment for a product of endonucleolytic cleavage destined for uridylyl addition. However, in the absence of REL1 at sites 88, 89 and 93 the greatest percentage of linked sequences contain one or two extra uridylyls, indicative of potential miscleavage events occurring after addition editing. In the absence of REL1, deletion sites at positions 67, 84, 89 and 92 the greatest percentage of linked

transcripts represented transcripts that had been cleaved ready for deletion editing (i.e. they contained the expected pre-edited number of uridylyls) or had the expected number of uridylyls removed from deletion editing. It should also be noted that the RT-PCR product responsible for the peak produced at position 112 (edited position 205, Figure 3.5) could not be identified on the gel as a physical product (Figure 3.7) and was treated as a technical anomaly of the sequencing process.

Figures 3.10 and 3.11 illustrate that Category 1 and 2 reads produce similar graphs. There are a number of unexpected uridylyl bases present within the first ten (Category 1) and 20 (Category 2) nucleotides of the 5' end of the T-stripped ND7 sequence, and these are likely to reflect the heterogeneity of the ND7 5' end. Misediting occurs throughout the entirety of the transcript, and can make up to 10% of the total editing events. It is of importance to note that on the percentage editing graphs dark grey bars correspond to editing at a non-editing site or the wrong number of uridylyls added at an editing site. The extent of correctly edited sequence within the editing blocks (areas of near consecutive editing) for Category 1 and 2 sets decreased from 3' to 5'. For –REL1 Category 1 data, correct editing events per site decreases from 3' to 5' at a greater rate than with REL1, and there are fewer correctly edited sites upstream from site 66 (T-stripped sequence). Since Category 1 data originated from the most 5' end (containing the 5'linker) and spanned the length of the transcript to the 3' end it may be more reliable for interpretation than Category 2 data. Category 2 reads follow a similar pattern to that of Category 1, however the percentage of misedited and correctly, fully edited transcripts in the absence of REL1 is diminished overall in comparison. This reduction of correctly edited sequence may represent an accumulation of transcripts representative of editing bottlenecks, which

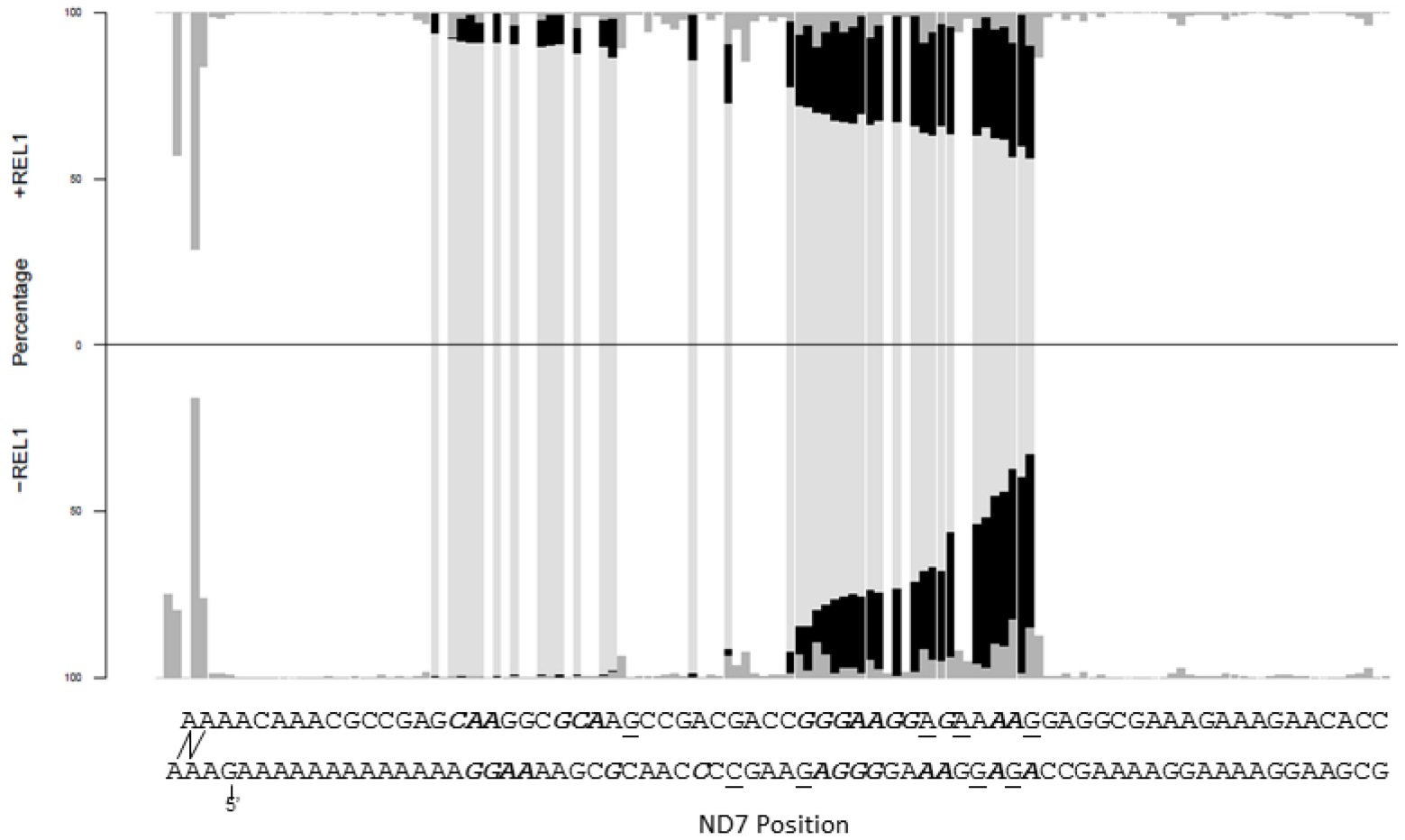


Figure 3.10. Percentage of each type of editing, Category 1 reads for 5' linker/ND7P3 RT-PCR
Percentages of editing were calculated from total reads and from the number of reads involved in observed and expected editing events. White – non-edited at non-editing site, light grey - non-edited at editing site, dark grey – edited at non-edited site/misedited, black – fully/correctly edited at editing site. The T- stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.

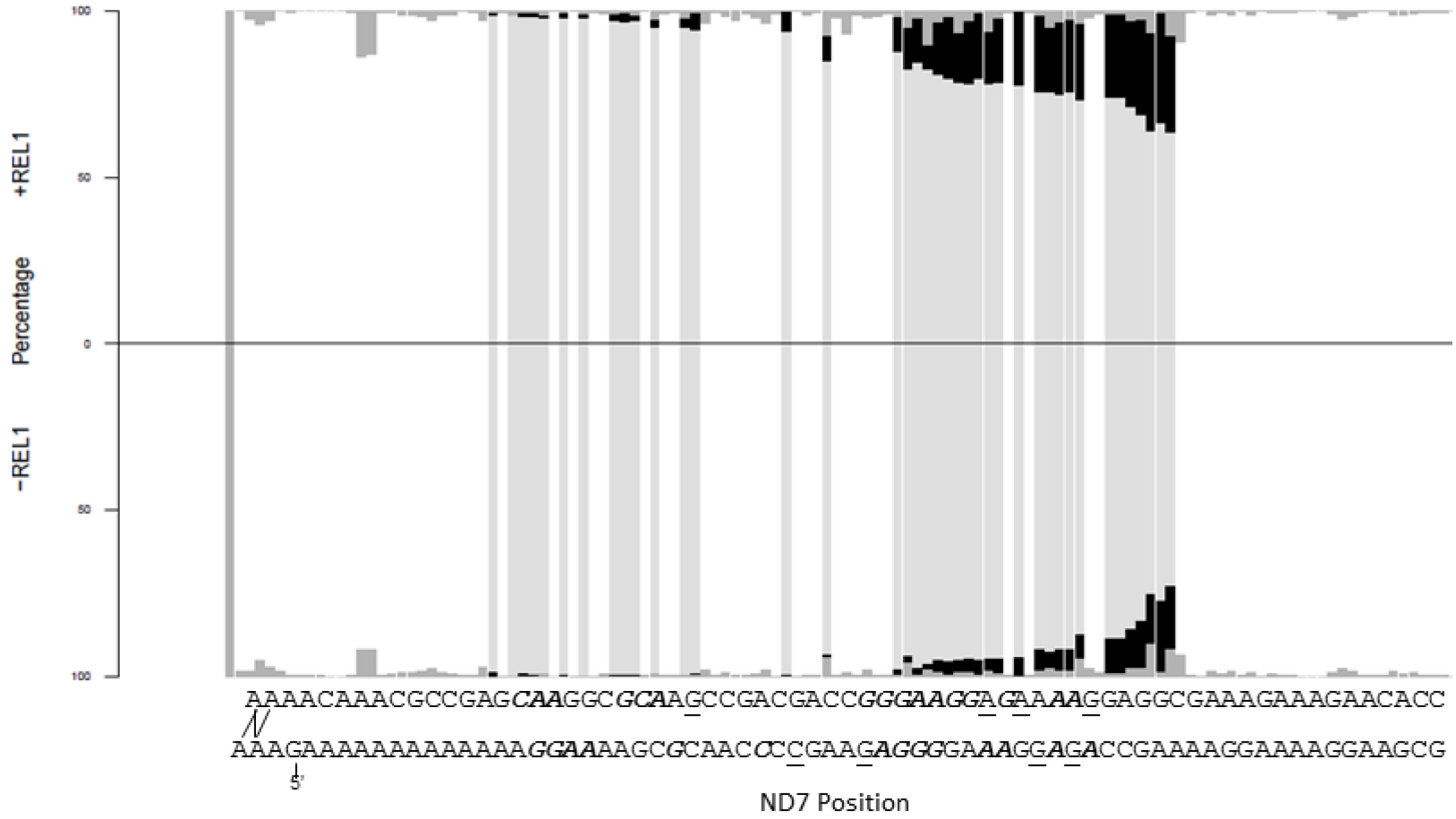


Figure 3.11. Percentage of each type of editing, Category 2 reads for 5' linker/ND7P3 RT-PCR
Percentages of editing were calculated from total reads and from the number of reads involved in observed and expected editing events. White – non-edited at non-editing site, light grey - non-edited at editing site, dark grey – edited at non-edited site/misedited, black – fully/correctly edited at editing site.
The T- stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.

are more apparent in the –REL1 sample.

3.3.4 Deep sequencing analysis of REL1 substrates through MiSeq™ sequencing

The sequences used in this analysis (tabulated in Appendix 6) were first subjected to the same pipeline as outlined in Figure 3.3. The MiSeq data was based upon RT-PCRs that selected for editing intermediates only (i.e ES1 for RPS12 and ES1 ND7) and therefore all sequences linked and analysed had entered into the RNA editing reaction. Due to the short length of the ND7 sequences there were no separate A and B Categories, as reads covered the whole length of the sequences from the most 5' end, to the reverse primer. Before sending to the Genepool for MiSeq analysis, cleaned RT-PCR products were visualised on a gel (below in Figure 3.12).

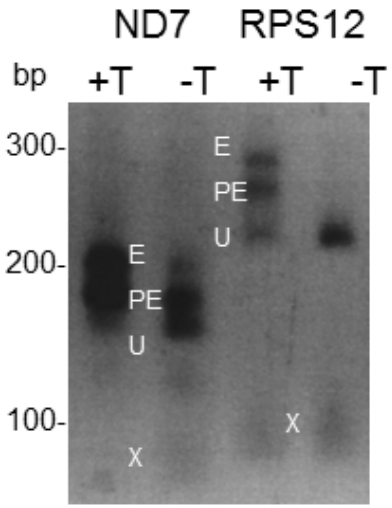


Figure 3.12. RT-PCR product for MiSeq

The RT-PCR product was generated using primers that selected for ND7 and RPS12 editing intermediates, and no non-edited transcripts. +T and –T correspond to cells growth in the presence and absence of REL1 respectively. E, PE and U X indicate fully edited sequence, partially edited sequence and un-edited sequence and unligated fragments, respectively.

The analysis of 5’ read starts (Figure 3.13) reveals the accumulation of 5' linked transcripts at deletion editing positions 67, 84 and 91 (positions 110, 157 and 172 on edited sequence, see Figure 3.5), which also correspond to some of the most prominent peaks in the Ion Torrent data for this transcript in the absence of REL1. Position 91 produced the highest peak, corresponding to editing site 2, the first deletion editing site. This indicates that the greatest number of small fragments accumulated in the absence of REL1 corresponds to this editing site, and of these fragments over 90% of transcripts have not had the uridylyl removed by the ExoUase. The first three most 5’ bases are the most commonly linked, and these full length transcripts are more abundant in the presence of REL1, indicating full length transcripts at various stages of editing. This echoes the results of the Ion Torrent sequencing and indicates that both deletion editing and misediting products are substrates for ligation by REL1. To determine whether more peaks could be visualised, the percentage of 5’ starts of the total reads were calculated and used to construct Figure 3.14. This method yielded a greater number of peaks and showed a clearer difference between + and – REL1 samples, which may have been occluded by adjusting for RT-PCR efficiency. As expected, the peaks pertain to the first three bases and deletion editing sites 67, 84 and 91 (positions 110, 157 and 172 on edited sequence, see Figure 3.5). Peaks were also apparent at positions 69 (addition editing site, position 117 edited sequence) and 74 (addition editing site, position 131 edited sequence) which are not apparent from the Ion Torrent data. At position 74, 65.52% and 96.91% of linked reads could be assigned to a fragment containing one terminal uridylyl, for +REL1 and –REL1 samples, respectively, indicative of potential miscleavage events occurring after addition editing.

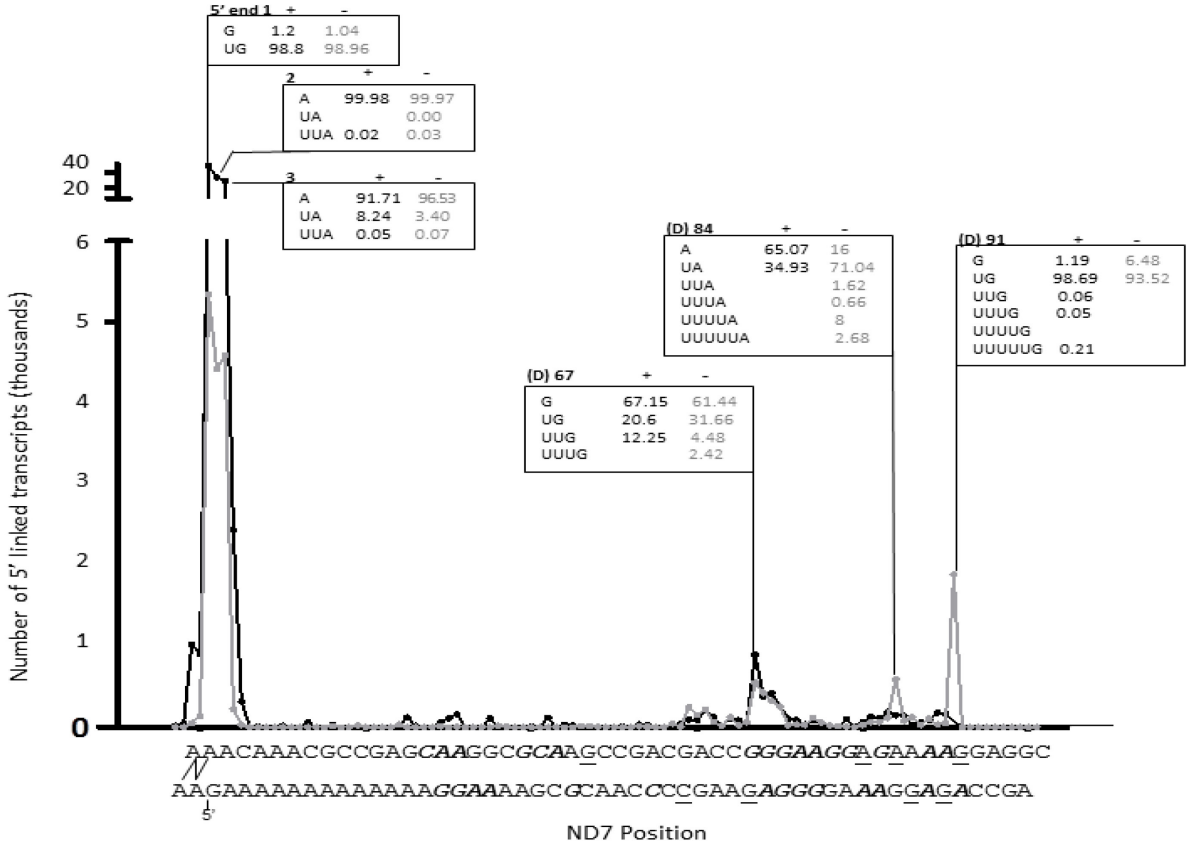


Figure 3.13. Marking the 5' ends of transcripts (5' linker/ND72 RT-PCR) from cells with REL1 induced/uninduced

The position of the 5' linker indicates the start of the sequenced transcripts in Category 1 reads. The number of starts (adjusted to the $\Delta\Delta C_t$ derived from Q-PCR results) was plotted against a T-stripped ND7 sequence. Starts had a heterogenous number of Us, expressed as percentages of the total within call-outs, for the peaks indicated. Black Line and numbering: +REL1 (+tet growth), grey line and numbering: -REL1 (-tet growth). The T- stripped sequence is given below. Underlined and bold/italic nucleotides refer to nucleotides immediately downstream of deletion and addition editing sites respectively.

Sites are marked as addition (A), deletion (D) or misediting (i.e. non canonical-editing sites) (M) editing sites and the relative position along the sequence is indicated. For the complete sequence, pre- and post-editing, see Figure 3.5.

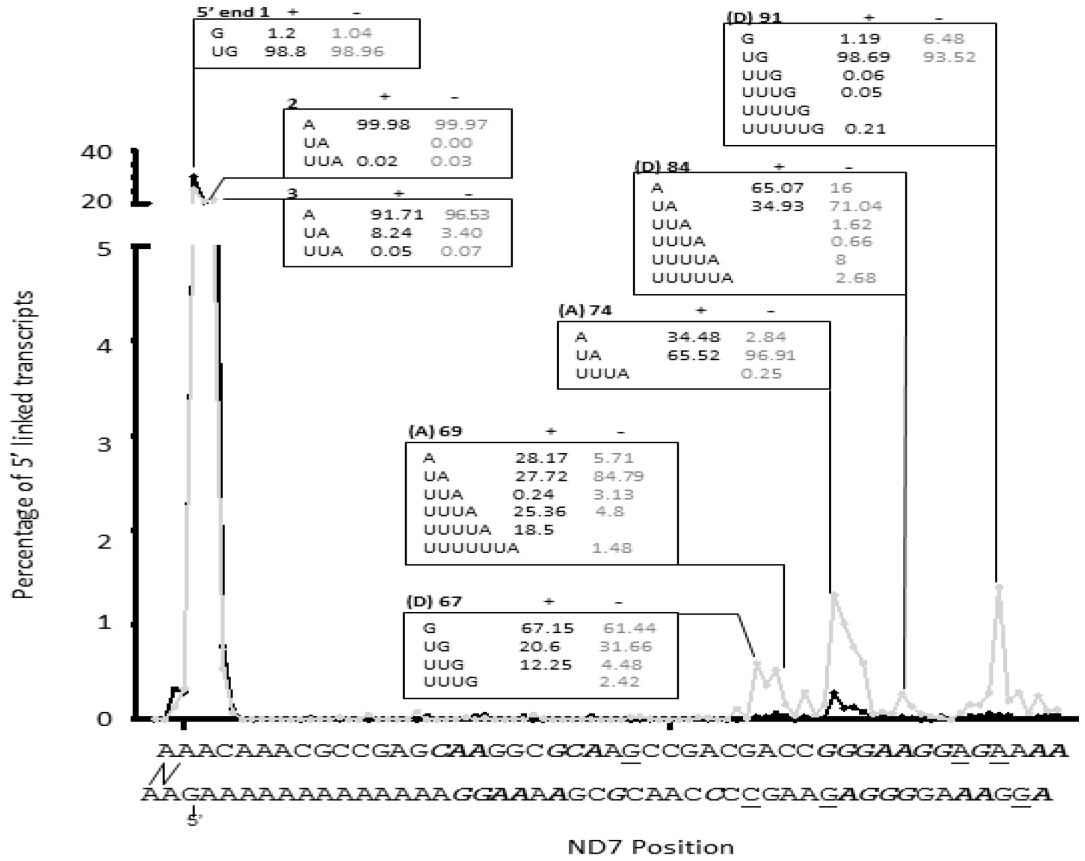


Figure 3.14. Percentage of 5' ends of transcripts (5' linker/ND7P2 RT-PCR) from cells with REL1 induced/uninduced
The position of the 5' linker indicates the start of the sequenced transcripts in Category 1 reads. The number of percentage starts of the total reads was calculated and plotted against a T-stripped ND7 sequence. Starts had a heterogenous number of Us, expressed as percentages of the total within call outs, which correspond to designated peaks. Black line and numbering: +REL1 (+tet growth), grey line and numbering: -REL1 (-tet growth).The T-stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.
Sites are marked as addition (A), deletion (D) or incorrect (M) editing sites and the relative position along the sequence is indicated.

Position 69, in the +REL1 sample, was linked via a highly variable number of uridylyls, whereas in the absence of REL1, 84.79% of 5' linked transcripts contained a single terminal uridylyl.

Analysis of the percentage editing of ND7 from MiSeq sequencing (Figure 3.15) revealed misediting events fluctuating between 0% and 30% at sites throughout the rest of the length of the transcript. The most 5' end reveals heterogeneity in the number of Ts present in the stretch of never-edited sequence, as previous described from Ion Torrent data. The percentage of total fully and partially edited sequence fluctuates between 100% (due to the primer selecting for transcripts that have already entered the editing process) and 10% at each position from the most 3' and 5' editing site of the edited region. The rate of decline is reduced for 3' to 5' in the presence of REL1, in comparison to the absence of REL1, as seen for the Category 1 Ion Torrent sequencing data, which may indicate editing bottle necks.

RPS12 transcripts were also analysed by MiSeq sequencing. Since no discernable 5' linked peaks from +REL1 or –REL1 samples could be visualised using the Q-PCR adjusted starts (save the most 5' end), a graph was constructed for the percentage of 5' starts within the total reads (Figure 3.16). This method of visualisation revealed heterogenous 5' linking of transcripts at the most 5' positions, as with ND7. The other sites that were linked were at positions 131 (misedited, position 164 of edited sequence, Figure 3.5), 132 (deletion editing site, position 168 edited sequence), 140 (addition editing site, position 182 edited sequence), 141 and 143 (addition editing sites at positions 183 and 185 edited sequence), and 145 (misedited site, position 188 edited sequence).

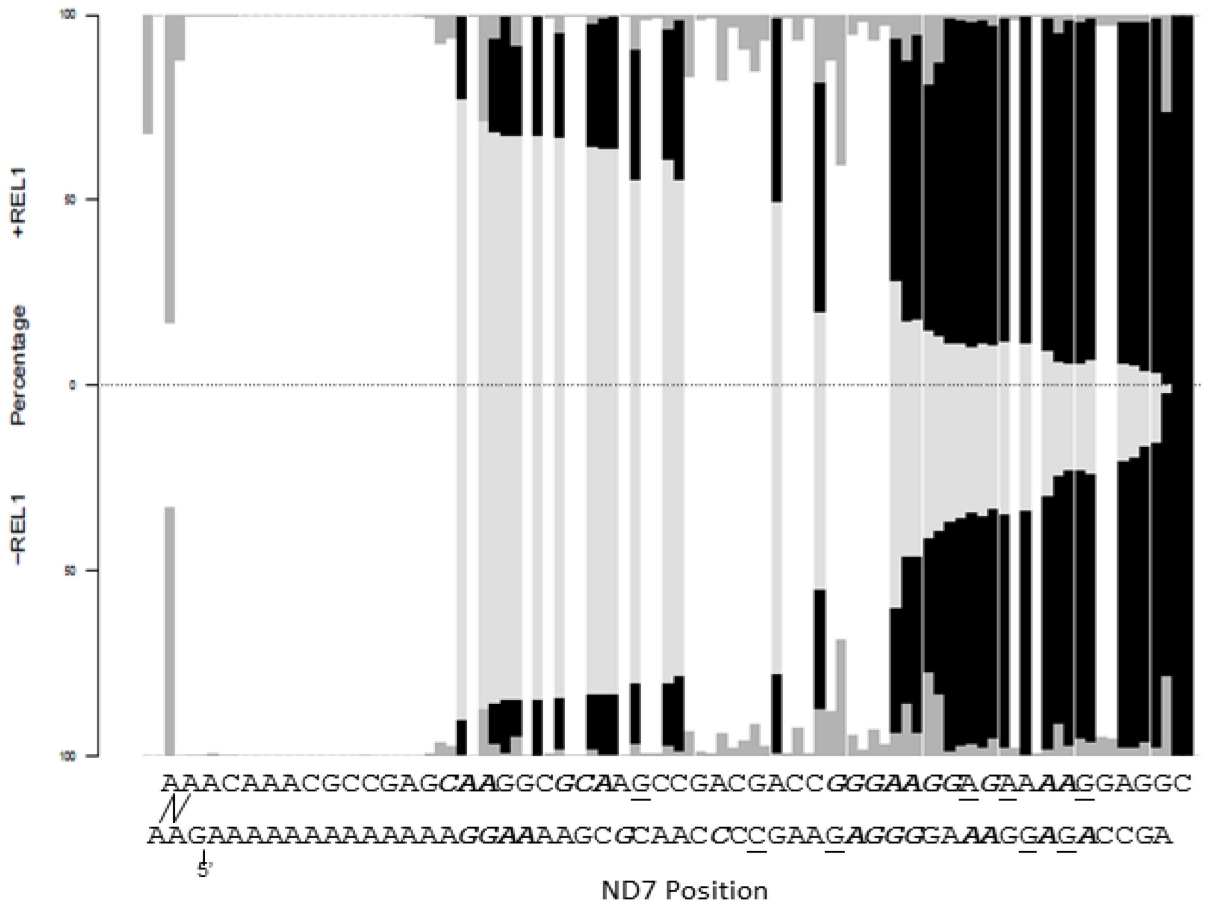


Figure 3.15. Percentage of each type of editing for 5' linker/ND7P2 RT-PCR
Percentages of editing were calculated from total reads and from the number of reads involved in observed and expected editing events. White – non-edited at non-editing site, light grey - non-edited at editing site, dark grey – edited at non-edited site/misedited, black – fully/correctly edited at editing site.
The T- stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.

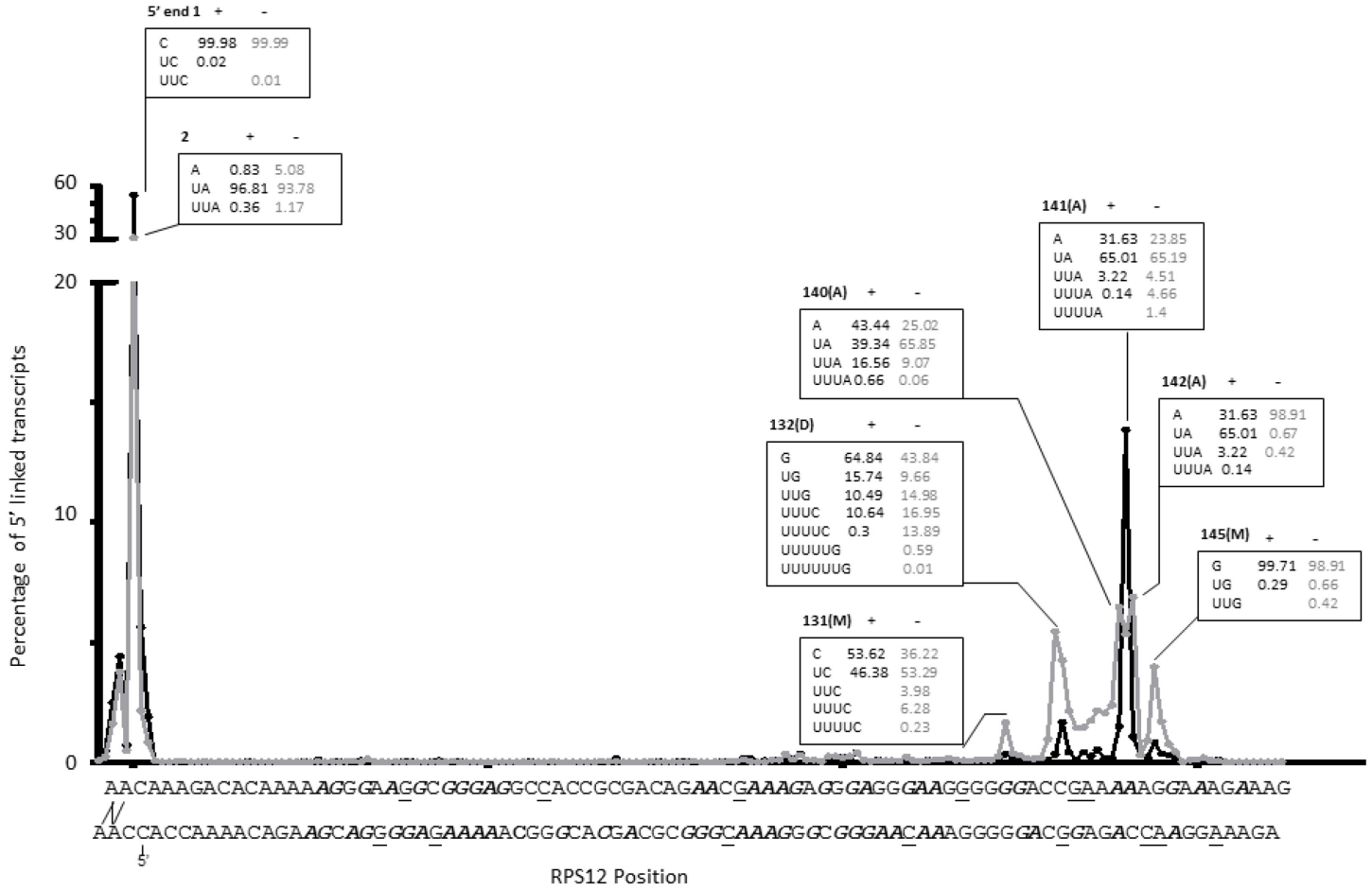


Figure 3.16. Percentage of 5' ends of transcripts (5' linker/RPS12 RT-PCR) from cells with REL1 induced/uninduced
The position of the 5' linker indicates the start of the sequenced transcripts in Category 1 reads. The number of percentage starts of the total reads was calculated and plotted against a T-stripped ND7 sequence. Starts had a heterogenous number of Us, expressed as percentages of the total within call outs, which correspond to designated peaks. Black line and numbering: +REL1 (+tet growth), grey line and numbering: -REL1 (-tet growth).The T-stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.
Sites are marked as addition (A), deletion (D) or incorrect (M) editing sites and the relative position along the sequence is indicated.

Linked transcripts at position 132 were representative of deletion editing substrates, with the greatest percentage of +REL1 and –REL1 derived transcripts having a terminal G-base. Misedited sites were products of mis-directed endonucleolytic cleavage, and contained either an expected non-edited uridylyl (position 131) or no terminal uridylyl (position 145). Transcripts linked at addition editing site 140 in the presence and absence of REL1 could be attributed to production of true addition editing substrates lacking a terminal uridylyl (43.44% and 25.02%, respectively) and of miscleaved substrates (39.34% and 65.85%, for +REL1 and –REL1 samples, respectively). The addition editing site linked, at position 141, revealed two thirds of +REL1 and –REL1 derived transcripts that contained a single terminal uridylyl, suggestive of miscleavage events occurring after addition editing. Contrary to this, 98.91% of transcript fragments produced in the absence of REL1 at position 142 could be considered addition editing substrates, since they contained no terminal uridylyl.

Figure 3.17 illustrates the percentage of editing type for RPS12 along the transcript. In the presence of REL1, fully edited transcripts decrease from 100% at the most 3’ end to less than 5% at the most 5’ editing site. This decrease is more gradual than with ND7. Misediting events at editing and non-editing sites does not exceed 25%, save at the most 5’ site, which may be indicative of a heterogenous 5’ end. In the absence of REL1 the presence of fully edited sites is drastically reduced below 5% throughout the length of the transcript, and examples of misediting increase proportionally from 5% (at most 5’ editing site) to 100% (most 3’ editing site). This clearly indicates that the absence of REL1 abolishes normal RNA editing in RPS12, and instead increases the number of misediting or miscleavage events.

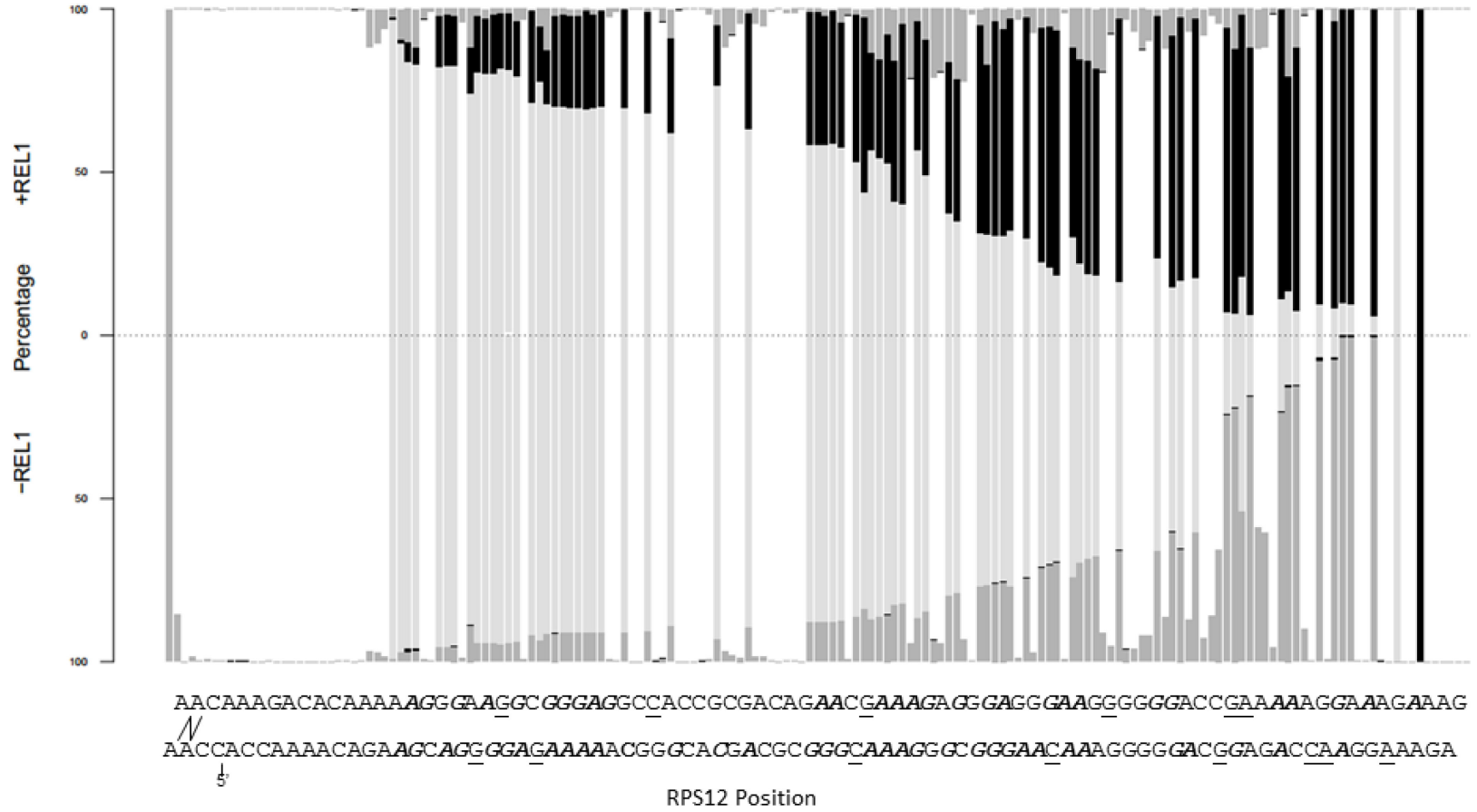


Figure 3.17. Percentage of each type of editing for RPS12

Percentages of editing were calculated from total reads and from the number of reads involved in observed and expected editing events. White – non-edited at non-editing site, light grey - non-edited at editing site, dark grey – edited at non-edited site/misedited, black – fully/correctly edited at editing site.

The T- stripped sequence is aligned below. Underlined and bold/italic nucleotides refer to deletion and addition editing sites respectively.

3.4 Discussion

This study has indicated potential substrates for REL1 ligation throughout ND7 and RPS12 using a novel sequencing approach. Full length fully, partially and unedited 5' linked transcripts could be cloned and sequenced in the presence of REL1. However, shallow sequencing of 5' linked transcripts revealed shorter fragments, which had failed to be re-ligated by REL1. These fragments represented substrates of typical deletion and erroneous, and unexpected addition, editing events. However, it was clear that the 5' linker attached to addition editing sites, albeit not to typical substrates of addition editing, but rather to products of unexpected endonucleolytic cleavage. Of these misedited products, most started with one or two uridylyl residues, indicating that addition editing has occurred and the transcripts later had been cleaved by an endonuclease. Interestingly, there were no examples of 5' linked transcripts indicative of true insertion editing. But, since only few sequences were available from limited cloning to describe this finding, deeper sequencing was undertaken. The cloned sequences reflect the gel patterns obtained from RT-PCR (also previously observed by Schnaufer *et al.*, 2001) and clearly indicate that REL1 ablation causes a reduction in fully edited mRNA populations to those of unedited or partially edited ones.

The ND7 RT-PCR samples sent for Ion Torrent and MiSeq revealed the same 5' linked deletion editing sites at positions 67, 84 and 91 in the absence of REL1. The 5' linker in this sample also adhered to addition editing sites (differing in position for Ion Torrent and MiSeq sequencing results), but few examples of expected addition editing products could be found at these sites, in comparison to products with terminal uridylyls. A key difference between the two RT-PCR samples

generated for Ion Torrent and MiSeq was the selection for transcripts that have entered into the editing cycle only in the latter sample. This was reflected by the greater percentage of overall correctly and fully edited sequences at the most 3' end of ND7 in both +REL1 and –REL1 samples.

The RPS12 RT-PCR samples sent for MiSeq analysis did not reveal any particular peaks relating to REL1 substrates when starts were initially analysed, even though clear bands corresponding to fully, partially and unedited transcripts were observed on the agarose gel. However, it was obvious that the preferred attachment site of the 5' linker was the previously reported 5' end of the transcript for +REL1 (Read *et al.*, 1992). To gain further insight into the sites of 5' linkage, the percentage of read starts of the total number of reads obtained was plotted, which revealed further addition and non-editing sites. The results here suggest that, whilst a small percentage of 5' ends reflect true, or expected, products of addition editing, most are indicative of deletion editing and misediting (or miscleavage) events.

Consistently, between all three separate runs of sequencing, the percentage of expected addition substrates at editing site positions was greater in the +REL1, in comparison to the –REL1 samples. This may be indicative of dual function of REL1 and REL2, since REL1 had not re-ligated these addition editing substrates. A number of previous studies have suggested that REL1 and REL2 have distinct, but not mutually exclusive functions in editing (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz *et al.*, 2002; Schnaufer *et al.*, 2003; Gao and Simpson, 2003). Whilst this suggests distinct functions of the ligases in deletion and addition editing, respectively, it also suggests that the essentiality of REL1 is linked to its function in ligating erroneously cleaved transcripts (Huang *et al.*, 2001).

Misediting, or erroneous cleavage of RNA substrates happens frequently *in vitro* (Kable *et al.*, 1996; S D Seiwert *et al.*, 1996; Rusché *et al.*, 1997). It has previously been suggested that REL1 may have had a key role as an RNA repair enzyme, as its closest relative is T4Rnl2, and may be involved in the repair of erroneously cleaved substrates born of misguiding by the incorrect gRNA (Koslowsky *et al.*, 1991; Huang *et al.*, 2001; Ho and Shuman, 2002; Palazzo *et al.*, 2003; Yin *et al.*, 2003; Ho *et al.*, 2004; Nandakumar *et al.*, 2004; Deng *et al.*, 2004). These fragments are not re-ligated by REL2 in this scenario, since endogenous REL2 is present in the REL1 cKO cell lines. This is likely due to the substrate requirements of REL2, since this ligase requires a perfectly nicked duplex produce from addition editing directed cleavage and cannot ligate overhangs like REL1 (Blanc *et al.*, 1999; Igo *et al.*, 2000; Schnaufer *et al.*, 2001; Cruz-Reyes *et al.*, 2002; Palazzo *et al.*, 2003). This study provides evidence of this, through the identification of unligated RNA substrates. This curious observation is frequently linked to the presence of additional uridylyls at an addition editing site that may suggest endonucleolytic cleavage occurred after the addition editing event, and as a result the transcript was cleaved erroneously. It cannot be ruled out that these observed bases are due to the *in vivo* formation of gRNA-mRNA chimeras through the action of an endonuclease and subsequently a ligase, or through transesterification (Blum *et al.*, 1991; Rusché *et al.*, 1995). However, it was previously observed that this chimera formation was decreased in conditions that also decreased ligase activity, it would be expected that such instances would decrease in the absence of REL1, and this is not shown here (Rusché *et al.*, 1995). These chimeras formed of truncated mRNAs and gRNAs would be a dead end process formed of RNA editing, and would initiate some form

of editing block (Sturm and Simpson, 1990b). It is known that addition uridylyls are added to an editing transcript by a TUTase, before potential excision by a 3' exoUase (Byrne *et al.*, 1996; McManus *et al.*, 2000; Igo Jr. *et al.*, 2002; Zhelonkina *et al.*, 2006). Additional uridylyls present at deletion editing sites may be a product of multi-cycle RNA editing and reflect the addition of extra uridylyls by a TUTase, independent of gRNA interaction (Feagin, Abraham, *et al.*, 1988; Alatortsev *et al.*, 2008)

Taken together, results of limited, Ion Torrent and MiSeq sequencing reveal that the most frequently linked sequences, for both +REL1 and –REL1 samples, (as indicated by the number of read starts) correspond to the first three nucleotides starting from the 5' end of ND7. This may reflect a naturally occurring "micro" heterogeneity of the 5' end, which has not been described before in ND7 (Koslowsky *et al.*, 1990), but has been described before in cloned DNA sequences from CR6 (RPS12) (Read *et al.*, 1992b). This linkage does not provide information on the editing state that these transcripts are in, however, which would differ greatly between +REL1 and –REL1 samples, taking into consideration the results of shallow sequencing (see Appendix 3). It is clear from the ND7 data that there is an accumulation of transcript fragments starting at editing sites 1 (position 91), 2 (position 84) and 3 (position 67) after REL1 ablation, which may suggest REL1 has a preference for these deletion editing substrates, over the products of misediting and addition editing. The same cannot be said for RPS12 in the absence of REL1, since the favoured sites for 5' linking were at addition editing sites, and contained a mix of expected addition editing substrates and potential products of miscleavage. This, together with the percentage editing graph, indicates that REL1 has an important role

in the processing of miscleaved mRNAs and perhaps have a more active role in re-ligation of addition editing products in RPS12.

The most unexpected peaks were those seen in +REL1 which were most apparent in the MiSeq data. These peaks were omitted from the ND7 5' linked graph generated (due to peak intensity), but shown for RPS12 (Figure 3.16) as a peak at position 141. These peaks were probably not genuine, however, as they could not be correlated with bands on the gel. Studies have ascertained that the greatest biases have been introduced by the addition of 5’ and 3’ adapters to the transcripts, through the preferences of T4 RNA ligases to certain substrates and the secondary and tertiary structure of the RNAs sequenced (Amitsur *et al.*, 1987; Hafner *et al.*, 2011; Jayaprakash *et al.*, 2011). This is important to note since it could have introduced linking biases, and given rise to unexplained peaks in the sequence data, although the reasons behind this discrepancy between +REL1 and –REL1 samples is unknown.

The percentage editing graphs illustrate a general 3’ to 5’ progression of editing, which has previously been reported in the literature. Prior studies have revealed the most abundant mRNAs in the population were partially edited transcripts (Abraham *et al.*, 1988; Decker and Sollner-Webb, 1990; Sturm *et al.*, 1992). Sequencing also revealed apparent junctions between these blocks of editing, suggesting that editing does not strictly occur from 3’ to 5’, with unexpected gRNAs often found along the stretch of pre-mRNAs (Feagin, Abraham, *et al.*, 1988; Decker and Sollner-Webb, 1990; Koslowsky *et al.*, 1990; Sturm and Simpson, 1990a; Sturm and Simpson, 1990b; Koslowsky *et al.*, 1991; Read, Corell, *et al.*, 1992; Sturm *et al.*, 1992). However, the in-depth analysis of percentage type of editing at each site present here confirms that editing does progress generally in a 3' to 5' direction. This

3’ to 5’ editing efficiency (equated to the percentage of fully edited transcripts per T-stripped nucleotide site), however, is diminished in the absence of REL1. REL1 knock-down abolishes normal RNA editing in RPS12, and instead proportionally increases the number of misediting or miscleavage events, throughout the length of the transcript. The discrepancy between Category 1 and 2 Ion Torrent data has most likely arisen from the lower quality score of the Category 2 data, since these reads were compiled from short fragments of potentially overlapping sequence originated from the most 3’ end of the ND7 transcript. Category 1 Ion Torrent and MiSeq sequence analysis illustrates a sharper diminution of fully edited transcripts from 3’ to 5’ in the absence of REL1, in comparison to +REL1 samples. The effect of REL1 knock-down on RNA editing was more pronounced for RPS12, although the reason for this is unclear, as both proteins would be required for the cells function and are preferentially edited in BSFs (Koslowsky *et al.*, 1990; Read, *et al.*, 1992b).

3.5 Conclusions and Outlook

The main findings of this Chapter were (i) REL1 is involved in the ligation of products born of deletion editing, more frequently in misediting, and rarely in addition editing events, (ii) the 5' ends of pan-edited transcripts exhibit heterogeneity, (iii) the efficiency of 3' to 5' editing is reduced, or diminished in the absence of REL1, (iv) the 5' monophosphate born from endonucleolytic cleavage can be used to capture REL1 substrates using an RNA linker, which provides a basis for the sequencing of ligation substrates. This study is the first deep sequencing analysis of RNA editing intermediates to date, and provides an interesting insight into REL1 substrates. There however, remains much potential to expand on these results.

For example, the material used to generate RT-PCR products used all three sequencing approaches was limited to a single biological replicate, due to time constraints. It would be important to future studies to repeat the approaches with RNA extracted from separate biological samples.

To dissect the potential RNA repair function of REL1, a novel assay could be derived from existing addition and deletion editing assays (Igo Jr. *et al.*, 2002). This method would utilise pre-cleaved mRNAs (representative of miscleaved substrates) bridged to gRNAs as the substrates of interest.

Further to this it would be important to verify that the absence of REL1 would not be causing any secondary defects to editing, by for example by destabilising other editosome components (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002). This has not been reported in other studies, however (Stuart *et al.*, 2002). This potential influence could be tested by utilising RNA isolated from cell lines

expressing a catalytically dead REL1 in its stead (such as used by Huang *et al.*, 2001).

Since results here indicate that REL1 is involved primarily in misediting and deletion editing, it would be interesting to ascertain if REL2 is involved solely in addition editing. This could be achieved by extracting poly(A)+ RNA from an available REL2 RNAi line (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Cruz-Reyes *et al.*, 2002; O'Hearn *et al.*, 2003). Since few instances of expected addition editing substrates were observed here, we would expect the editing fragments produced from REL2 knock-down to represent these specific ligation substrates. Further investigation here is required to validate our working hypothesis that REL2 may be functioning as a proofreading ligase, along side REL1 that ligates productions of deletion editing and erroneous cleavage.

The scope of this PhD study only permitted the analysis of ND7 and RPS12 transcripts, but this approach could be applied to the other nine mitochondrial transcripts that are edited in BSFs, which in turn would reveal the extent of REL1 involvement in the final step of RNA editing. This would be interesting to follow up since the data presented here suggest that they are transcript-specific effects of REL1 knock-down.

Further more, since the 5' linking of cleaved transcripts with the Solexa-based RNA linker was robust enough, a more global approach could be pursued. This was the original intention of this study, and the reason why a tagged hexamer primer was initially used to generate the RT material used in subsequent sequencing reactions. The tagged hexamer could subsequently be used to select for all 5' linked

mitochondrial transcript in an enrichment PCR reaction containing a primer based upon the hexamer tag and the 5' linker sequence. From the data presented here, Ion Torrent may produce more meaningful data. Whilst the long poly-U stretches may introduce biases during genome sequencing and increase error in the raw dataset, the ND7 sequencing results presented here suggest reproducibility of MiSeq to that of Ion Torrent sequencing results (Quail *et al.*, 2012).

To date, much information regarding the transcript specific effects of RNA editing has relied on the generation of large clone libraries (Benne *et al.*, 1986; Feagin *et al.*, 1987; Feagin, Abraham, *et al.*, 1988; Bhat *et al.*, 1990; Koslowsky *et al.*, 1990; Read, Myler, *et al.*, 1992; Souza *et al.*, 1992; Read, Wilson, *et al.*, 1994; Ochsenreiter and Hajduk, 2006; Ochsenreiter *et al.*, 2008). This global approach would allow the direct investigation into the transcripts involved in REL-based ligation, and may allow the quantification of the involvement of REL1 in RNA editing as a whole.

Chapter 4

Discerning the importance of REL1 and REL2 through evolutionary analysis

4.1 Introduction to project

As discussed in Chapters 1 and 2, it has been ascertained that REL1 is essential to cell viability and the editing process, where as REL2 is not. The assumption that REL2 is non-essential to the cell is based on RNAi data, and specifically through analysis of growth phenotypes (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz *et al.*, 2002; Gao and Simpson, 2003); however, microscopic images of REL2 RNAi cell lines reveal subtle cell deformation (O'Hearn *et al.*, 2003). If REL2 is not essential, then what is its purpose at its position within the addition subcomplex? Initially, we wanted to create a REL2 null mutant within a cKO REL1 environment to aid the integration of a tagged chimeric ligase (with a REL2 ID) into the editosome (Chapter 2). However, we also wanted to create a null mutant within a wild type environment, since RNAi is never fully complete and only a small amount of REL2 may be required for editing to function as normal.

Aside from the direct genetic manipulation of *T. brucei* to ascertain if a residual amount of REL2 within the cell allows editing to function, we also wanted to determine if REL2's presence within the cell could be explained evolutionarily. It is known that proteins are subject to strong purifying selective forces when they have stringent substrate or structural requirements, and so amino acid changes can be directly correlated to catalytic importance. A classic measure of the type of selective forces acting upon a protein is the calculated ratio of the number of synonymous and non-synonymous substitutions per site (dN/dS) (Hurst, 2002). Catalytically important proteins evolve more slowly than would be expected under neutrality. For example bacterial proteins which are essential and mammalian house keeping genes are more conserved over evolutionary time than non-essential, or tissue specific proteins

(Jordan *et al.*, 2002; Zhang and Li, 2004). These observations can be applied to the editing ligases of *T. brucei* to determine whether REL1 and REL2 have been subjected to the same kind of evolutionary forces. For example, do selective forces acting upon the ligases promote diversification or conservation of the sequence, in relation to protein function? This Chapter addresses these questions through evolutionary analysis of REL1 and REL2 using an open source program for Molecular Evolutionary Genetic Analysis (MEGA).

It has been suggested that the components of the editosome have been subjected to the forces of constructive neutral evolution, where the absence of positive selection allowed the evolution of the protein machinery under neutral conditions (Lukes *et al.*, 2005; Lukes *et al.*, 2009; Gray, 2012). Under this neutrality the editing process established itself and subsequently became essential and from this theory it could be expected that proteins are now under purifying selection to maintain protein sequence. Although the evolutionary forces acting upon the RNA editing machinery as a whole have been discussed at length, selection upon individual components has not yet been addressed. Sequences from KREPA3 have also been used in the following analysis, since this protein main role in editing is providing stability and not catalysis to the editing process. This protein connects the two subcomplexes of the editosome and is essential to the integrity of the whole complex (Brecht *et al.*, 2005; Guo *et al.*, 2008; Niemann *et al.*, 2009; Gao *et al.*, 2010). Although this approach cannot clearly distinguish whether a protein is essential or non-essential, since the latter may still confer a selective advantage to the cell, we can still ascertain a proteins importance to the cell evolutionarily.

4.2 Methods

4.2.1 Identification of REL1, REL2 and KREPA3 genes

REL1 and REL2 genes present in the genomes of *T. brucei, T. vivax, T. congolense* and *T. cruzi* were successfully identified using the TriTryp online database (http://www.tritrypdb.org). Sequences for *T. evansi* (Chinese strain: STIB805) REL1 and 2 were provided by Achim Schnaufer (unpublished data). BLAST analyses for RELs for *Leishmania tarentolae* were performed using the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Genes and their annotations used in the following analyses are summarised below in Table 4.1. Sequences used in subsequent analyses can be found in Appendix 7. The number of nucleotide sites analysed were as follows: REL1, 1404 bp, REL2, 1244 bp and KREPA3 1331 bp.

Table 4.1. ORFs used in this study

Species	Systematic ID REL1	Systematic ID REL2	Systematic ID A3
T. brucei 927	Tb09.1.60.2970	Tb927.1.3030	Tb927.08.620
T. brucei 427	Tb427tmp.160.2970	Tb427.01.3030	Tb947.08.620
T. brucei gambiense	Tbg972.9.2300	Tbg972.1.1840	Tbg972.8.220
T. evansi STIB805	-unpublished	-unpublished	-unpublished
T. congolense	TcIL3000.9.1420	TcIL3000.1.1450	TcIL3000.8.100
T. vivax	TvY486_0901490	TvY486_0101350	TvY486_0800080
T. cruzi (*Brener Esmeraldo-like)*	Tc00.1047053511585.20	Tc00.1047053506363.110	Tc00.1047053510857.40
L. tarentolae	AY148476.1	AY148475.1	LtaP07.1150

4.2.1 Phylogenetic tree construction of REL1 and REL2

The evolutionary history between trypanosome species and strains of *T. brucei* (totalling 8 amino acid sequences, listed in Table 4.1) were inferred using the neighbour-joining method and the bootstrap consensus trees shown in Section 4.3 were inferred from 1000 replicates (Felsenstein, 1985; Saitou and Nei, 1987). The evolutionary distances were computed using the JTT matrix-based method (Jones *et al.*, 1992). Editosome protein ORFs were translated in accordance to the Standard Genetic Code. Evolutionary analyses were conducted in MEGA (version 5.1) (Tamura *et al.*, 2011).

4.2.2 Analysis of selection acting upon RELs using codon-based Z-test

The variance of the ratio of non-synonymous to synonymous mutations (variance dN/dS) was computed for all 8 sequences using the bootstrap method (1000 replicates). Analyses were conducted using the Kumar method, using MEGA (version 5.1). P values were considered significant at the 5% level.

The following null hypotheses were considered for the analysis of selection:

1) Neutral selection: dN = dS
2) Positive selection: The probability of rejecting the null hypothesis of strict neutrality (dN = dS) in favour of the alternative hypothesis (dN > dS)
3) Purifying selection: The probability of rejecting the null hypothesis of strict neutrality (dN = dS) in favour of the alternative hypothesis (dN < dS).

Here, dS and dN are the numbers of synonymous and non-synonymous substitutions per site, respectively. The first of these tests was considered a more stringent two-tailed test, since for neutrality to be rejected the dN/dS ratio could go either way, signifying purifying or positive selection. Tests for positive and purifying selection were one-tailed as the direction of the tests was already defined. The computational chi-squared analysis assumed normal distribution of dN/dS. All significant results are asterisked in the Results Section (Section 4.3).

4.2.3 Testing for sites of positive selection

As very few sequence sites may be under positive selection at any one time, positive selection may be easily overlooked in evolutionary analysis (Pond and Frost, 2005). Six of the eight sequences available (excluding Tb427tmp.160.2970 and Tbg972.9.2300 for REL and Tb427.01.3030 and Tbg972.1.1840 for REL2 and Tb947.08.620 and Tbg972.8.220 for KREPA3) were used in the analysis, using the open source HyPhy package (Hypothesis testing using Phylogenies) available at the free public server, http://www.Datamonkey.org. Sequences that were omitted represented strains and sub-species of six sent for HyPhy analysis. All start and stop codons were removed from the sequences prior to analysis.

We used Random Effective Likelihood and a general Reversible model of nucleotide substitution (REV). Random Effective Likelihood is a codon-based maximum likelihood selection method developed by Nielsen and Yang (1998). This method allows both synonymous and non-synonymous rate variation across different sites and can be used with low divergence alignments (Pond and Frost, 2005). REV

allowed correction for the different biases of nucleotides.

4.2.4 Calculation of dN/dS ratios of RELs to infer rate of evolution

If REL2 is non-essential, it will be subject to neutral evolution and therefore accumulate a greater proportion of non-synonymous mutations over time compared with REL1. This analysis is also a measure of selective pressure, but does not create a test statistic as with the analysis of selection (Nielsen and Yang, 1998; Yang, 2000). All 8 nucleotide sequences were subjected to separate synonymous and non-synonymous analyses, conducted using the Kumar model (Tamura *et al.*, 2011). Evolutionary analyses were conducted in MEGA5.1, where all positions containing gaps and missing data were eliminated (leaving a total of 456 positions). The dN/dS ratios for REL1 and REL2 were subjected further to a Mann-Whitney U test using Minitab 16 and dN/dS ratios were expressed graphically as a box plot (Minitab 16) and scatter plot (Graphpad prism 6).

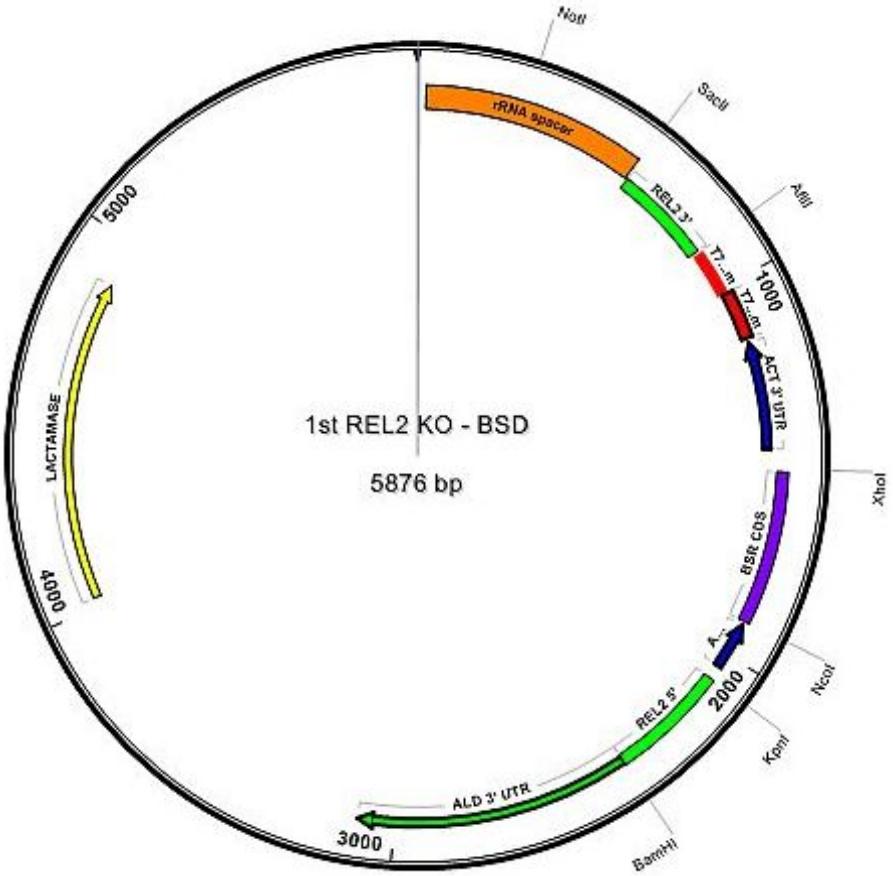
4.2.5 Pairwise comparisons of REL evolution

This approach was taken to determine whether the rate of divergence of REL1 and REL2 differed. MEGA5.1 was used to compare the number of amino acid substitutions per site for each of the 8 gene sequences shown in Table 4.1. Analyses were conducted using a JTT based matrix model (Jones *et al.*, 1992).

4.2.6 Attempted construction of a REL2 null mutant

The following approach was undertaken to ascertain whether there was a possibility of REL2 being essential in *T. brucei* BSFs, since RNAi is never complete in ablating gene expression. No ectopic copy was constructed or transfected into 427 or cKO REL1 cells, since REL2 is thought to be non-essential. This method aimed to replace the REL2 alleles (gene ID Tb.11.01.8470) by a drug resistant cassette, initially for blasticidin and nourseothricin and subsequently for puromycin, using ends out recombination, employing a plasmid (see Figure 4.1) and later a PCR-based approach to disrupt gene loci (see Table 4.2).

Drug resistant cassettes for nourseothricin and puromycin were amplified from plasmids pLEXSY-hyg-SAT2 and pGEM-PURO (kind gifts from Sean Dean and Matt Gould) using primers sets D and E and F and G, respectively. Primer sets are listed in Figure 4.1. REL2 flanking regions were amplified from wt427 gDNA using standard 50 µl Expand High Fidelity PCR (Roche). Briefly, 2.6 U Taq Expand, 100 mM of each of dNTPs, 300 nM of each primer and 1 × buffer (with $MgCl_2$) were used for each PCR reaction. PCR reactions were subjected to the following conditions: 95 °C 120s and 35 cycles of (95 °C 30s, 50 °C 60s and 72 °C 120s), followed by incubation at 72 °C for 8 minutes. PCR products were verified by electrophoresis before Nucleospin clean up, ligation into pGEM. Miniprep DNA was sent for Sanger sequencing (big dye reaction) to verify the sequences. 500 ng of pLEW100v51d-BSD plasmid and 250 ng of 5'REL2-pGEM were digested with 10 U



Name	Sequence	Details
A	5′-ATAGCGGCCGCTCCTTGCTGAAGATGTTGCGTC	F' 5' REL2 flanking region
B	5′-ATACTCGAGGCAACTCAGGGATCAATATATGATAG	R' 5' REL2 flanking region
C	5′-ATACCCGGGAAGGAGGAAATGACACCAGACCAG	F' 3' REL2 flanking region
D	5′-ATACCATGGCACCACCAATTCTCCGCAG	R' 3' REL2 flanking region
E	5′-ATACCATGGATGAAGATTTCGGTGATCCCTG	F' SAT
F	5′-ATACTCGAGGTTAGGCGTCATCCTGTGCTC	R' SAT
G	5′-ATACCATGGATGACCGAGTACAAGCCCAC	F' PURO
H	5′-ATACTCGAGTCAGGCACCGGGCTTG	R' PURO

Figure 4.1. pLEW100v51d-BSD based REL2 KO constructs

This approach made use of ends out recombination events during cell division to replace the ORF of each REL2 with that of a nourseothricin (SAT) or puromycin (PURO) drug resistant cassette. Blasticidin (BSD) was already present in this vector. Only the 1st KO construct is shown here (blasticidin-based). The 2nd KO constructs were derived from the 1st through excision of the BSD gene with *Nco*I and *Xho*I and ligation of the SAT or PURO genes in its stead.

of *Bam*HI and *Kpn*I (NEB) as per manufacturer's instruction. A total of 10 µg of digested plasmid (firstly *Not*I and subsequently *Bam*HI/*Age*I) was used to transfect log phase cKO REL1 and wt 427 bloodstream trypanosomes (for transfection protocol, see Section 2.2.4).

Full digestion was verified by gel electrophoresis, gel bands corresponding to pLEW100vBSD backbone and the 5'REL2 fragment were excised and purified. 0.5 ng of 5' REL2 flanking sequence was ligated into 10 ng plasmid. Minipreps were screened for correct inserts before this process was repeated with the insertion of the REL2 3' flanking region. 10ng the pLEW100vBSD-5'REL2 plasmid and 250 ng of 3'REL2 were digested with 10 U *Afl*II and *Sac*II (NEB). Once 1st knock-out (KO) construct (pLEW100vBSD-KO1) sequences were verified by Sanger sequencing, plasmids were cloned and extracted maxiprep DNA was used for transfections. To create the two versions of the 2nd KO construct the 1st KO construct was doubly digested, alongside pGEM-PURO or pGEM-SAT with 10 U *Nco*I and *Xho*1, this directly replaced the blasticidin resistance cassette with puromycin or nourseothricin, respectively. Details of cloning procedures used here can be found in Sections 2.2.2 and 3.2.10.

For the PCR-based knock out approach, blasticidin and puromycin-specific primers (Table 4.2) were used to amplify drug resistant cassettes from available vectors pLEW100v51d-BSD, and pGEM-PURO, kind gifts from Sean Dean and Matt Gould. This strategy was designed using the techniques described by Gaud *et al.*, 1997 and Oberholzer *et al.*, 2006. Initially a 50 bp overhang was used, and subsequently a further PCR step was used to lengthen these to 100 bp. PCR fragments were amplified from the primer sets in Table 4.2 in 50 µl reactions using

expand High fidelity PCR system (Roche) and a gradient PCR reaction. Puromycin and blasticidin were amplified from 2 ng of pGEM plasmids containing the respective drug resistant cassettes, using standard reaction condition described before. PCR reactions were subjected to the following conditions: 95 °C 120s and 30 cycles of (95 °C 60s, 55 °C +/- 10 °C and 72 °C 120s), followed by incubation at 72 °C for 8 minutes. Cleaned PCR products were sent for Sanger sequencing (Big dye) reaction to verify they were the correct sequence. 2 µg of PCR product and 10 µg of digested plasmid was used for each transfection. Prior to transfection, the PCR reactions were cleaned with Nucleospin II (Machery-Nagel) and by ethanol precipitation in the presence of glycogen.

Aside from the amount of PCR product used, transfections with prepared, digested plasmids and PCR products were performed as previously described in Chapter 2, Section 2.2.2. Both plasmid and PCR product transfectants were selected for with 5 µg/ml of blasticidin and 0.1 µg of puromycin, alongside positive controls (pLEW79-TDP1-BSD, kind gift Roberta Carloni) containing the blasticidin drug resistance cassette. Clones were picked after 5 days of growth in HMI-9 media with selective drugs.

Table 4.2. Forward and reverse primers used in the PCR REL2 KO strategy

	Blasticidin 1st KO
1	F’ 5’ ***GTTTGTTGGGCGCCGGTGTGGTCAGGGGGGGCGCCTCCTTGCTGAAG***ATGGCCAAGCCTTTGTCTCAAGAAGAATC
2	R’ 5′-***TCCGCAGCCATGCACTATTTCCACGCTCTCTGGGGAGCCATATCTTTTCA***GCCCTCCCACACATAACCAGAG
	Puromycin 2nd KO
3	F’ 5′-***GTTTGTTGGGCGCCGGTGTGGTCAGGGGGGGCGCCTCCTTGCTGAAG***ATGACCGAGTACAAGCCCACGGTG
4	R’ 5′-***TCCGCAGCCATGCACTATTTCCACGCTCTCTGGGGAGCCATATCTTTTCA***GGCACCGGGCTTGCGGGTC
	REL2 UTR extension
5	F' 5'-***TGTTCAACCCTCTGCTTCCCTTTGCTTCACCTAACTACGTCGTGATTAGTGTTTGTTGGGCGCCG***
6	R' 5'-***CGTGCCAAGAGGTGCTTTTCTTTTTCTTTGCTTCCCTCCCACCACCAATTCTCCGCAGCCATGCAC***

The 5' and 3' gene flanking regions of REL2 are highlighted in bold/italic. Primers 1 - 4 provided a 50 bp overhang enabling recombination of a drug resistant cassette between the REL2 flanking regions in place of the ORF. Primers 5 and 6 were used to further amplify the PCR products of 1 and 2, and 3 and 4, to introduce 100 bp of homologous rejoin.

4.3 Results

4.3.1 Phylogenetic trees of REL1 and REL2

Neighbour-joining trees using JTT for REL1, REL2 and KREPA3 are shown respectively in Figure 4.2. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The evolutionary distances (relative to bottom scale) are representative of the number of amino acid substitutions per site. It should be noted that the evolutionary scale of KREPA3 is twice that of the RELs.

For all trees constructed, *L. tarentolae* is placed at the root. The near-collapsed branches of the *T. brucei* clade (i) group indicate a very recent divergence, and bootstrap values relaxing to 48, 66 and 70 (Figure 4.2A - C respectively) also pertain to this. In general the divergence of the proteins occurred in an identical species order, however there were exceptions to this. Figure 4.2A shows that REL1 diverged later in *T. cruzi* than in *T. vivax*, and Figure 4.2B illustrates the opposite pattern of divergence for REL2. Figure 4.2C, on the other hand, indicates a common point of divergence in KREPA3 from *T. vivax* and *T. cruzi.*

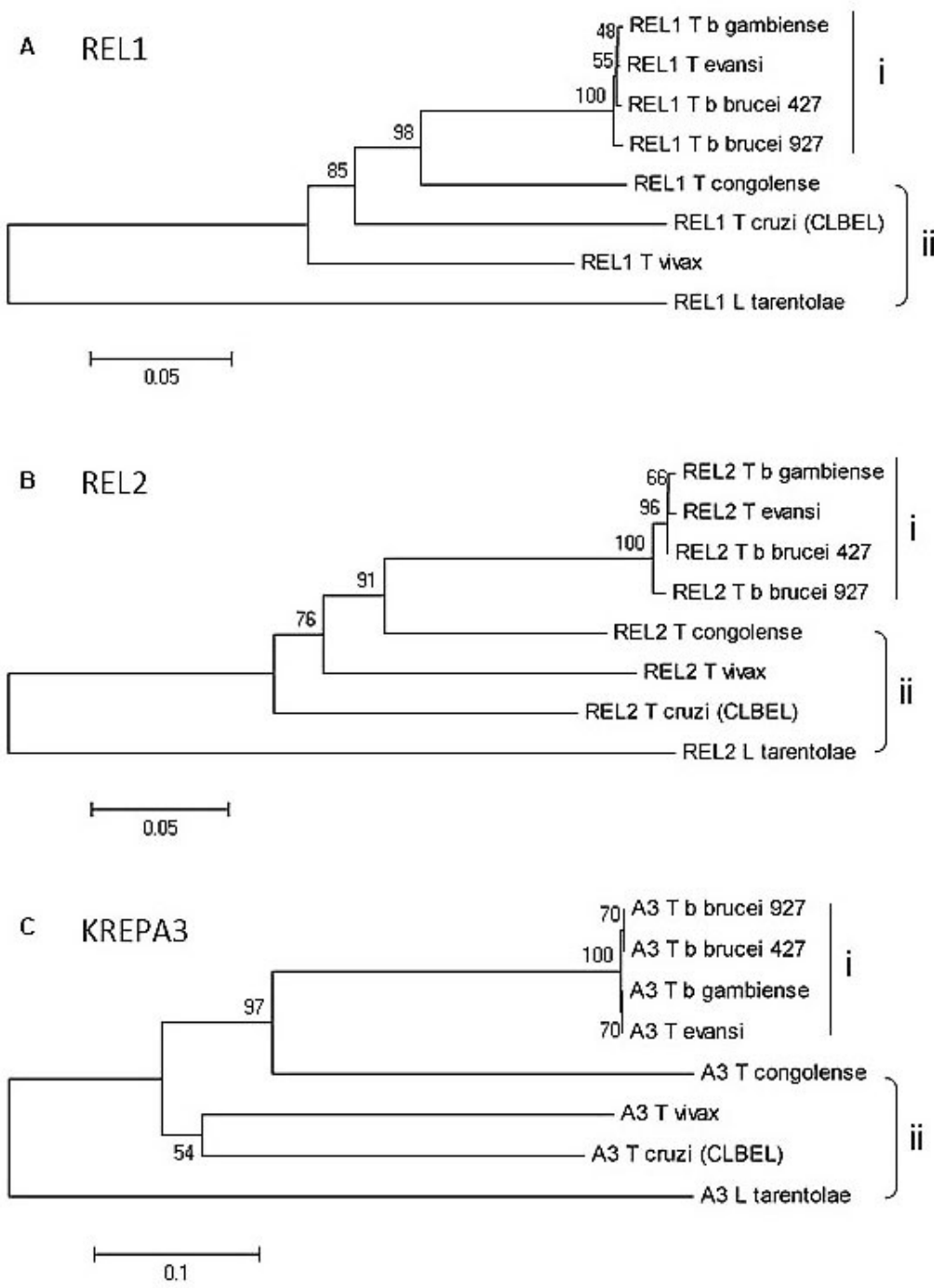


Figure 4.2. Neighbour-joining trees showing phylogenetic relationships

Numerical values pertain to bootstrap values. Trees were constructed for REL1 (A), REL2 (B) and KREPA3 (C). The evolutionary distances (expressed as a scale below the trees) is expressed as the observed proportions of amino acid pairs, between a pair of sequences, with their divergence time. For ease, the *T. brucei* clade is marked i and all other trypanosomatids ii.

4.3.2 Codon-based Z-tests to test for selection

This test was selected to determine the kind of selective pressure(s), if any, acting upon REL1 (Table 4.3), REL2 (Table 4.4) or KREPA3 (Table 4.5) by specifically comparing the relative abundance, or variance, of synonymous and non-synonymous mutations within the gene sequence. From this variance a Z-score, a standard score based upon normal distribution, is given and a P value assigned.

In the case of REL1, the null hypothesis of neutrality could be rejected (at 5% level) in 19/24 pairwise comparisons (as $p < 0.05$). The hypothesis of neutrality cannot be rejected in favour of positive selection for REL1, since no pairwise examples displayed a significant P value for their corresponding test statistic. However, the same 19/24 comparisons showed a significant P value after a test of purifying selection. This indicates that in these instances neutrality may be rejected in favour of purifying selection. In addition to these 19 comparisons there were a further two (*T. b brucei* 927 vs. *T. b brucei* 427, *T. b gambiense*) with borderline significance.

A similar pattern can be seen for REL2 and KREPA3 (Table 4.4 and 4.5), where 20/24 and 21/26 cases of neutrality could be rejected in favour of purifying selection at the 5% level. As with REL1, no cases of positive selection acting upon REL2 or KREPA3 could be found. It is clear from Tables 4.3 – 4.5 that a number of pairwise comparisons produced non-significant test statistic (in favour of the null hypothesis). These sequences can be assigned to the *T. brucei* clade (Figure 4.2), meaning a Z-score and corresponding P-value will reflect the few nucleotide changes between species and sub-species.

Table 4.3. Tests for neutrality, positive selection and purifying selection acting upon REL1 across trypanosome species and strains

Comparison		Neutral Evolution 2-tailed		Positive Selection 1-tailed		Purifying Selection 1-tailed	
Species 1	Species 2	Z-score	P value	Z-score	P value	Z-score	P value
T b gambiense	*T. evansi*	-0.523	0.602	-0.526	1.000	0.513	0.304
T b gambiense	*T. congolense*	-2.813	0.006*	-2.973	1.000	2.941	0.002*
T b gambiense	*T. vivax*	-2.733	0.007*	-2.752	1.000	2.690	0.004*
T b gambiense	*T. cruzi*	-3.018	0.003*	-3.013	1.000	3.119	0.001*
T brucei 427	*T b gambiense*	1.469	0.144	1.505	0.067	-1.517	1.000
T brucei 427	*T. evansi*	-0.416	0.678	-0.406	1.000	0.402	0.344
T brucei 427	*T. congolense*	-2.818	0.006*	-2.961	1.000	2.909	0.002*
T brucei 427	*T. vivax*	-2.757	0.007*	-2.771	1.000	2.694	0.004*
T brucei 427	*T. cruzi*	-3.112	0.002*	-3.142	1.000	3.326	0.001*
T brucei 927	*T brucei* 427	-1.572	0.118	-1.566	1.000	1.584	0.058+
T brucei 927	*T b gambiense*	-1.601	0.112	-1.588	1.000	1.619	0.054+
T brucei 927	*T. evansi*	-2.037	0.044*	-2.030	1.000	2.040	0.022*
T brucei 927	*T. congolense*	-2.827	0.006*	-2.955	1.000	2.922	0,002*
T brucei 927	*T. vivax*	-2.803	0.006*	-2.792	1.000	2.738	0.004*
T brucei 927	*T. cruzi*	-2.734	0.007*	-2.771	1.000	2.846	0.003*
T. congolense	*T. vivax*	-6.924	0.000*	-6.393	1.000	6.131	0.000*
T. congolense	*T. cruzi*	-2.158	0.033*	-2.180	1.000	2.167	0.016*
T. congolense	*L. tarentolae*	-7.150	0.000*	-6.992	1.000	7.163	0.000*
T. cruzi	*L. tarentolae*	-3.612	0.000*	-4.073	1.000	4.404	0.000*
T. evansi	*T. congolense*	-2.671	0.009*	-2.759	1.000	2.774	0.003*
T. evansi	*T. vivax*	-2.620	0.010*	-2.606	1.000	2.576	0.006*
T. evansi	*T. cruzi*	-2.841	0.005*	-2.757	1.000	2.924	0.002*
T. vivax	*T. cruzi*	-2.347	0.021*	-2.411	1.000	2.389	0.009*
T. vivax	*L. tarentolae*	-2.303	0.023*	-2.304	1.000	2.304	0.011*

* denotes significant P value (5% level) based upon variance of dN-dS.

+ denotes borderline significance at 5% level.

In all cases the null hypothesis is that the protein is subject to neutrality (dN = dS). Therefore the asterisks present under neutral and purifying selection, pertaining to a significant P value, and requires rejection of the null hypothesis in favour of purifying selection, where dN<dS. Pairwise comparisons that could not be calculated are not shown.

Table 4.4. Tests for neutrality, positive selection and purifying selection acting upon REL2 across trypanosome species and strains

Comparison		Neutral Evolution 2-tailed		Positive Selection 1-tailed		Purifying Selection 1-tailed	
Species 1	Species 2	Z-Score	P value	Z-Score	P value	Z-Score	P value
T b gambiense	*T. congolense*	-5.639	0.000*	-5.673	1.000	5.445	0.000*
T b gambiense	*T. vivax*	-2.211	0.029*	-2.183	1.000	2.189	0.015*
T b gambiense	*T. cruzi*	-3.471	0.001*	-3.605	1.000	3.529	0.000*
T b gambiense	*L. tarentolae*	-2.075	0.040*	-2.095	1.000	2.079	0.020*
T brucei 427	*T.evansi*	1.003	0.318	1.043	0.150	0.000	1.000
T brucei 427	*T. congolense*	-5.700	0.000*	-5.748	1.000	-1.036	0.000*
T brucei 427	*T. vivax*	-2.234	0.027*	-2.203	1.000	5.649	0.015*
T brucei 427	*T. cruzi*	-3.558	0.001*	-3.682	1.000	2.203	0.000*
T brucei 427	*L. tarentolae*	-2.095	0.038*	-2.110	1.000	3.620	0.019*
T brucei 927	*T brucei* 427	-1.869	0.064	-1.837	1.000	2.102	0.034*
T brucei 927	*T b gambiense*	-1.690	0.094	-1.662	1.000	1.842	0.048+
T brucei 927	*T. evansi*	-1.639	0.104	-1.633	1.000	1.675	0.053+
T brucei 927	*T. congolense*	-6.504	0.000*	-6.572	1.000	1.625	0.000*
T brucei 927	*T. vivax*	-2.252	0.026*	-2.219	1.000	6.339	0.014*
T brucei 927	*T. cruzi*	-3.313	0.001*	-3.381	1.000	2.223	0.001*
T brucei 927	*L. tarentolae*	-2.378	0.019*	-2.378	1.000	3.365	0.009*
T. congolense	*T. vivax*	-2.179	0.031*	-2.145	1.000	2.391	0.017*
T. congolense	*T. cruzi*	-2.169	0.032*	-2.164	1.000	2.156	0.016*
T. cruzi	*L. tarentolae*	-2.067	0.041*	-2.064	1.000	2.172	0.021*
T. evansi	*T. congolense*	-5.695	0.000*	-5.743	1.000	0.744	0.000*
T. evansi	*T. vivax*	-2.234	0.027*	-2.203	1.000	2.064	0.015*
T. evansi	*T. cruzi*	-3.557	0.001*	-3.681	1.000	5.643	0.000*
T. evansi	*L. tarentolae*	-2.095	0.038*	-2.110	1.000	2.203	0.019*
T. vivax	*T. cruzi*	-3.572	0.001*	-3.721	1.000	3.671	0.000*

* denotes significant P value (at 5% level) based upon variance of dN-dS.

+ denotes borderline significance at 5% level.

In all cases the null hypothesis is that the protein is subject to neutrality (dN = dS). Therefore the asterisks present under neutral and purifying selection, pertaining to a significant P value, and requires rejection of the null hypothesis in favour of purifying selection, where dN<dS. Pairwise comparisons that could not be calculated are not shown.

Table 4.5. Tests for neutrality, positive selection and purifying selection acting upon KREPA3 across trypanosome species and strains

Comparison		Neutral Evolution 2-tailed		Positive Selection 1-tailed		Purifying Selection 1-tailed	
Species 1	Species 2	Z-Score	P value	Z-Score	P value	Z-Score	P value
T. b. gambiense	*T. evansi*	-1.047	0.297	-1.030	1.000	1.050	0.148
T. b. gambiense	*T. congolense*	-2.161	0.033*	-2.151	1.000	2.147	0.017*
T. b. gambiense	*T. vivax*	-2.072	0.040*	-2.075	1.000	2.108	0.019*
T. b. gambiense	*T. cruzi*	-2.146	0.034*	-2.154	1.000	2.156	0.017*
T. b. gambiense	*L. tarentolae*	-2.176	0.032*	-2.170	1.000	2.171	0.016*
T. b. gambiense	*T.b. brucei* 427	-0.474	0.636	-0.473	1.000	0.494	0.311
T. b. brucei 427	*T.evansi*	1.054	0.294	1.042	0.150	-1.045	1.000
T brucei 427	*T. congolense*	-2.247	0.026*	-2.218	1.000	2.221	0.014*
T brucei 427	*T. vivax*	-2.051	0.042*	-2.062	1.000	2.080	0.020*
T brucei 427	*T. cruzi*	-2.078	0.040*	-2.096	1.000	2.119	0.018*
T brucei 427	*L. tarentolae*	-2.080	0.040*	-2.081	1.000	2.067	0.020*
T brucei 927	*T brucei* 427	-1.046	0.298	-1.074	1.000	1.070	0.143
T brucei 927	*T b gambiense*	-1.074	0.285	-1.092	1.000	1.123	0.132
T brucei 927	*T. evansi*	-0.474	0.636	-0.479	1.000	0.491	0.312
T brucei 927	*T. congolense*	-2.271	0.025*	-2.247	1.000	2.246	0.013*
T brucei 927	*T. vivax*	-2.041	0.043*	-2.053	1.000	2.068	0.020*
T brucei 927	*T. cruzi*	-2.063	0.041*	-2.055	1.000	2.065	0.021*
T brucei 927	*L. tarentolae*	-2.079	0.040*	-2.080	1.000	2.068	0.020*
T. congolense	*T. vivax*	-2.589	0.011*	-2.564	1.000	2.540	0.006*
T. congolense	*T. cruzi*	-4.202	0.000*	-3.844	1.000	4.180	0.000*
T. congolense	*L. tarentolae*	-2.193	0.030*	-2.200	1.000	2.198	0.015*
T. cruzi	*L. tarentolae*	-2.104	0.037*	-2.067	1.000	2.090	0.019*
T. evansi	*T. congolense*	-2.249	0.026*	-2.223	1.000	2.227	0.014*
T. evansi	*T. vivax*	-2.053	0.042*	-2.056	1.000	2.080	0.020*
T. evansi	*T. cruzi*	-2.076	0.040*	-2.089	1.000	2.113	0.018*
T. evansi	*L. tarentolae*	-2.106	0.037*	-2.116	1.000	2.091	0.019*
T. vivax	*T. cruzi*	-2.419	0.017*	-2.350	1.000	2.394	0.009*

* denotes significant P value (at 5% level) based upon variance of dN-dS.

+ denotes borderline significance at 5% level.

In all cases the null hypothesis is that the protein is subject to neutrality (dN = dS). Therefore the asterisks present under neutral and purifying selection, pertaining to a significant P value, and requires rejection of the null hypothesis in favour of purifying selection, where dN<dS. Pairwise comparisons that could not be calculated are not shown.

4.3.3 HyPhy analysis to test for positive selection

Since positive selection is easily overlooked, http://www.Datamonkey.org was used to determine if any sites were under positive selection (Pond and Frost, 2005).

Analysis of REL2 sequences revealed that no sites had a rate where dN > dS, inferring that all sites were under purifying selection (see Appendix 8 for analysis read out). Analysis of REL1 sequences found no sites under positive selection, and 298 (out of the 489 sites) were found to be under purifying selection. This indicates that, even with 31% of sites neutrally evolving, purifying selection remains a strong evolutionary force on this ligase at these sites. When subjecting KREPA3 sequences to the same analysis, no sites were found to be under positive selection, or with dN>dS, inferring that all sites were under purifying selection.

4.3.4 Using dN/dS to discern rate of evolution of editosome proteins

The dN/dS ratio was used in this study, since it is widely accepted as a measure of selection pressure (Nielsen and Yang, 1998; Yang, 2000).

Since the dN/dS ratio in all calculable cases (Table 4.6) was between 0.058 and 0.491, it was determined that the number of synonymous mutations occurring over time is considerably higher than that of non-synonymous mutations. In all but one case (*T. b brucei* 927 vs. *T. b gambiense*) KREPA3 portrayed a greater accumulation of non-synonymous compared to synonymous mutations over time than REL1 or REL2. In this example, dN/dS is greater for REL2 than for KREPA3. Mann-Whitney U test revealed that the median dN/dS ratios of REL2 and REL1

were 0.096 and 0.077, respectively (displayed as a box plot Figure 4.3 A). The difference between these two medians was significant at 0.027 (Mann-Whitney U test; 95% CI values of 0.002 and 0.027, respectively). This indicates a statistically significant difference between the rates of evolution between the two ligases. REL1 is evolving significantly slower than REL2.

The dN/dS values for the two ligases reveal a general trend for more relaxed purifying selection acting upon REL2 in comparison to REL1, shown by the deviation of the regression line towards the REL2 axis (Figure 4.3 B, left panel). Since data points are matched pairwise, this relationship between the two datasets is not related to phylogeny; rather the slope represents a functional relationship between the two ligases. This indicates that coevolution may have occurred, although the R^2 value of 0.681 does not indicate a robust relationship. This dN/dS relationship between REL1 and KREPA3 is more difficult to interpret from the graphical data (Figure 4.3 B, right panel), but from the dN/dS values given in Table 4.5 it is clear that the purifying selection is a weaker force on KREPA3 in comparison to REL1. The absence of a clear line does not indicate the coevolution of these two proteins.

Comparison		REL2			REL1			KREPA3		
Species 1	Species 2	dN	dS	dN/dS ratio	dN	dS	dN/dS ratio	dN	dS	dN/dS ratio
T b gambiense	*T.evansi*	0.002	/	/	0.001	0.002	0.427	0.000	0.003	/
T b gambiense	*T. congolense*	0.101	1.287	0.078	0.088	1.502	0.058	0.264	2.126	0.124
T b gambiense	*T. vivax*	0.151	1.781	0.085	0.129	1.696	0.076	0.337	2.101	0.161
T b gambiense	*T. cruzi*	0.143	1.423	0.100	0.133	1.728	0.077	0.303	1.992	0.152
T b gambiense	*L. tarentolae*	0.282	2.155	0.131	0.247	2.168	0.114	0.462	2.516	0.184
T brucei 427	*T b gambiense*	0.001	/	/	0.001	0.000	/	0.001	0.003	0.492
T brucei 427	*T.evansi*	0.001	0.000	/	0.000	0.002	/	0.001	0.000	/
T brucei 427	*T. congolense*	0.100	1.277	0.078	0.088	1.504	0.058	0.264	2.206	0.119
T brucei 427	*T. vivax*	0.150	1.760	0.085	0.130	1.693	0.077	0.337	2.143	0.157
T brucei 427	*T. cruzi*	0.142	1.422	0.100	0.134	1.727	0.077	0.304	2.027	0.150
T brucei 427	*L. tarentolae*	0.280	2.129	0.132	0.248	2.204	0.112	0.464	2.388	0.194
T brucei 927	*T brucei* 427	0.005	0.016	0.294	0.002	0.012	0.162	0.000	0.003	/
T brucei 927	*T b gambiense*	0.006	0.016	0.352	0.003	0.012	0.233	0.001	0.005	0.246
T brucei 927	*T.evansi*	0.006	0.016	0.368	0.002	0.014	0.138	0.001	0.003	0.491
T brucei 927	*T. congolense*	0.098	1.233	0.079	0.088	1.491	0.059	0.264	2.242	0.118
T brucei 927	*T. vivax*	0.148	1.750	0.085	0.130	1.668	0.078	0.337	2.169	0.155
T brucei 927	*T. cruzi*	0.139	1.431	0.097	0.133	1.791	0.074	0.304	2.056	0.148
T brucei 927	*L. tarentolae*	0.281	1.997	0.141	0.248	2.121	0.117	0.464	2.356	0.197
T. congolense	*T. vivax*	0.133	1.758	0.076	0.129	1.334	0.097	0.351	1.862	0.189
T. congolense	*T. cruzi*	0.120	1.703	0.070	0.125	1.848	0.068	0.358	1.493	0.240
T. congolense	*L. tarentolae*	/	/	/	0.261	1.405	0.186	0.494	1.839	0.269
T. cruzi	*L. tarentolae*	0.277	2.207	0.125	0.267	1.618	0.165	0.420	2.477	0.169
T.evansi	*T. congolense*	0.101	1.277	0.079	0.088	1.527	0.058	0.264	2.207	0.120
T.evansi	*T. vivax*	0.151	1.760	0.086	0.130	1.716	0.076	0.337	2.143	0.157
T.evansi	*T. cruzi*	0.143	1.422	0.101	0.134	1.757	0.076	0.303	2.035	0.149
T.evansi	*L. tarentolae*	0.282	2.129	0.133	/	/	/	0.462	2.441	0.189
T. vivax	*T. cruzi*	0.134	1.417	0.094	0.133	1.796	0.074	0.292	1.865	0.157
T. vivax	*L. tarentolae*	/	/	/	0.256	1.945	0.132	/	/	/

Table 4.6. Using dN/dS ratios to quantify selection pressure

e presence of / in the results denotes cases in which it was not possible to estimate evolutionary distances. Highlighted in blue is the only example where dN/dS is greater for the RELs in comparison to the A3 protein.

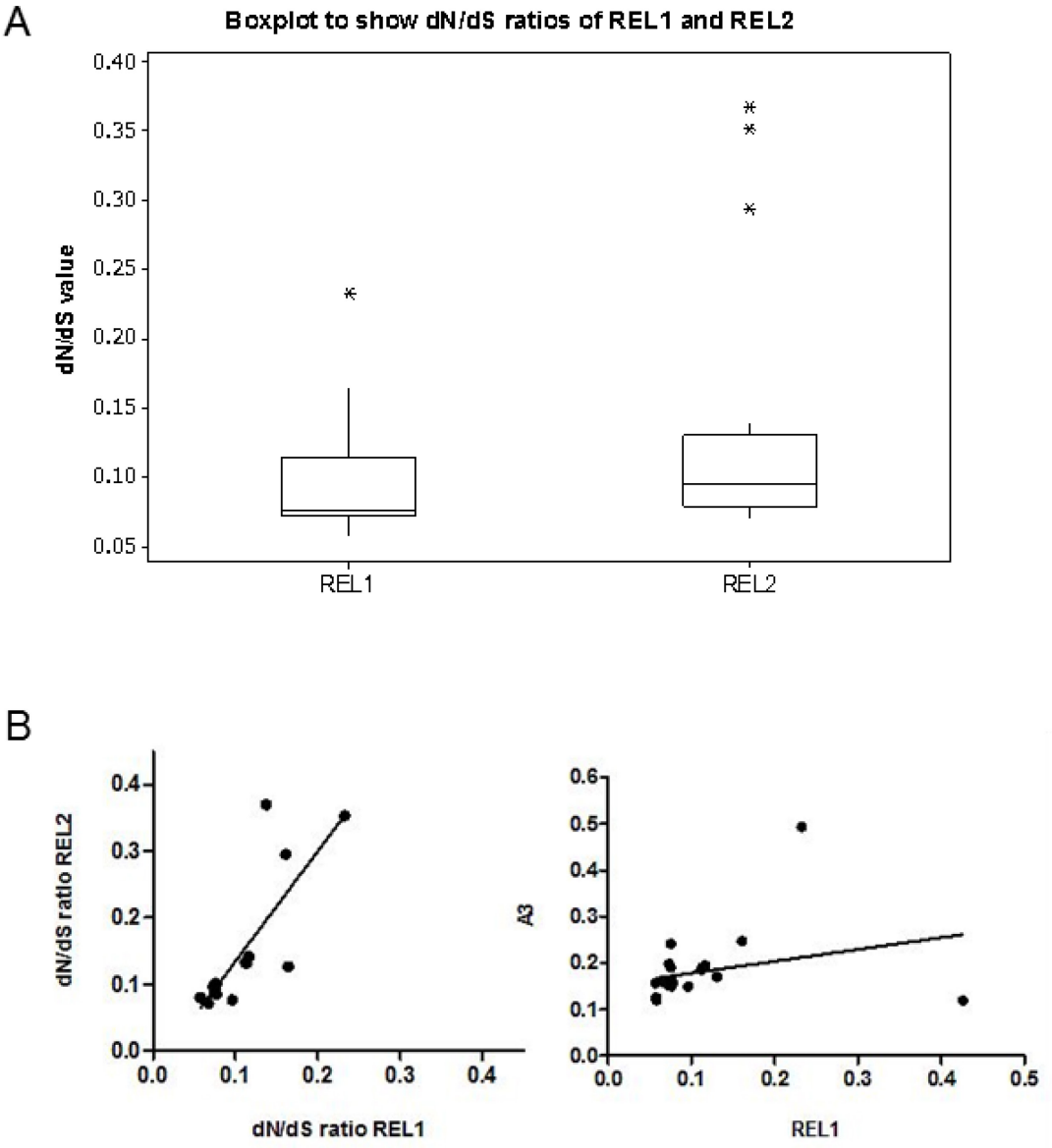


Figure 4.3. Graphical comparisons of dN/dS

A) Blot plots were drawn for REL1 vs. REL2 and a Mann-Whitney U test was carried out giving: medians for REL2 0.096 and REL1 0.077, P = 0.027.

B) dN/dS values were plotted for two separate protein comparisons and a regression line was drawn. Left panel: REL1 vs. REL2 (R^2 = 0.681, P = < 0.0001). Right panel: REL1 vs. A3 (R^2 = 0.073, P = 0.2248).

4.3.5 Pairwise comparison of the RELs

This method was used to crudely infer the differences in the evolutionary divergence (or the variance of speed of evolution) between the two ligases. The distributions of the mean evolutionary divergence were found to differ slightly across all examined trypanosomatid species (see Table 4.7 below). If the ligases were evolving at exactly the same pace, then the difference between the mean would be 0, but the difference in the means indicates greater sequence divergence of REL2 over time, in comparison to REL1.

Table 4.7 Pairwise comparisons of RELs to infer evolutionary divergence

Comparison		REL1			REL2			Difference in means
Species 1	Species 2	Dist	Std. Err	-2SE	Dist	Std. Err	+2SE	(REL2-1)
T brucei 927	*T.evansi*	0.004	0.003	-0.002	0.012	0.005	0.023	0.025
T brucei 927	*T. congolense*	0.131	0.018	0.096	0.183	0.020	0.224	0.128
T brucei 927	*T. vivax*	0.185	0.023	0.140	0.238	0.025	0.288	0.148
T brucei 927	*T. cruzi*	0.199	0.022	0.156	0.255	0.026	0.306	0.150
T brucei 927	*L. tarentolae*	0.430	0.035	0.361	0.484	0.041	0.566	0.206
T. congolense	*T. vivax*	0.184	0.023	0.138	0.222	0.023	0.268	0.130
T. congolense	*T. cruzi*	0.207	0.023	0.161	0.230	0.024	0.277	0.115
T. congolense	*L. tarentolae*	0.423	0.035	0.353	0.460	0.041	0.542	0.189
T.evansi	*T. congolense*	0.131	0.017	0.096	0.187	0.021	0.229	0.133
T.evansi	*T. vivax*	0.185	0.023	0.140	0.242	0.025	0.293	0.153
T.evansi	*T. cruzi*	0.199	0.022	0.156	0.261	0.026	0.313	0.158
T.evansi	*L. tarentolae*	0.428	0.035	0.358	0.485	0.041	0.567	0.209
T. vivax	*T. cruzi*	0.214	0.024	0.167	0.233	0.024	0.282	0.115
T. vivax	*L. tarentolae*	0.432	0.037	0.358	0.485	0.040	0.566	0.221
T. cruzi	*L. tarentolae*	0.460	0.037	0.385	0.451	0.040	0.531	0.145

4.3.6 Creation of a REL2 null mutant

All attempts at creating a null mutant using plasmid or PCR derived constructs were unsuccessful in both the cKO REL1 and wt 427 parental backgrounds. Unsuccessful transfections are tabulated below in Table 4.8. Due to time constraints this part of the study was not pursued further, but from the evolutionary analyses REL2 would be expected to confer a selective advantage to cell growth and survival, as REL1 does.

Positive controls were used throughout from a blasticidin containing plasmids (pLEW79-TDP1-BSD, kind gift Roberta Carloni), which allowed us to determine that at least one of the selective drugs was working as expected.

Table 4.8 Transfections for REL2, using ends out recombination approach

Starting construct	Construct name (drug cassette)	Failed transfections
Plasmid REL2 3'UTRs pLEW100v5b1d	1st KO (Nourseothrycin)	4 attempts in wt/cKO REL1 cell lines
	1st KO (Blasticidin)	3 attempts in wt/cKO REL1 cells
	2nd KO (Puromycin)	3 attempts in wt cells
PCR product REL2 3'UTRs	1st KO 50 bp overhang (Blasticidin)	2 attempts wt cells
	St KO 100 bp overhang (Blasticidin)	1 attempt wt cells
	2nd KO 50 bp overhang (Puromycin)	2 attempts wt cells

4.4 Discussion

This Chapter aimed to assess the essentiality of REL2, in comparison to REL1, through evolutionary analysis. Prior hypotheses have suggested that the editosome proteins had previously evolved neutrally, whereby the accumulation of synonymous mutations has lead to molecular complexity. Once editing became an essential process, these proteins could be conserved through the actions of purifying selection (Lukeš *et al.*, 2005; Lukeš *et al.*, 2009; Gray, 2012). There are two potential opposing evolutionary forces acting upon the editosome proteins when neutral evolution is not acting. These consist of purifying selection aiming to preserve the protein function, and therefore sequence, and positive selection driving protein evolution over time.

Protein evolution was visualised using phylogenetic methods. A branch corresponding to partitions reproduced with less than 50% bootstrap replicates is considered collapsed, and may be indicative of close genetic distance (Felsenstein, 1985). This phenomenon can be clearly seen in the *T. brucei* clade, and suggests a very recent speciation event.

The results summarised in Table 4.3 suggest REL1 evolves neutrally in certain pairwise comparisons. However, it must be recognised that the small genetic distances between *T. b brucei* 927, *T. b brucei* 427, *T. b gambiense* and *T. evansi* confounds the dN/dS calculations employed in this analysis. Of the five non-significant cases (*T. b gambiense* vs. *T. evansi, T. b brucei* 427*, T. b brucei* 927 and *T. b brucei* 427 vs. *T. b brucei* 927*, T. evansi*) all were present in the *T. brucei* clade and exhibited near-collapsed branches (Figure 4.2). The near-collapsed branches

indicate that these sequences are not divergent enough to confidently analyse the sequence difference, which was taken into consideration with the subsequent comparative evolutionary comparisons. Even taking this into account, the RELs of *T. b gambiense* and *T. evansi* were found to be as closely related to *T. b brucei* strains as 927 and 427 RELs are to each other. The pattern of evolution illustrated by the editosome genes generally adhere to the pattern of trypanosome evolution seen in the major housekeeping genes; glycosomal glyceraldehyde phosphate dehydrogenase (gGAPDH) and small subunit (SSU) rRNA. These studies suggested a monophyletic origin for *T. brucei* and distinct clustering of *T. brucei, T. equiperdum* and *T. evansi* sequences (Maslov *et al.*, 1996; Brun *et al.*, 1998; Hughes and Piontkivska, 2003; Claes *et al.*, 2005; Hamilton *et al.*, 2007). Results presented here indicate a recent speciation event between *T. brucei* and *T. evansi* (Figure 4.2). A more recent study, using SL RNA repeat sequences, presented by Lai *et al.* (2008) suggests that *T. brucei, T. equiperdum* and *T. evansi* have also undergone a recent divergence, but does not suggest a monophyletic origin. However, results presented here suggest, as with the housekeeping genes, monophyly, and also reveal a *T. brucei* clade that encompasses *T. evansi* (Maslov *et al.*, 1996; Brun *et al.*, 1998; Hughes and Piontkivska, 2003; Claes *et al.*, 2005; Hamilton *et al.*, 2007). The only difference across the three trees pertained to the positioning of *T. vivax* and *T. cruzi*. The *T. vivax* sequences used here are from an African isolate. However *T. vivax* escaped Africa due to its ability to be mechanically transmitted and this discrepancy in tree positioning may reflect this, as it has been observed that species of trypanosomes cluster phylogenetically according to their hosts reflecting adaptation of proteins and genes to the host environment (Hamilton *et al.*, 2007).

It was taken into consideration that the number of sequences involved in this study was small, which may have decreased the robustness of the Z-score value, and corresponding P-value, calculated from the test of selection. But it was clear from the tests for selection that purifying selection was acting upon all three proteins used in this analysis. Assumptions of normality were also made when generating the test for selection P-value. To determine the strength of this selective force, dN/dS ratios were calculated separately, without a test statistic. What can be obtained from these analyses is that positive selective forces appear to be the weaker forces acting upon these editosome proteins, and so the proteins are not evolving neutrally. It should also be noted that the strength of purifying selection pressure seems to increase on the proteins in the order: KREPA3, REL2, and REL1.

Mann-Whitney U test results indicate that the difference in the median dN/dS values were significantly greater for REL2 than REL1. Not all cases of dN/dS could be calculated, most likely due to a very recent evolutionary divergence (in the case of within the *T. brucei* clade) or a highly divergent lineage as found with the comparison between the salivarian trypanosomes and the *L. tarentolae* out-group. Nonetheless, the evolutionary distance calculated from both REL1 and REL2 sequence data is low, indicating that both genes have important roles due to the small number of non-synonymous mutations picked up over time. Put differently, the strength of purifying selection acting upon the RELs is high, as the dN/dS ratios are low, indicating a need for trypanosomatids to retain REL function, and consequently, sequence over evolutionary time. This dN/dS ratio is noticeably higher in KREPA3 in comparison to the RELs, indicating a greater accumulation of non- synonymous mutations in this protein over time in all but a single case (highlighted in Table 4.6).

It should also be noted that there were several instances of dS being greater than 1, inferring saturation of the test parameter, which is indicative of multiple substitutions at the same site (Graur and Li, 2000). As a consequence, these calculated dN/dS ratios may obscure the magnitude of positive selection acting upon the editosome proteins. Because of this, and since positive selection can be easily overlooked in such evolutionary analyses, it was important to subject the editosome sequences to HyPhy analysis. This was used to demonstrate that all REL2 and KREPA3 codons are sites where purifying selection occurs. In comparison 61% of REL1 sites are under purifying selection. On closer inspection of the alignment of the 6 REL1 genes, it was apparent that the sites indicated to be under purifying selection were located across the 5 motifs involved in ATP and substrate binding. From this evolutionary perspective, it can be inferred that functional REL2, like KREPA3, confers a selective advantage to the cell, since all codons of this ligase have been selected for conservation across the trypanosome species analysed. Hyphy is a sophisticated method to detect codons under positive selection as it allows transition and transversions to take place at different rates, since the mutation rate of a nucleotide sequence is affected by the sequence itself (Charlesworth and Charlesworth, 2010). Hyphy also takes the whole phylogenetic tree generated from sequence alignment, increasing the robustness of results, even if few sites are analysed and data points cannot be treated separately due to multiple pairwise comparisons. This is important to note because multiple testing can lead to false positive and negative results (Charlesworth and Charlesworth, 2010).

The distance values, as calculated from Pairwise Analysis, for REL1 and REL2 also suggest a greater divergence of sequence, through the accumulation of

non-synonymous mutations, for REL2.

HyPhy analysis seemingly contradicts the other pairwise comparisons of selection, since the dN/dS values were greater for KREPA3 and REL2 than for REL1. The Datamonkey output suggested that 298 sites, but not 100% of sites were under negative selection for REL1. It is important to note that, whilst these sites were flagged for negative selection, these do not reflect the catalytically essential sites, since the codons influence this result. What should be focused on is the absence of sites under positive selection, which can be easily overlooked in dN/dS pairwise testing (Pond and Frost, 2005).

REL1 may also have a role in general RNA repair (See Chapter 3), having substrates born of deletion and miscleavage events. In this model, REL2 is discriminating of its substrates and therefore may be essential to work alongside REL1 as a proofreading ligase (Cruz-Reyes *et al.*, 2002). It is possible that both plasmid and PCR-based approaches failed due technical difficulties, such as the presence of two stabilising actin UTRs flanking the drug resistance cassettes. It is also possible that transfections using PCR-products may require optimisation, since it has been reported that this approach has varying efficiency and requires varying amounts of PCR-product to produce stable transfectants (Oberholzer *et al.*, 2006). The viability of REL2 RNAi cells, which show signs of deformation in one instance (O'Hearn *et al.*, 2003), may indicate that only a small amount of REL2 enzyme may be required to function within the editosome as a proofreading enzyme. Indeed, it is also possible that the difficulties in obtaining a REL2 null mutant may also reflect its importance to the cell. However, the approaches described here did not yield a single, stable KO cell line, which is unusual since, RNAi lines described in the

literature repress REL2 mRNA by more than 50% (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz *et al.*, 2002; Gao and Simpson, 2003; O'Hearn *et al.*, 2003).

4.5 Conclusions and Outlook

All of these analyses highlight the importance of the REL2 protein in an evolutionary context, and may add weight to the working hypothesis that both REL1 and REL2 function are important to the cell. This may be due to their distinct catalytic properties, as previously described *in vitro* (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002; Palazzo *et al.*, 2003; Schnaufer *et al.*, 2003). In this hypothesis, which is perhaps contrary to current RNAi data on REL2, both ligases have their own important role in RNA editing.

It would be important to try and obtain a REL2 null mutant to validate a non-essential role of REL2, which did not succeed in the time frame available, especially since the evolutionary data presented here suggests importance of REL2 to the cell.

Interestingly, KREPA3 also appears to be under purifying selection, as with both ligases. This data on its own suggests that the editing machinery has been under the purifying selection in the genuses *Leishmania* and *Trypanosoma*. This is not in itself contrary to the idea of constructive neutral evolution, since the editosome machinery may have been evolving neutrally long before the branching of *Leishmania* and *Trypanosoma*. What is important to note is that the evolution of the editosome since this split has not been spurned by positive selection. It would be interesting to determine whether constructive neutral evolution can be proven. To discern this, more editosome genes from more species at the foot of the Euglenozoa should be subjected to the same evolutionary analyses, once the sequence data becomes available.

Chapter 5

Concluding Discussion

5.1 Summary of main findings

This study aimed to provide fundamental knowledge concerning the essentiality of REL1 (Chapter 1). Genetic complementation was used to assess whether it was the subcomplex positioning or catalytic properties that made REL1 essential (Chapter 2). Next, the 5' ends of ND7 and RPS12 transcripts were mapped, via a 5' linker ligation, from cells before and after REL1 ablation to determine what editing events were governed by REL1 ligation, hypothesising that this would provide clues to its essential role in the process (Chapter 3). Finally, REL1, REL2 and KREPA3 protein sequences were subjected to evolutionary analysis to determine the relative strength of selective forces acting upon them to conserve sequence, and hence catalytic activity, and thus determine the importance of the ligases through evolutionary analysis (Chapter 4). The three complementary experimental approaches provided scope to assess this topic from the level of protein function and RNA processing, as well as providing an evolutionary perspective on the presence of two ligases within the 20S editosome.

The present suggest that the REL1 catalytic domain (CD) is essential at its specific position within the deletion subcomplex, since only full length, tagged REL1 (REL1-TAP) could rescue the growth phenotype caused by ectopic REL1 ablation. Whilst data presented in Chapter 2 shows (as suggested from glycerol gradient sedimentation) that all tagged chimeric ligases integrated into their intended positions within the editosome, the CD and interactive domain (ID) swap had greatly reduced the proteins catalytic activities of the proteins compared to REL1-TAP *in vitro*. It is known from studies with REL1, T4Rnl1 and T4Rnl2 that the CDs of these ligases can form covalent AMP-enzyme ("adenylylated") intermediates, and hence

are at least partially active, without the presence of their full C-terminal domain (Ho *et al.*, 2004; Deng *et al.*, 2004; Wang *et al.*, 2007). However, results presented in Chapter 2 indicate more cross-talk than expected between the REL CDs and IDs, which made it difficult to create catalytically competent chimeric ligases than expected. The catalytically repressed chimeric ligases confounded the interpretation of the growth curve data slightly: it cannot be ruled out that the failure to rescue the growth phenotype from ablating REL1 were caused by insufficient catalytic activity of the chimeric ligases, rather than a strict requirement for the REL1 CD in the deletion subcomplex. As with some previous studies involving integration of ligases into native editosomes, the RECC lost integrity when REL1 was ablated (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002). This was apparent only after glycerol gradient and subsequent TAP analysis and the significance of this observation *in vivo* is uncertain, but is important to take into consideration since this may have also affected the ability of these editosomes in processing RNA.

Shallow and deep sequencing analyses of REL1 substrates revealed frequent products of deletion editing at deletion editing sites and misediting at editing and non-editing sites, in the absence of REL1. Peaks illustrated by the 5' start graphs correspond to the substrates that had accumulated in the absence of REL1 and indicate the products of endonucleolytic cleavage that REL2 is unable to re-ligate. To add to this, there were fewer examples of expected addition editing substrates at editing sites, in the absence of REL1 than in its presence. Misediting events seem to be the most abundant type unable to be re-ligated efficiently by endogenous REL2. Partially edited cDNAs and truncated sequences indicative of miscleavage during RNA editing has also previously been reported in the literature (Decker and Sollner-

Webb, 1990; Koslowsky *et al.*, 1990; Sturm and Simpson, 1990b; Koslowsky *et al.*, 1991). Results from this Chapter also confirm the general, but not strict, 3' to 5' progression of editing, as has previously been reported in the literature (Decker and Sollner-Webb, 1990; Sturm and Simpson, 1990b; Koslowsky *et al.*, 1991). Thus, the essentiality of REL1 may be due to this role as a general RNA repair enzyme, which has previously been suggested (Huang *et al.*, 2001). When the percentage editing was taken at each position along the Category 2 ND7 sequences, an apparent reduction in fully edited sequence was observed at each position on knock-down of REL1. This reduction in fully edited sites on REL1 knock-down was more apparent across RPS12, where RNA editing was severely repressed. This study demonstrates a novel sequencing approach that is successful in capturing the true 5' ends of RNA substrates, and in doing so, emphasise the types of editing event that REL1 is responsible for.

Finally, we were able to assess the essentiality of REL2 in relation to REL1 and KREPA3 through the detection of selection pressures acting upon the proteins. Results from this Chapter emphasize the importance of maintaining the protein sequence and catalytic activity of REL2, through the actions of purifying selection. All three editosome proteins were strongly influenced by purifying selection, this may suggest that after neutral evolution and expansion of the editing process, proteins retained sequences once editing became essential (Lukeš *et al.*, 2009). Results also suggest coevolution has occurred between REL1 and REL2, further emphasising their relative importance in RNA editing, presumably by fulfilling differential roles of ligation.

5.2 Recommendation for further research

Potential plans for future research have already been discussed in the relevant Chapters and are outlined in more detail below. Sections 5.2.1 and 5.2.2 describe methods mentioned in Chapter 2, Sections 5.2.3 and 5.2.4 expand upon ideas discussed in Chapter 3, and Section 5.2.5 discusses the potential creation of a REL2 null mutant, the current attempts of which are discussed in Chapter 4.

5.2.1 Addition of catalytically dead REL1 to cKO REL1/TAP cells

It has been previously reported in the literature that the absence of REL1 causes instability of the editosome in PCFs (Huang *et al.*, 2001; Cruz-Reyes *et al.*, 2002). The results described here also support a role of REL1 in 20S editosome stabilisation, and suggest that the presence of either a full length REL1, or chimeric REL2CD:1ID chimera at REL1s original site in the deletion subcomplex is enough to protect editosome structure.

Prior experimental design for introducing an over-expressed and catalytically dead REL1 ligase has involved site directed mutagenesis of REL1 to introduce a lysine to arginine mutation within a KXXY motif (Huang *et al.*, 2001). This mutated ligase could be then introduced into BSFs using a plasmid, such as a newly constructed pHD309-BSD (adapted from pHD309-HYG-PUR, http://tryps.rockefeller.edu/trypsru2_plasmids.html) allowing the mutated REL1 to be expressed in cKO cell lines that also possessed REL2ID containing tagged ligases. This should alleviate any potential instability within the editosome caused by REL1s

loss from the deletion subcomplex, however, it is not known if this approach will result in a dominant-negative effect until it is attempted.

5.2.2 Further quantification of TAP tagged ligase activity through auto-adenylylation and *in vitro* U-insertion and U-deletion assays

Once experimental conditions have been established wherein comparable amount of TAP tagged ligases can be isolated and subjected to *in* vitro assays, the ligases could be subjected to full round activity or autoadenylylation assays. Although a FRET based, full round activity assay provides a means to specifically monitor rREL1 activity *in vitro* (Hall and Schnaufer, in preparation) it cannot distinguish the activities of REL1 and REL2 within an isolated editosome. This is yet another reason to knock-out REL2 in REL cKO cell lines. Auto-adenylylation, either after miniTAP purification or subcomplex isolation from glycerol gradient sedimentation, can be used to distinguish the two relative activities of REL1 and REL2, which can be quantitated. If U-insertion and U-deletion assays, using pre-cleaved mRNA substrates (synthesised *in vitro* using T7 RNA polymerase), were performed at various ATP concentrations, then quantification would reveal whether the tagged ligases were acting as expected, since REL1 and REL2 have different affinities for ATP (see Cruz-Reyes *et al.*, 2002 and Igo Jr. *et al.*, 2002). It would be important to determine if the REL2CD/1ID-TAP ligase behaved catalytically as REL2 in the deletion subcomplex, and REL1CD/2ID-TAP chimeras behaved catalytically as REL1 when in the addition subcomplex.

5.2.3 RNA Repair Assays

There are many proteins involved in RNA repair in different organisms (Amitsur *et al.*, 1987; Martins and Shuman, 2004; Martins and Shuman, 2005; Nandakumar *et al.*, 2008; Tanaka and Shuman, 2011). One of the best studied RNA repair systems is that of T4 bacteriophage, whereby the polynucleotide kinase-phosphatase (PnKp)/Hen1 heterodimer and T4Rnl1 work to repair RNA fragments cleaved by ribotoxins (Amitsur *et al.*, 1987; Wang *et al.*, 2012). This system of repair has been used to create assays specific to tRNA repair *in vitro* (Zhang *et al.*, 2012). Building on the methods mentioned in Section 5.2.2, *in vitro* editing assays could be modified to ascertain the ability of REL1 in RNA repair. These assays would be performed on mitochondrial extracts and use of synthesised substrates that mimic the products of mismediting cleavage directed by endonucleases, which could be prepared by T7 transcription from PCR templates (Igo Jr. *et al.*, 2002; Cruz-Reyes *et al.*, 2002). These RNA fragments could consist of unexpected cleavage events at both editing and non-editing sites.

5.2.4 Application of RNA sequencing strategy

Since meaningful sequences were obtained through capture of 5' monophosphates born of endonucleolytic cleavage, it is possible to expand this study to encompass different transcripts and RNA examples. So far only ND7 and RPS12 REL1 substrates were successfully captured and subjected to deep sequencing, but there remain another ten transcripts to be studied, which are edited in BSFs. Since the RTs for ND7 and RPS12 were generated using the same tagged hexamer primer, the generation of RT-PCRs for deep sequencing will be more streamlined. The tag was originally used to aid enrichment of 5' linked transcripts after the initial RT stage to allow the global sequencing of mitochondrial transcripts. This approach could be pursued, and PCRs could be optimised, which have been designed with adapters for direct attachment to the Illumina sequencing flow cell.

Another possibility would be to extract RNA for sequencing from REL2 RNAi, or null mutant cell lines (see Section 5.2.5 below). This would also reveal the substrates of REL2, through comparison of samples prepared from cells expressing normal and reduced levels of REL2, or none at all. From the results in Chapter 3 it would be expected that the absence of REL2 would create fragments indicative of expected addition editing substrates. If REL2 is functioning *in vivo* as a proof-reading enzyme, as well as in insertion editing as this study and others have suggested, then its ablation may also give rise to stretches of misedited sequences in the presence of REL1.

5.2.5 Creation of a REL2 null mutant

Originally the REL2 null mutant was attempted to aid integration of the REL1CD/2ID tagged chimeric ligases into native editosomes. This experiment was also attractive because the RNAi lines available do not completely knock-out the protein, and a small amount of REL2 may be sufficient in the cell to maintain function in RNA editing (Huang *et al.*, 2001; Schnaufer *et al.*, 2001; Drozdz *et al.*, 2002; Gao and Simpson, 2003; O'Hearn *et al.*, 2003). The creation of a null mutant was not possible in the time frame of this study; however. Both plasmid and PCR based disruption, the latter of which was also adopted by Maciej Drozdz (personal communication) were attempted, with no success. Whilst the PCR based approach tends to produce varying transfection efficiencies, it is surprising that a single allele could not be effectively replaced by a drug resistance marker (Oberholzer *et al.*, 2006). The best future approach would be to construct a plasmid-based REL2 null in wt 427 and cKO REL1 cells using robust drug selection markers, such as puromycin and blasticidin.

5.3 Outlook

The findings described here provide significant and novel insight into the roles of the RNA editing ligases and, in addition, have elucidated a broad range of potential future studies within the field of RNA editing biology. One important observation of this study is the apparent cross talk between the CD and ID specifically of REL1 and REL2, which led to the lower activity of the chimeric ligases. This has not been well characterised, and may, if pursued, give further insight into transcript specific activities of REL1 and REL2. The results of this research has also elucidated a novel role for REL1 in RNA editing, providing insight into its essentiality to the cell, and has provided an approach to sequence the substrates of the final step of RNA editing, which in turn can discern the roles of REL1 and REL2 biochemically.

This future work is of great importance to trypanosome biology, since REL1 has been revealed as a potential drug target, and so it is of importance to clarify the down-stream effects of its ablation and to discriminate the function of REL2 to the cell.

Reference list

Abraham, J.M., Feagin, J.E., and Stuart, K. (1988) Characterization of cytochrome c oxidase III transcripts that are edited only in the 3' region. *Cell* **55**: 267–272.

Acestor, N., Panigrahi, A.K., Carnes, J., Zikova, A., and Stuart, K.D. (2009) The MRB1 complex functions in kinetoplastid RNA processing. *RNA* **15**: 277–286.

Alatortsev, V.S., Cruz-Reyes, J., Zhelonkina, A.G., and Sollner-Webb, B. (2008) *Trypanosoma brucei* RNA editing: coupled cycles of U deletion reveal processive activity of the editing complex. *Molecular and Cellular biology* **28**: 2437–2445.

Alibu, V.P., Storm, L., Haile, S., Clayton, C., and Horn, D. (2004) A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **139**: 75–82.

Amaro, R.E., Baron, R., and McCammon, J.A. (2008a) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design* **22**: 693–705.

Amaro, R.E., Schnaufer, A., Interthal, H., Hol, W., Stuart, K.D., and McCammon, J.A. (2008b) Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 17278–17283.

Amitsur, M., Levitz, R., and Kaufmann, G. (1987) Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *The EMBO Journal* **6**: 2499–2503.

Ammerman, M.L., Downey, K.M., Hashimi, H., Fisk, J.C., Tomasello, D.L., Faktorová, D., *et al.* (2012) Architecture of the trypanosome RNA editing accessory complex, MRB1. *Nucleic Acids Research* **40**: 5637–5650.

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

Ammerman, M.L., Presnyak, V., Fisk, J.C., Foda, B.M., and Read, L.K. (2010) TbRGG2 facilitates kinetoplastid RNA editing initiation and progression past intrinsic pause sites. *RNA* **16**: 2239–2251.

Aphasizhev, R., and Aphasizheva, I. (2011) Mitochondrial RNA processing in trypanosomes. *Research in Microbiology* **162**: 655–663.

Aphasizhev, R., Aphasizheva, I., and Simpson, L. (2003) A tale of two TUTases. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 10617–10622.

Aphasizhev, R., Sbicego, S., Peris, M., Jang, S.H., Aphasizheva, I., Simpson, A.M., *et al.* (2002) Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase):

the key enzyme in U-insertion/deletion RNA editing. *Cell* **108**: 637–648.

Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. *Molecular Cell* **42**: 106–117.

Aphasizheva, I., Ringpis, G.E., Weng, J., Gershon, P.D., Lathrop, R.H., and Aphasizhev, R. (2009) Novel TUTase associates with an editosome-like complex in mitochondria of *Trypanosoma brucei*. *RNA* **15**: 1322–1337.

Babbarwal, V.K., Fleck, M., Ernst, N.L., Schnaufer, A., and Stuart, K. (2007) An essential role of KREPB4 in RNA editing and structural integrity of the editosome in *Trypanosoma brucei*. *RNA* **13**: 737–744.

Baker, N., Koning, H.P. de, Mäser, P., and Horn, D. (2013) Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. *Trends in Parasitology* **29**: 110–118.

Balasegaram, M., Harris, S., Checchi, F., Ghorashian, S., Hamel, C., and Karunakara, U. (2006) Melarsoprol versus eflornithine for treating late-stage Gambian trypanosomiasis in the Republic of the Congo. *Bulletin of the World Health Organization* **84**: 783–791.

Barbet, A.F., and Kamper, S.M. (1993) The importance of mosaic genes to trypanosome survival. *Parasitology Today* **9**: 63–66.

Barrett, M.P., Burchmore, R.J., Stich, A., Lazzari, J.O., Frasch, A.C., Cazzulo, J.J., and Krishna, S. (2003) The trypanosomiases. *Lancet* **362**: 1469–1480.

Barry, J.D., and McCulloch, R. (2001) Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite. *Advances in Parasitology* **49**: 1–70.

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E., and Rudenko, G. (2004) Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome research* **14**: 2319–29.

Benne, R., den, B.J. van, Brakenhoff, J.P., Sloof, P., Boom, J.H. Van, and Tromp, M.C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**: 819–826.

Bernards, A., Ploeg, L.H.T. Van der, Frasch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S., and Cross, G.A.M. (1981) Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3' end. *Cell* **27**: 497–505.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D., *et al.* (2005) The genome of the African trypanosome, *Trypanosoma brucei*. *Science* **309** : 416–422.

Besteiro, S., Barrett, M.P., Riviere, L., and Bringaud, F. (2005) Energy generation in insect stages of *Trypanosoma brucei*: metabolism in flux. *Trends in Parasitology* **21**: 185–191.

Bhat, G.J., Koslowsky, D.J., Feagin, J.E., Smiley, B.L., and Stuart, K. (1990) An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. *Cell* **61**: 885–894.

Bhat, G.J., Souza, A.E., Feagin, J.E., and Stuart, K. (1992) Transcript-specific developmental regulation of polyadenylation in *Trypanosoma brucei* mitochondria. *Molecular and Biochemical Parasitology* **52**: 231–240.

Birkenmeyer, L., Sugisaki, H., and Ray, D.S. (1987) Structural characterization of site-specific discontinuities associated with replication origins of minicircle DNA from *Crithidia fasciculata. Journal of Biological Chemistry* **262**: 2384–2392.

Blanc, V., Alfonzo, J.D., Aphasizhev, R., and Simpson, L. (1999) The mitochondrial RNA ligase from *Leishmania tarentolae* can join RNA molecules bridged by a complementary RNA. *Journal of Biological Chemistry* **274**: 24289–24296.

Blum, B., Bakalara, N., and Simpson, L. (1990) A model for RNA editing in kinetoplastid mitochondria: “guide” RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**: 189–198.

Blum, B., and Simpson, L. (1992) Formation of guide RNA/messenger RNA chimeric molecules in vitro, the initial step of RNA editing, is dependent on an anchor sequence. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 11944–11948.

Blum, B., Sturm, N.R., Simpson, A.M., and Simpson, L. (1991) Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. *Cell* **65**: 543–550.

Boothroyd, C.E., Dreesen, O., Leonova, T., Ly, K.I., Figueiredo, L.M., Cross, G.A.M., and Papavasiliou, F.N. (2009) A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature* **459**: 278–81.

Borst, P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. *Annual review of biochemistry* **55**: 701–32.

Borst, P. (1991) Why kinetoplast DNA works? *Trends in Genetics* **7(5)**: 139–141.

Borst, P., Fase-Fowler, F., and Gibson, W.C. (1987) Kinetoplast DNA of *Trypanosoma evansi*. *Molecular and Biochemical Parasitology* **23**: 31–38.

Brecht, M., Niemann, M., Schlüter, E., Müller, U.F., Stuart, K., and Göringer, H.U. (2005) TbMP42, a protein component of the RNA editing complex in African trypanosomes has endo-exoribonuclease activity. *Molecular Cell* **17**: 621–630.

Bruderer, T., Tu, L.-C., and Lee, M.G.-S. (2003) The 5′ end structure of transcripts derived from the rRNA gene and the RNA polymerase I transcribed protein coding genes in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **129**: 69–77.

Bruhn, D.F., Mozeleski, B., Falkin, L., and Klingbeil, M.M. (2010) Mitochondrial DNA polymerase POLIB is essential for minicircle DNA replication in African trypanosomes. *Molecular Microbiology* **75**: 1414–1425.

Brun, R., Hecker, H., and Lun, Z.R. (1998) *Trypanosoma evansi* and *T. equiperdum*: distribution, biology, treatment and phylogenetic relationship (a review). *Veterinary Parasitology* **79**: 95–107.

Burkard, G., Fragoso, C.M., and Roditi, I. (2007) Highly efficient stable transformation of blood stream fors of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **153**: 220–223.

Byrne, E.M., Connell, G.J., and Simpson, L. (1996) Guide RNA-directed uridine insertion RNA editing in vitro. *The EMBO Journal* **15**: 6758–6765.

Capewell, P., Veitch, N.J., Turner, C.M.R., Raper, J., Berriman, M., Hajduk, S.L., and MacLeod, A. (2011) Differences between *Trypanosoma brucei gambiense* groups 1 and 2 in their resistance to killing by trypanolytic factor 1. *PLoS neglected tropical diseases* **5**: e1287.

Carnes, J., Lewis Ernst, N., Wickham, C., Panicucci, B., and Stuart, K. (2012c) KREX2 is not essential for either procyclic or bloodstream form *Trypanosoma brucei*. *PloS one* **7**: e33405.

Carnes, J., Schnaufer, A., McDermott, S.M., Domingo, G., Proff, R., Steinberg, A.G., *et al.* (2012a) Mutational analysis of *Trypanosoma brucei* editosome proteins KREPB4 and KREPB5 reveals domains critical for function. *RNA (New York, NY)* **18**: 1897–909.

Carnes, J., Schnaufer, A., McDermott, S.M., Domingo, G., Proff, R., Steinberg, A.G., *et al.* (2012b) Mutational analysis of *Trypanosoma brucei* editosome proteins KREPB4 and KREPB5 reveals domains critical for function. *RNA* **18**: 1897–1909.

Carnes, J., and Stuart, K. (2008a) Nucleic Acids and Molecular Biology Volume 20: RNA Editing Working Together: the RNA Editing Machinery in *Trypanosoma brucei*. **20**: 143–164.

Carnes, J., Trotter, J.R., Ernst, N.L., Steinberg, A., and Stuart, K. (2005) An essential RNase III insertion editing endonuclease in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 16614–16619.

Carnes, J., Trotter, J.R., Peltan, A., Fleck, M., and Stuart, K. (2008b) RNA editing in *Trypanosoma brucei* requires three different editosomes. *Molecular and Cellular Biology* **28**: 122–130.

Carter, N., Berger, B.J., and Fairlamb, A.H. (1995) Uptake of Diamidine Drugs by the P2 Nucleoside Transporter in Melarsen-sensitive and -resistant *Trypanosoma brucei brucei*. *Journal of Biological Chemistry* **270**: 28153–28157.

Carter, N.S., and Fairlamb, A.H. (1993) Arsenical-resistant trypanosomes lack an unusual adenosine transporter. *Nature* **361**: 173–176.

Cavalier-Smith, T. (1997) Cell and genome coevolution: facultative anaerobiosis, glycosomes and kinetoplastan RNA editing. *Trends in Genetics* **13**: 6–9.

Cech, T.R. (1991) RNA editing: World's smallest introns. *Cell* **64**: 667–669.

Chan, C.M., Zhou, C., and Huang, R.H. (2009) Reconstituting Bacterial RNA Repair and Modification in Vitro. *Science* **326**: 247–247.

Charlesworth, B., and Charlesworth, D. (2010) *Elements of Evolutionary Genetics*. Roberts & Company Publishers.

Chaudhuri, M., and Hill, G.C. (1996) Cloning, sequencing, and functional activity of the *Trypanosoma brucei brucei* alternative oxidase. *Molecular and biochemical parasitology* **83**: 125–9.

Cheng, L.K.L., and Unrau, P.J. (2010) Closing the circle: replicating RNA with RNA. *Cold Spring Harbor perspectives in biology* **2**: a002204.

Claes, F., Buscher, P., Touratier, L., and Goddeeris, B.M. (2005) *Trypanosoma equiperdum*: master of disguise or historical mistake? *Trends in Parasitology* **21**: 316–321.

Clarkson, A.B., Bienen, E.J., Pollakis, G., and Grady, R.W. (1989) Respiration of bloodstream forms of the parasite *Trypanosoma brucei brucei* is dependent on a plant-like alternative oxidase. *The Journal of biological chemistry* **264**: 17770–6.

Clark-Walker, G.D., Hansbro, P.M., Gibson, F., and Chen, X.J. (2000) Mutant residues suppressing ρ0-lethality in *Kluyveromyces lactis* occur at contact sites between subunits of F1-ATPase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1478**: 125–137.

Clement, S.L., Mingler, M.K., and Koslowsky, D.J. (2004) An intragenic guide RNA location suggests a complex mechanism for mitochondrial gene expression in *Trypanosoma brucei*. *Eukaryotic cell* **3**: 862–9.

Corell, R.A., Feagin, J.E., Riley, G.R., Strickland, T., Guderian, J.A., Myler, P.J., and Stuart, K. (1993) *Trypanosoma brucei* minicircles encode multiple guide RNAs

which can direct editing of extensively overlapping sequences. *Nucleic Acids Research* **21**: 4313–4320.

Corell, R.A., Read, L.K., Riley, G.R., Nellissery, J.K., Allen, T.E., Kable, M.L., *et al.* (1996) Complexes from *Trypanosoma brucei* that exhibit deletion editing and other editing-associated properties. *Molecular and Cellular Biology* **16**: 1410–1418.

Courtin, F., Sidibé, I., Rouamba, J., Jamonneau, V., Gouro, A., and Solano, P. (2009) Population growth and global warming: impacts on tsetse and trypanosomoses in West Africa. *Parasite* **16**: 3–10.

Coustou, V., Besteiro, S., Biran, M., Diolez, P., Bouchaud, V., Voisin, P., *et al.* (2003) ATP generation in the *Trypanosoma brucei* procyclic form: cytosolic substrate level is essential, but not oxidative phosphorylation. *Journal of Biological Chemistry* **278**: 49625–49635.

Coustou, V., Biran, M., Breton, M., Guegan, F., Riviere, L., Plazolles, N., *et al.* (2008) Glucose-induced remodeling of intermediary and energy metabolism in procyclic *Trypanosoma brucei*. *Journal of Biological Chemistry* **283**: 16342–16354.

Covello, P.S., and Gray, M.W. (1993) On the evolution of RNA editing. *Trends in Genetics* **9**: 265–268.

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

Cross, G.A.M., Klein, R.A., and Linstead, D.J. (1975) Utilization of amino acids by *Trypanosoma brucei* in culture: L-threonine as a precursor for acetate. *Parasitology* **71**: 311.

Cruz-Reyes, J., and Sollner-Webb, B. (1996) Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 8901–8906.

Cruz-Reyes, J., Zhelonkina, A.G., Huang, C.E., and Sollner-Webb, B. (2002) Distinct functions of two RNA ligases in active *Trypanosoma brucei* RNA editing complexes. *Molecular and Cellular Biology* **22**: 4652–4660.

Damper, D., and Patton, C. (1976) Pentamidine Transport and Sensitivity in brucei-Group Trypanosomes. *The Journal of Eukaryotic Microbiology* **23**: 349–356.

Dean, S., Gould, M., Dewar, C. and Schnaufer, A. (2013) Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes. *Proceedings of the National Academy of Science* **110**: 14741-14746.

Decker, C.J., and Sollner-Webb, B. (1990) RNA editing involves indiscriminate U changes throughout precisely defined editing domains. *Cell* **61**: 1001–1011.

Deng, J., Schnaufer, A., Salavati, R., Stuart, K.D., and Hol, W.G. (2004) High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. *Journal of Molecular Biology* **343**: 601–613.

Deschamps, P., Lara, E., Marande, W., López-García, P., Ekelund, F., and Moreira, D. (2011) Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids. *Molecular Biology and Evolution* **28**: 53–58.

Domingo, G.J., Palazzo, S.S., Wang, B., Panicucci, B., Salavati, R., and Stuart, K.D. (2003) Dyskinetoplastic *Trypanosoma brucei* contain functional editing complexes. *Eukaryotic Cell* **2**: 569–577.

Drew, M.E., and Englund, P.T. (2001) Intramitochondrial location and dynamics of *Crithidia fasciculata* kinetoplast minicircle replication intermediates. *The Journal of Cell Biology* **153**: 735–44.

Drozdz, M., Palazzo, S.S., Salavati, R., O'Rear, J., Clayton, C., and Stuart, K. (2002) TbMP81 is required for RNA editing in *Trypanosoma brucei*. *The EMBO Journal* **21**: 1791–1799.

Durieux, P.O., Schutz, P., Brun, R., and Kohler, P. (1991) Alterations in Krebs cycle enzyme activities and carbohydrate catabolism in two strains of *Trypanosoma brucei* during in vitro differentiation of their bloodstream to procyclic stages. *Molecular and Biochemical Parasitology* **45**: 19–27.

Durrant, J.D., Hall, L., Swift, R. V, Landon, M., Schnaufer, A., and Amaro, R.E. (2010) Novel Naphthalene-Based Inhibitors of *Trypanosoma brucei* RNA Editing Ligase 1. *PLoS neglected tropical diseases* **4**: e803.

Englund, P.T., and Marini, J.C. (1980) The replication of kinetoplast DNA. *American Journal of Tropical Medicine and Hygiene, The* **29**: 1064–1069.

Ernst, N.L., Panicucci, B., Carnes, J., and Stuart, K. (2009) Differential functions of two editosome exoUases in *Trypanosoma brucei*. *RNA* **15**: 947–957.

Ernst, N.L., Panicucci, B., Igo Jr., R.P., Panigrahi, A.K., Salavati, R., and Stuart, K. (2003) TbMP57 is a 3' terminal uridylyl transferase (TUTase) of the *Trypanosoma brucei* editosome. *Molecular Cell* **11**: 1525–1536.

Estévez, A.M., and Simpson, L. (1999) Uridine insertion/deletion RNA editing in trypanosome mitochondria-a review. *Gene* **240**: 247–260.

Etheridge, R.D., Aphasizheva, I., Gershon, P.D., and Aphasizhev, R. (2008) 3' adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. *The EMBO Journal* **27**: 1596–1608.

Feagin, J.E., Abraham, J.M., and Stuart, K. (1988) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* **53**: 413–422.

Feagin, J.E., Jasmer, D.P., and Stuart, K. (1985) Apocytochrome b and other mitochondrial DNA sequences are differentially expressed during the life cycle of *Trypanosoma brucei*. *Nucleic Acids Research* **13**: 4577–4596.

Feagin, J.E., Jasmer, D.P., and Stuart, K. (1986) Differential mitochondrial gene expression between slender and stumpy bloodforms of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **20**: 207–214.

Feagin, J.E., Shaw, J.M., Simpson, L., and Stuart, K.D. (1987) Addition of nucleotides within cytochrome b transcripts of kinetoplastids. *Molecular Biology of Mitochondria and Chloroplasts Meeting at Cold Springs Harbor Laboratory* 113.

Feagin, J.E., Shaw, J.M., Simpson, L., and Stuart, K.D. (1988) Transcript alteration by mRNA editing in kinetoplastid mitochondria. *Journal of Cellular Biochemistry* **12D**: 19.

Feagin, J.E., and Stuart, K. (1985) Differential expression of mitochondrial genes between life cycle stages of *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 3380–3384.

Felsenstein, J. (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**: 783–791.

Fenn, K., and Matthews, K.R. (2007) The cell biology of *Trypanosoma brucei* differentiation. *Current Opinion in Microbiology* **10**: 539–546.

Fernandes, A.P., Nelson, K., and Beverley, S.M. (1993) Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: Perspectives on the age and origins of parasitism. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 11608–11612.

Fèvre, E.., Coleman, P., Oditt, M., Magona, J., Welburn, S., and woolhouse, M. (2001) The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *The Lancet* **358**: 625–628.

Fèvre, E.., Coleman, P.., Welburn, S.C., and Maudlin, I. (2004) Reanalyzing the 1900–1920 Sleeping Sickness Epidemic in Uganda. *Emerging Infectious Diseases* **10**: 567–573.

Fisk, J.C., Ammerman, M.L., Presnyak, V., and Read, L.K. (2008) TbRGG2, an essential RNA editing accessory factor in two *Trypanosoma brucei* life cycle stages. *Journal of Biological Chemistry* **283**: 23016–23025.

Fisk, J.C., Presnyak, V., Ammerman, M.L., and Read, L.K. (2009) Distinct and Overlapping Functions of MRP1/2 and RBP16 in Mitochondrial RNA Metabolism. *Molecular and Cellular Biology* **29**: 5214–5225.

Foda, B.M., Downey, K.M., Fisk, J.C., and Read, L.K. (2012) Multifunctional G-rich

and RRM-containing domains of TbRGG2 perform separate yet essential functions in trypanosome RNA editing. *Eukaryotic cell* **11**: 1119–31.

Ford, L.B. (2007) Civil conflict and sleeping sickness in Africa in general and Uganda in particular. *Conflict and Health* **1**.

Frommel, T.O., and Balber, A.E. (1987) Flow cytofluorimetric analysis of drug accumulation by multidrug- resistant *Trypanosoma brucei brucei* and T. b rhodesiense. *Molecular and Biochemical Parasitology* **26**: 183–191.

Gao, G., Rogers, K., Li, F., Guo, Q., Osato, D., Zhou, S.X., *et al.* (2010) Uridine Insertion/Deletion RNA Editing in Trypanosomatids: Specific Stimulation in vitro of *Leishmania tarentolae* REL1 RNA Ligase Activity by the MP63 Zinc Finger Protein. *Protist* **161**: 489–496.

Gao, G., Simpson, A.M., Kang, X., Rogers, K., Nebohacova, M., Li, F., and Simpson, L. (2005) Functional complementation of *Trypanosoma brucei* RNA in vitro editing with recombinant RNA ligase. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 4712–4717.

Gao, G., and Simpson, L. (2003) Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? *Journal of Biological Chemistry* **278**: 27570–27574.

Gaud, A., Carrington, M., Deshusses, J., and Schaller, D. (1997) Polymerase chain reaction-based gene disruption in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* 113–115.

Genovese, G., Friedman, D.J., and Pollak, M.R. (2013) APOL1 variants and kidney disease in people of recent African ancestry. *Nature reviews Nephrology* **9**: 240–4.

Gibson, W. (2009) Species-specific probes for the identification of the African tsetse-transmitted trypanosomes. *Parasitology* **136**: 1501–7.

Golas, M.M., Bohm, C., Sander, B., Effenberger, K., Brecht, M., Stark, H., and Goringer, H.U. (2009) Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages in vivo. *The EMBO Journal* **28**: 766–778.

Golden, D.E., and Hajduk, S.L. (2005) The 3'-untranslated region of cytochrome oxidase II mRNA functions in RNA editing of African trypanosomes exclusively as a cis guide RNA. *RNA* **11**: 29–37.

Golden, D.E., and Hajduk, S.L. (2006) The importance of RNA structure in RNA editing and a potential proofreading mechanism for correct guide RNA: pre-mRNA binary complex formation. *Journal of Molecular Biology* **359**: 585–96.

Göringer, H.U. (2012) "Gestalt," Composition and Function of the *Trypanosoma brucei* Editosome. *Annual Review of Microbiology* **66**: 65–82.

Grams, J., McManus, M.T., and Hajduk, S.L. (2000) Processing of polycistronic guide RNAs is associated with RNA editing complexes in *Trypanosoma brucei*. *The EMBO Journal* **19**: 5525–32.

Grams, J., Morris, J. C., Drew, M. E., Wang, Z., Englund, P. T. and hajduk, S. L. (2002) A Trypanosome Mitochondrial RNA Polymerase Is Required for Transcription and Replication. *The Journal of Biological Chemistry* **277**: 16952-16959.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009) Identification of protein binding sites on U3 snoRNA and Pre-rRNA by UV crosslinking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 9613–9618.

Graur, D., and Li, W.-H. (2000) *Fundamentals of molecular evolution (2nd Ed.) GRAUR Dan, LI Wen-Hsiung: Librairie Lavoisier*. .

Gray, M.W. (2012) Evolutionary origin of RNA editing. *Biochemistry* **51**: 5235–42.

Guo, X., Carnes, J., Ernst, N.L., Winkler, M., and Stuart, K. (2012) KREPB6, KREPB7, and KREPB8 are important for editing endonuclease function in Trypanosoma brucei. *RNA (New York, NY)* **18**: 308–20.

Guo, X., Ernst, N.L., and Stuart, K.D. (2008) The KREPA3 zinc finger motifs and OB-fold domain are essential for RNA editing and survival of Trypanosoma brucei. *MolCell Biol* **28**: 6939–6953.

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., *et al.* (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA (New York, NY)* **17**: 1697–712.

Hajduk, S., and Ochsenreiter, T. (2010) RNA editing in kinetoplastids. *RNABiol* **7**: 229–236.

Hajduk, S.L., Siqueira, A.M., and Vickerman, K. (1986) Kinetoplast DNA of *Bodo caudatus*: a Noncatenated Structure. *Molecular and Cellular Biology* **6**: 4372–4378.

Hamilton, P.B., Gibson, W.C., and Stevens, J.R. (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Molecular Phylogenetics and Evolution* **44**: 15–25.

Hancock, K., and Hajduk, S.L. (1990) The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. *Journal of Biological Chemistry* **265**: 19208–19215.

Harris, M.E., and Hajduk, S.L. (1992) Kinetoplastid RNA editing: In vitro formation of cytochrome b gRNA-mRNA chimeras from synthetic substrate RNAs. *Cell* **68**: 1091–1099.

Hashimi, H., Cicova, Z., Novotna, L., Wen, Y.Z., and Lukeš, J. (2009) Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. *RNA* **15**: 588–599.

Hashimi, H., Zikova, A., Panigrahi, A.K., Stuart, K.D., and Lukeš, J. (2008) TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. *RNA* **14**: 970–980.

Hashimi, H., Zimmer, S.L., Ammerman, M.L., Read, L.K., and Lukeš, J. (2013) Dual core processing: MRB1 is an emerging kinetoplast RNA editing complex. *Trends in Parasitology* **29**: 91–99.

Hayman, M.L., and Read, L.K. (1999) *Trypanosoma brucei* RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. *Journal of Biological Chemistry* **274**: 12067–12074.

Hellemond, J.J. Van, Opperdoes, F.R., and Tielens, A.G. (2005) The Extraordinary Mitochondrion and Unusual Citric Acid Cycle in *Trypanosoma brucei*. *Biochemical Society Transactions* **33**: 967–971.

Hernandez, A., Madina, B.R., Ro, K., Wohlschlegel, J.A., Willard, B., Kinter, M.T., and Cruz-Reyes, J. (2010) REH2 RNA helicase in kinetoplastid mitochondria: ribonucleoprotein complexes and essential motifs for unwinding and guide RNA (gRNA) binding. *The Journal of Biological Chemistry* **285**: 1220–1228.

Hertz-Fowler, C., Figueiredo, L.M., Quail, M.A., Becker, M., Jackson, A., Bason, N., *et al.* (2008) Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PloS one* **3**: e3527.

Ho, C.K., and Shuman, S. (2002) Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 12709–12714.

Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S. (2004) Structure and mechanism of RNA ligase. *Structure* **12**: 327–339.

Hoare, C.A. (1937) Morphological and taxonomic studies on mammalian trypanosomes. III. Spontaneous occurrence of strains of *Trypanosoma evansi* devoid of the kinetonucleus. *Parasitology* **29**.

Hong, M., and Simpson, L. (2003) Genomic Organization of Kinetoplast DNA Minicircles. *Protist* **154**: 265–279.

Horn, D., and Barry, J.D. (2005) The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Research* **13**: 525–33.

Horn, D., and McCulloch, R. (2010) Molecular mechanisms underlying the control

of antigenic variation in African trypanosomes. *Current Opinion in Microbiology* **13**: 700–5.

Huang, C.E., Cruz-Reyes, J., Zhelonkina, A.G., O'Hearn, S., Wirtz, E., and Sollner-Webb, B. (2001) Roles for ligases in the RNA editing complex of *Trypanosoma brucei*: band IV is needed for U-deletion and RNA repair. *The EMBO Journal* **20**: 4694–4703.

Huang, C.E., O'Hearn, S.F., and Sollner-Webb, B. (2002) Assembly and function of the RNA editing complex in *Trypanosoma brucei* requires band III protein. *Molecular and Cellular Biology* **22**: 3194–3203.

Huber, L.A. (2003) Is proteomics heading in the wrong direction? *Nature reviews Molecular cell biology* **4**: 74–80.

Hughes, A.L., and Piontkivska, H. (2003) Phylogeny of trypanosomatidae and bodonidae (kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. *Molecular Biology and Evolution* **20**: 644–652.

Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics* **18**: 486–487.

Igo Jr., R.P., Data, !Lost, Ernst, N.L., Panigrahi, A.K., Salavati, R., and Stuart, K. (2002) Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing . *Eukaryotic Cell* **1**: 112–118.

Igo, R.P., Palazzo, S.S., Burgess, M.L., Panigrahi, A.K., and Stuart, K. (2000) Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing. *Molecular and Cellular Biology* **20**: 8447–57.

Jamonneau, V., Ilboudo, H., Kaboré, J., Kaba, D., Koffi, M., Solano, P., *et al.* (2012) Untreated human infections by *Trypanosoma brucei gambiense* are not 100% fatal. *PLoS neglected tropical diseases* **6**: e1691.

Jasmer, D.P., Feagin, J.E., and Stuart, K. (1985) Diverse patterns of expression of the cytochrome c oxidase subunit I gene and unassigned reading frames 4 and 5 during the life cycle of *Trypanosoma brucei*. *Molecular and Cellular Biology* **5**: 3041–3047.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011) Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research* **39**: e141.

Jensen, R.E., and Englund, P.T. (2012) Network news: the replication of kinetoplast DNA. *Annual Review of Microbiology* **66**: 473–91.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987) Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. *Cell 51* 273–281.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**: 275–282.

Jordan, A.M. (1986) *Trypanosomiasis control and African rural development.* Longman Group Ltd., .

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., and Koonin, E. V. (2002) Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. *Genome Research* **12**: 962–968.

K.R., M., Sherwin, T., and Gull, K. (1995) Mitochondrial Genome Repositioning During the Differentiation of the African Trypanosome Between Life Cycles is Microtubule Mediated. *Journal of Cell Science* **108**: 2231–2239.

Kable, M.L., Seiwert, S.D., Heidmann, S., and Stuart, K. (1996) RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* **273**: 1189–1195.

Kang, X., Rogers, K., Gao, G., Falick, A.M., Zhou, S., and Simpson, L. (2005) Reconstitution of uridine-deletion precleaved RNA editing with two recombinant enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 1017–1022.

Kieft, R., Capewell, P., Turner, C.M.R., Veitch, N.J., MacLeod, A., and Hajduk, S. (2010) Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 16137–41.

Koning, H.P. de, Watson, C.J., Sutcliffe, L., and Jarvis, S.M. (2000) Differential regulation of nucleoside and nucleobase transporters in Crithidia fasciculata and *Trypanosoma brucei brucei*. *Molecular and Biochemical Parasitology* **106**: 93–107.

Koslowsky, D.J., Bhat, G.J., Perrollaz, A.L., Feagin, J.E., and Stuart, K. (1990) The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. *Cell* **62**: 901–911.

Koslowsky, D.J., Bhat, G.J., Read, L.K., and Stuart, K. (1991) Cycles of progressive realignment of gRNA with mRNA in RNA editing. *Cell* **67**: 537–546.

Koslowsky, D.J., Göringer, H.U., Morales, T.H., and Stuart, K. (1992b) In vitro guide RNA/mRNA chimaera formation in *Trypanosoma brucei* RNA editing. *Nature* **356**: 807–809.

Koslowsky, D.J., Riley, G.R., Feagin, J.E., and Stuart, K. (1992a) Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of *Trypanosoma brucei*. *Molecular and Cellular Biology* **12**: 2043–2049.

Lai, D.H., Hashimi, H., Lun, Z.R., Ayala, F.J., and Lukeš, J. (2008) Adaptations of

Trypanosoma brucei to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of T. brucei. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 1999–2004.

Landweber, L.F., Fiks, A.G., and Gilbert, W. (1993) The boundaries of partially edited transcripts are not conserved in kinetoplastids: Implications for the guide RNA model of editing. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 9242–9246.

Landweber, L.F., and Gilbert, W. (1994) Phylogenetic analysis of RNA editing: A primitive genetic phenomenon. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 918–921.

Law, J.A., O'Hearn, S., and Sollner-Webb, B. (2007) In *Trypanosoma brucei* RNA editing, TbMP18 (band VII) is critical for editosome integrity and for both insertional and deletional cleavages. *Molecular and Cellular Biology* **27**: 777–787.

Law, J.A., O'Hearn, S.F., and Sollner-Webb, B. (2008) *Trypanosoma brucei* RNA editing protein TbMP42 (band VI) is crucial for the endonucleolytic cleavages but not the subsequent steps of U-deletion and U-insertion. *RNA* **14**: 1187–1200.

Lee, S.H., Stephens, J.L., Paul, K.S., and Englund, P.T. (2006) fatty acid synthesis by elongases in trypanosomes. *Cell* **126**: 691–699.

Legros, D., Ollivier, G., Gastellu-Etchegorry, M., Paquet, C., Burri, C., Jannin, J., and Büscher, P. (2002) Treatment of human African trypanosomiasis - present situation and needs for research and development. *Lancet Infectious Diseases* **2**: 437–440.

Lerch, M., Carnes, J., Acestor, N., Guo, X., Schnaufer, A., and Stuart, K. (2012) Editosome accessory factors KREPB9 and KREPB10 in *Trypanosoma brucei*. *Eukaryotic cell* **11**: 832–43.

Li, F., Herrera, J., Zhou, S., Maslov, D.A., and Simpson, L. (2011) Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3542–7.

Li, Z., Lindsay, M.E., Motyka, S.A., Englund, P.T., and Wang, C.C. (2008) Identification of a bacterial-like HslVU protease in the mitochondria of Trypanosoma brucei and its role in mitochondrial DNA replication. *PLoS pathogens* **4**: e1000048.

Liu, B., Molina, H., Kalume, D., Pandey, A., Griffith, J.D., and Englund, P.T. (2006) Role of p38 in replication of *Trypanosoma brucei* kinetoplast DNA. *Molecular and Cellular Biology* **26**: 5382–5393.

Liu, B., Wang, J., Yaffe, N., Lindsay, M.E., Zhao, Z., Zick, A., *et al.* (2009)

Trypanosomes have six mitochondrial DNA helicases with one controlling kinetoplast maxicircle replication. *Molecular Cell* **35**: 490–501.

Liu, B., Yildirir, G., Wang, J., Tolun, G., Griffith, J.D., and Englund, P.T. (2010) TbPIF1, a *Trypanosoma brucei* mitochondrial DNA helicase, is essential for kinetoplast minicircle replication. *Journal of Biological Chemistry* **285**: 7056–7066.

Liu, Y., and Englund, P.T. (2007) The rotational dynamics of kinetoplast DNA replication. *Molecular Microbiology* **64**: 676–690.

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

Lukeš, J., Arts, G.J., Burg, J. Van den, Haan, A. De, Opperdoes, F., Sloof, P., and Benne, R. (1994) Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. *The EMBO Journal* **13**: 5086–5098.

Lukeš, J., Hashimi, H., and Zikova, A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Current Genetics* **48**: 277–299.

Lukeš, J., Leander, B.S., and Keeling, P.J. (2009) Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 9963–9970.

Lythgoe, K.A., Morrison, L.J., Read, A.F., and Barry, J.D. (2007) Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 8095–100.

MacGregor, P., Szöőr, B., Savill, N.J., and Matthews, K.R. (2012) Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nature Reviews Microbiology* **10**: 431–8.

Madina, B.R., Kuppan, G., Vashisht, A.A., Liang, Y.-H., Downey, K.M., Wohlschlegel, J.A., *et al.* (2011) Guide RNA biogenesis involves a novel RNase III family endoribonuclease in *Trypanosoma brucei*. *RNA* **17**: 1821–1830.

Magez, S., Radwanska, M., Beschin, A., Sekikawa, K., and Baetselier, P. De (1999) Tumor Necrosis Factor Alpha Is a Key Mediator in the Regulation of Experimental *Trypanosoma brucei* Infections. *Infection and Immunity* **67**: 3128–3132.

Magnus, E., Vervoort, T., and Meirvenne, N. Van (1978) A card-agglutination test with stained trypanosomes (c.a.t.t.) for the serological diagnosis of *T.b. gambiense* trypanosomiasis. *Annales de la Societe Belge de Medicin Tropicale* **58**: 169–176.

Malvy, D., and Chappuis, F. (2011) Sleeping sickness. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **17**: 986–95.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martins, A., and Shuman, S. (2004) Characterization of a baculovirus enzyme with RNA ligase, polynucleotide 5'-kinase, and polynucleotide 3'-phosphatase activities. *The Journal of biological chemistry* **279**: 18220–31.

Martins, A., and Shuman, S. (2005) An end-healing enzyme from *Clostridium thermocellum* with 5' kinase, 2',3' phosphatase, and adenylyltransferase activities. *RNA* **11**: 1271–80.

Mäser, P., Lüscher, A., and Kaminsky, R. (2003) Drug transport and drug resistance in African trypanosomes. *Drug Resistance Updates* **6**: 281–290.

Maslov, D.A., Avila, H.A., Lake, J.A., and Simpson, L. (1994) Evolution of RNA editing in kinetoplastid protozoa. *Nature* **368**: 345–348.

Maslov, D.A., Lukes, J., Jirku, M., and Simpson, L. (1996) Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. *Molecular and Biochemical Parasitology* **75**: 197–205.

Matovu, E., Stewart, M.L., Geiser, F., Brun, R., Maser, P., Wallace, L.J., *et al.* (2003) Mechanisms of arsenical and diamidine uptake and resistance in *Trypanosoma brucei*. *Eukaryotic Cell* **2**: 1003–1008.

Matsumoto, J., Dewar, K., Wasserscheid, J., Wiley, G.B., Macmil, S.L., Roe, B.A., *et al.* (2010) High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: Alternative expression modes and gene function correlates. *Genome Research* **20**: 636–645.

Matthews, K.R., Ellis, J.R., and Paterou, A. (2004) Molecular regulation of the life cycle of African trypanosomes. *Trends in Parasitology* **20**: 40–47.

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

Mayho, M., Fenn, K., Craddy, P., Crosthwaite, S., and Matthews, K. (2006) 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 Acid Research* **34**: 5312–24.

McIntosh, L. (1994) Molecular biology of the alternative oxidase. *Plant physiology*

105: 781–6.

McManus, M.T., Adler, B.K., Pollard, V.W., and Hajduk, S.L. (2000) *Trypanosoma brucei* guide RNA poly(U) tail formation is stabilized by cognate mRNA. *Molecular and Cellular Biology* **20**: 883–891.

McManus, M.T., Shimamura, M., Grams, J., and Hajduk, S.L. (2001) Identification of candidate mitochondrial RNA editing ligases from *Trypanosoma brucei*. *RNA* **7**: 167–175.

Medina, D., Moskowitz, N., Khan, S., Christopher, S., and Germino, J. (2000) Rapid purification of protein complexes from mammalian cells. *Nucleic Acids Research* **28**: E61.

Michels, P.A.M., Bringaud, F., Herman, M., and Hannaert, V. (2006) Metabolic functions of glycosomes in trypanosomatids. *BBA - Molecular Cell Research* **1763**: 1463–1477.

Militello, K.T., and Read, L.K. (1999) Coordination of kRNA editing and polyadenylation in *Tryanosoma brucei* mitochondria: complete editing is not required for long poly(A) tract addition. *Nucleic Acids Research* **27**: 1377–1385.

Milman, N., Motyka, S.A., Englund, P.T., Robinson, D., and Shlomai, J. (2007) Mitochondrial origin-binding protein UMSBP mediates DNA replication and segregation in trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19250–5.

Nandakumar, J., Ho, C.K., Lima, C.D., and Shuman, S. (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *Journal of Biological Chemistry* **279**: 31337–31347.

Nandakumar, J., Schwer, B., Schaffrath, R., and Shuman, S. (2008) RNA repair: an antidote to cytotoxic eukaryal RNA damage. *Molecular cell* **31**: 278–86.

Nandakumar, J., and Shuman, S. (2004) How an RNA ligase discriminates RNA versus DNA damage. *Molecular cell* **16**: 211–21.

Nandakumar, J., and Shuman, S. (2005) Dual mechanisms whereby a broken RNA end assists the catalysis of its repair by T4 RNA ligase 2. *The Journal of biological chemistry* **280**: 23484–9.

Nandakumar, J., Shuman, S., and Lima, C.D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell* **127**: 71–84.

Nebohácová, M., Kim, C.E., Simpson, L., and Maslov, D.A. (2009) RNA editing and mitochondrial activity in promastigotes and amastigotes of *Leishmania donovani*. *International journal for parasitology* **39**: 635–44.

Nielsen, R., and Yang, Z. (1998) Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics* **148**: 929–936.

Niemann, M., Kaibel, H., Schlüter, E., weitzel, K., Brecht, M., and Ulrich Göringer, H. (2009) Kinetoplastid RNA editing involves a 3' nucleotidyl phosphatase activity. *Nucleic Acids Research* **37**: 1897–1906.

Niemann, M., Schneider, A., and Cristodero, M. (2011) Mitochondrial translation in trypanosomatids: a novel target for chemotherapy? *Trends in parasitology* **27**: 429–33.

O'Hearn, S., Huang, C.E., Hemann, M., Zhelonkina, A., and Sollner-Webb, B. (2003) *Trypanosoma brucei* RNA editing complex: band II is structurally critical and maintains band V ligase, which is nonessential. *Molecular and Cellular Biology* **23** : 7909–7919.

Oberholzer, M., Morand, S., Kunz, S., and Seebeck, T. (2006) A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. *Molecular and Biochemical Parasitology* **145**: 177–120.

Ochsenreiter, T., Cipriano, M., and Hajduk, S.L. (2007) KISS: the kinetoplastid RNA editing sequence search tool. *RNA* **13**: 1–4.

Ochsenreiter, T., Cipriano, M., and Hajduk, S.L. (2008) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS one* **3**.

Ochsenreiter, T., and Hajduk, S.L. (2006) Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Reports* **7**: 1128–1133.

Odell, M., Sriskanda, V., Shuman, S., and Nikolov, D.B. (2000) Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Molecular Cell* **6**: 1183–1193.

Oeffinger, M., Wei, K.E., Rogers, R., DeGrasse, J.A., Chait, B.T., Aitchison, J.D., and Rout, M.P. (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. *Nature Methods* **4**: 951–956.

Oli, M.W., Cotlin, L.F., Shiflett, A.M., and Hajduk, S.L. (2006) Serum resistance-associated protein blocks lysosomal targeting of trypanosome lytic factor in *Trypanosoma brucei*. *Eukaryotic Cell* **5**: 132–139.

Opperdoes, F.R. (1987) Compartmentation of carbohydrate metabolism in trypanosomes. *Annual Review of Microbiology* **41**: 127–151.

Palazzo, S.S., Panigrahi, A.K., Igo Jr., R.P., Salavati, R., and Stuart, K. (2003)

Kinetoplastid RNA editing ligases: Complex association, characterization, and substrate requirements . *Molecular and Biochemical Parasitology* **127**: 161–167.

Panigrahi, A.K., Ernst, N.L., Domingo, G.J., Fleck, M., Salavati, R., and Stuart, K.D. (2006) Compositionally and functionally distinct editosomes in *Trypanosoma brucei*. *RNA* **12**: 1038–1049.

Panigrahi, A.K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R.A., Acestor, N., *et al.* (2009) A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. *Proteomics* **9**: 434–450.

Panigrahi, A.K., Schnaufer, A., Carmean, N., Igo Jr., R.P., Gygi, S.P., Ernst, N.L., *et al.* (2001) Four related proteins of the *Trypanosoma brucei* RNA editing complex . *Molecular and Cellular Biology* **21**: 6833–6840.

Panigrahi, A.K., Schnaufer, A., Ernst, N.L., Wang, B., Carmean, N., Salavati, R., and Stuart, K. (2003) Identification of novel components of *Trypanosoma brucei* editosomes . *RNA* **9**: 484–492.

Panigrahi, A.K., Zikova, A., Dalley, R.A., Acestor, N., Ogata, Y., Anupama, A., *et al.* (2008) Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Molecular Cell Proteomics* **7**: 534–545.

Park, Y.-J., Budiarto, T., Wu, M., Pardon, E., Steyaert, J., and Hol, W.G.J. (2012b) The structure of the C-terminal domain of the largest editosome interaction protein and its role in promoting RNA binding by RNA-editing ligase L2. *Nucleic Acids Research* **40**: 6966–6977.

Park, Y.-J., Pardon, E., Wu, M., Steyaert, J., and Hol, W.G.J. (2012a) Crystal structure of a heterodimer of editosome interaction proteins in complex with two copies of a cross-reacting nanobody. *Nucleic acids research* **40**: 1828–40.

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Perez-Morga, D. (2001) The VSG expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Molecular and Biochemical Parasitology* **114**: 1–16.

Pays, E., Vanhollebeke, B., Vanhamme, L., Paturiaux-Hanocq, F., Nolan, D.P., and Pérez-Morga, D. (2006) The trypanolytic factor of human serum. *Nature Reviews: Microbiology* **4**: 477–486.

Pelletier, M., and Read, L.K. (2003) RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in *Trypanosoma brucei*. *RNA* **9**: 457–468.

Pérez-Morga, D., Vanhollebeke, B., Paturiaux-Hanocq, F., Nolan, D.P., Lins, L., Homblé, F., *et al.* (2005) Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* **309**: 469–472.

Piller, K.J., Rusché, L.N., and Sollner-Webb, B. (1996) Trypanosoma brucei RNA editing: a full round of uridylate insertional editing in vitro mediated by endonuclease and RNA ligase. *Journal of Biological Chemistry* **271**: 4613–4619.

Pollard, V.W., Harris, M.E., and Hajduk, S.L. (1992) Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. *The EMBO Journal* **11**: 4429–4438.

Pollard, V.W., Rohrer, S.P., Michelotti, E.F., Hancock, K., and Hajduk, S.L. (1990) Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. *Cell* **63**: 783–790.

Pond, S.L.K., and Frost, S.D.W. (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**: 2531–2533.

Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., *et al.* (2009) Nifurtimox-eflornithine combination therapy for second-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *Lancet* **374**: 56–64.

Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Molecular and Cellular Biology* **27**: 6876–6888.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **13**: 341.

Read, L.K., Corell, R.A., and Stuart, K. (1992a) Chimeric and truncated RNAs in *Trypanosoma brucei* suggest transesterifications at non-consecutive sites during RNA editing. *Nucleic Acids Research* **20**: 2341–2347.

Read, L.K., Myler, P.J., and Stuart, K. (1992b) Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. *Journal of Biological Chemistry* **267**: 1123–1128.

Read, L.K., Stankey, K.A., Fish, W.R., Muthiani, A.M., and Stuart, K. (1994a) Developmental regulation of RNA editing and polyadenylation in four life cycle stages of *Trypanosoma congolense*. *Molecular and Biochemical Parasitology* **68**: 297–306.

Read, L.K., Wilson, K.D., Myler, P.J., and Stuart, K. (1994b) Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein coding sequences. *Nucleic Acids Research* **22**: 1489–1495.

Reifur, L., Yu, L.E., Cruz-Reyes, J., Vanhartesvelt, M., and Koslowsky, D.J. (2010) The impact of mRNA structure on guide RNA targeting in kinetoplastid RNA

editing. *PloS one* **5**: e12235.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology* **17**: 1030–1032.

Ringpis, G.-E., Aphasizheva, I., Wang, X., Huang, L., Lathrop, R.H., Hatfield, G.W., and Aphasizhev, R. (2010) Mechanism of U insertion RNA editing in trypanosome mitochondria: the bimodal TUTase activity of the core complex. *Journal of Molecular Biology* **399**: 680–695.

Riou, G., and Saucier, J. (1979) Characterisation of the molecular components in kinetoplast-mitochondrial DNA of *Trypanosoma equiperdum*. *Journal of Cell Biology* **82**: 248–263.

Roberts, L.S., Jr., and Janovy, J. (2000) *Foundations of Parasitology*. McGraw-Hill Higher Education.

Rogers, D.J., and Randolph, S.E. (1988) Tsetse Flies in Africa: Bane or Boon? *Conservation Biology* **2**: 57–65.

Rogers, K., Gao, G., and Simpson, L. (2007) Uridylate-specific 3’ 5'-exoribonucleases involved in uridylate-deletion RNA editing in trypanosomatid mitochondria. *Journal of Biological Chemistry* **282**: 29073–29080.

Romaniuk, P.J., and Uhlenbeck, O.C. (1983) Recombinant DNA Part B. *Methods in Enzymology* **100**: 52–59.

Roth, C., Jacquemot, C., Giroud, C., Bringaud, F., Eisen, H., and Baltz, T. (1991) Antigenic variation in *Trypanosoma equiperdum*. *Research in Microbiology* **142**: 725–730.

Roy, J., Faktorová, D., Lukeš, J., and Burger, G. (2007) Unusual Mitochondrial Genome Structures throughout the Euglenozoa. *Protist* **158**: 385–396.

Rusché, L.N., Cruz-Reyes, J., Piller, K.J., and Sollner-Webb, B. (1997) Purification of a functional enzymatic editing complex from *Trypanosoma brucei* mitochondria. *The EMBO Journal* **16**: 4069–4081.

Rusché, L.N., Huang, C.E., Piller, K.J., Hemann, M., Wirtz, E., and Sollner-Webb, B. (2001) The Two RNA Ligases of the *Trypanosoma brucei* RNA Editing Complex: Cloning the Essential Band IV Gene and Identifying the Band V Gene. *Molecular and Cellular Biology* **21**: 979–989.

Rusché, L.N., Piller, K.J., and Sollner-Webb, B. (1995) Guide RNA-mRNA chimeras, which are potential RNA editing intermediates, are formed by endonuclease and RNA ligase in a trypanosome mitochondrial extract. *Molecular and Cellular Biology* **15**: 2933–2941.

Ryan, C.M., Militello, K.T., and Read, L.K. (2003) Polyadenylation regulates the stability of *Trypanosoma brucei* mitochondrial RNAs. *Journal of Biological Chemistry* **278**: 32753–32762.

Sabatini, R., and Hajduk, S.L. (1995) RNA ligase and its involvement in guide RNA/mRNA chimera formation. Evidence for a cleavage-ligation mechanism of *Trypanosoma brucei* mRNA editing. *Journal of Biological Chemistry* **270**: 7233–7240.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.

Salavati, R., Ernst, N.L., O'Rear, J., Gilliam, T., Jr, T.S., and Stuart, K. (2006) KREPA4, an RNA binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. *RNA* **12**: 819–831.

Schnaufer, A., Clark-Walker, G.D., Steinberg, A.G., and Stuart, K. (2005) The F1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. *The EMBO Journal* **24**: 4029–4040.

Schnaufer, A., Domingo, G.J., and Stuart, K. (2002) Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *International Journal for Parasitology* **32**: 1071–1084.

Schnaufer, A., Ernst, N., O'Rear, J., Salavati, R., and Stuart, K. (2003) Separate Insertion and Deletion Sub-complexes of the *Trypanosoma brucei* RNA Editing Complex. *Molecular Cell* **12**: 307–319.

Schnaufer, A., Panigrahi, A.K., Panicucci, B., Igo Jr., R.P., Wirtz, E., Salavati, R., and Stuart, K. (2001) An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science* **291**: 2159–2162.

Schnaufer, A., Wu, M., Park, Y.-J., Nakai, T., Deng, J., Proff, R., *et al.* (2010) A protein-protein interaction map of trypanosome ~20S editosomes. *Journal of Biological Chemistry* **285**: 5282–5295.

Schneider, A. (2011) Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annual Review of Biochemistry* **80**: 1033–1053.

Schumacher, M.A., Karamooz, E., Zikova, A., Trantirek, L., and Lukeš, J. (2006) Crystal structures of T. brucei MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* **126**: 701–711.

Seed, J.R. (2001) African trypanosomiasis research: 100 years of progress, but questions and problems still remain. *International Journal for Parasitology* **31**: 434–442.

Seiwert, S., Wachal, M., and Stuart, K.D. (1996) Kinetoplastid RNA editing: direct

visualization of precursor, product, and intermediates suggests a mechanism for in vitro uridine deletion. *Cell* 831–841.

Seiwert, S.D., and Stuart, K. (1994) RNA editing: Transfer of genetic information from gRNA to precursor mRNA in vitro. *Science* **266**: 114–117.

Shaneh, A., and Salavati, R. (2010) Kinetoplastid RNA editing ligases 1 and 2 exhibit different electrostatic properties. *Journal of Molecular Modeling* **16**: 61–76.

Shapiro, T.A. (1993) Kinetoplast DNA maxicircles: Networks within networks. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 7809–7813.

Shapiro, T.A., and Englund, P.T. (1995) The structure and replication of kinetoplast DNA. *Annual Review of Microbiology* **49**: 117–143.

Sharma, R., Gluenz, E., Peacock, L., Gibson, W., Gull, K., and Carrington, M. (2009) The heart of darkness: growth and form of *Trypanosoma brucei* in the tsetse fly. *Trends in parasitology* **25**: 517–24.

Simarro, P.P., Diarra, A., Ruiz Postigo, J.A., Franco, J.R., and Jannin, J.G. (2011) The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000-2009: the way forward. *PLoS neglected tropical diseases* **5**: e1007.

Simpson, L. (1987) The mitochondrial genome of kinetoplastid protozoa: Genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology* **41**: 363–382.

Simpson, L., Neckelmann, N., la Cruz, V.F. de, Simpson, A.M., Feagin, J.E., Jasmer, D.P., and Stuart, K. (1987) Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. *Journal of Biological Chemistry* **262**: 6182–6196.

Simpson, L., and Shaw, J. (1989) RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* **57**: 355–366.

Sollner-Webb, B. (1991) RNA editing. *Current Opinion in Cell Biology* **3**: 1056–1061.

Sollner-Webb, B., Rusché, L.N., and Cruz-Reyes, J. (2001) Ribonucleases - Part A. *Methods in Enzymology* **341**: 154–174.

Songa, E.B., Paindavoine, P., Wittouck, E., Viseshakul, N., Muldermans, S., Steinert, M., and Hamers, R. (1990) Evidence for kinetoplast and nuclear DNA homogeneity in *Trypanosoma evansi* isolates. *Molecular and Biochemical Parasitology* **43**: 167–180.

Souza, A.E., Myler, P.J., and Stuart, K. (1992) Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited, developmentally regulated, and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. *Molecular and Cellular Biology* **12**: 2100–2107.

Speijer, D. (2006) Is kinetoplastid pan-editing the result of an evolutionary balancing act? *IUBMBLife* **58**: 91–96.

Steinert, M., and Assel, S. Van (1980) Sequence heterogeneity in kinetoplast DNA: reassociation kinetics. *Plasmid* **3**: 7–17.

Steverding, D. (2010) The development of drugs for treatment of sleeping sickness: a historical review. *Parasites & Vectors* **3**: 15.

Stuart, K., deVos, T., and Myler, P.J. (1997) Maxicircle replication in *Trypanosoma brucei*. *Acta Parasitologica Turcica* **21**: 52.

Stuart, K., and Feagin, J.E. (1992) Mitochondrial DNA of kinetoplastids. *International Review of Cytology* **141**: 65–88.

Stuart, K., Feagin, J.E., and Abraham, J.M. (1989) RNA editing: the creation of nucleotide sequences in mRNA - A minireview. *Gene* **82**: 155–160.

Stuart, K., and Gelvin, S.R. (1980) Kinetoplast DNA of normal and mutant *Trypanosoma brucei*. *American Journal of Tropical Medicine and Hygiene, The* **29**: 1075–1081.

Stuart, K.D. (1971) Evidence for the retention of kinetoplast DNA in an acriflavin-induced dyskinetoplastic strain of *Trypanosoma brucei* which replicates the altered central element of the kinetoplast. *Journal of Cellular Biochemistry* **49**: 189–195.

Stuart, K.D., Panigrahi, A.K., Schnaufer, A., Drozdz, M., Clayton, C., and Salavati, R. (2002) Composition of the editing complex of *Trypanosoma brucei*. *Philosophical Transactions of the Royal Society of LondonSeries B, Biological Sciences* **357**: 71–79.

Stuart, K.D., Schnaufer, A., Ernst, N.L., and Panigrahi, A.K. (2005) Complex management: RNA editing in trypanosomes. *Trends in Biochemical Sciences* **30**: 97–105.

Sturm, N.R., Maslov, D., Blum, B., and Simpson, L. (1992) Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: misediting produced by misguiding. *Cell* **70**: 469–476.

Sturm, N.R., and Simpson, L. (1990a) Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. *Cell* **61**: 871–878.

Sturm, N.R., and Simpson, L. (1990b) Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell* **61**: 879–884.

Surve, S., Heestand, M., Panicucci, B., Schnaufer, A., and Parsons, M. (2012) Enigmatic presence of mitochondrial complex I in *Trypanosoma brucei* bloodstream forms. *Eukaryotic Cell* **11**: 183–93.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–2739.

Tanaka, N., and Shuman, S. (2011) RtcB is the RNA ligase component of an Escherichia coli RNA repair operon. *The Journal of Biological Chemistry* **286**: 7727–7731.

Tarun Jr., S.Z., Schnaufer, A., Ernst, N.L., Proff, R., Deng, J., Hol, W., and Stuart, K. (2008) KREPA6 is an RNA-binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. *RNA* **14**: 347–358.

Tasker, M., Wilson, J., Sarkar, M., Hendriks, E., and Matthews, K. (2000) A novel selection regime for differentiation defects demonstrates an essential role for the stumpy form in the life cycle of the African trypanosome. *Molecular Biology of the Cell* **11**: 1905–1917.

Tessier, D.C., Brousseau, R., and Vernet, T. (1986) Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA ligase. *Analytical Biochemistry* **158**: 171–178.

Thon, G., Baltz, T., and Eisen, H. (1989) Antigenic diversity by the recombination of pseudogenes. *Genes and Development* **3**: 1247–1254.

Tielens, A.G., and Hellemond, J.J. Van (1998) Differences in energy metabolism between Trypanosomatidae. *Parasitology Today* **14**: 265–272.

Tielens, A.G., and Hellemond, J.J. Van (2009) Surprising variety in energy metabolism within Trypanosomatidae. *Trends in Parasitology* **25**: 482–490.

Tielens, A.G.M., Rotte, C., Hellemond, J.J. Van, and Martin, W. (2002) Mitochondria as we don't know them. *Trends in Biochemical Sciences* **27**: 564–572.

Tobie, E.J. (1951) Loss of the kinetoplast in a strain of *Trypanosoma equiperdum*. *Transactions of the American Microscopical Society* **70**: 251–254.

Trotter, J., Ernst, N., Carnes, J., Panicucci, B., and Stuart, K.D. (2005) A Deletion Site Editing Endonuclease in *Trypanosoma brucei*. **20**: 403–412.

Turner, C. (1999) Antigenic variation in *Trypanosoma brucei* infections: an holistic

view. *Journal Cell Science* **112**: 3187–3192.

Turner, C.M.R., and Barry, J.D. (1989) High frequency of antigenic variation in *Trypanosoma brucei rhodesiense* infections. *Parasitology* **99**: 67–75.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Dan Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J.P., De Baetselier, P., Braisser, R. and Pays, E. (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**: 83–87.

Vanhamme, L., Perez-Morga, D., Marchal, C., Speijer, D., Lambert, L., Geuskens, M., Alexandre, S., Ismaïli, N., Göringer, U., Benne, R. and Pays, E.. (1998) *Trypanosoma brucei* TBRGG1, a mitochondrial oligo(u)-binding protein that co-localizes with an in vitro RNA editing activity. *Journal of Biological Chemistry* **273**: 21825–21833.

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997) Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of Cell Science* **110**: 2661–2671.

Verner, Z., Cermáková, P., Skodová, I., Kriegová, E., Horváth, A., and Lukeš, J. (2011) Complex I (NADH:ubiquinone oxidoreductase) is active in but non-essential for procyclic *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **175**: 196–200.

Verner, Z., Skodová, I., Poláková, S., Durišová-Benkovičová, V., Horváth, A., and Lukeš, J. (2013) Alternative NADH dehydrogenase (NDH2): intermembrane-space-facing counterpart of mitochondrial complex I in the procyclic *Trypanosoma brucei*. *Parasitology* **140**: 328–337.

Vickerman, K. (1965) Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature* **208**: 762–766.

Vickerman, K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin* **41**: 105–114.

Vincendeau, P., and Bouteille, B. (2006) Immunology and immunopathology of African trypanosomiasis. *Anais da Academia Brasileira de Ciências* **78**: 645–665.

Wang, B., Ernst, N.L., Palazzo, S.S., Panigrahi, A.K., Salavati, R., and Stuart, K. (2003a) TbMP44 is essential for RNA editing and structural integrity of the editosome in *Trypanosoma brucei*. *Eukaryotic Cell* **2**: 578–587.

Wang, L.K., Ho, C.K., Pei, Y., and Shuman, S. (2003b) Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction. *The Journal of Biological Chemistry* **278**: 29454–29462.

Wang, L.K., Nandakumar, J., Schwer, B., and Shuman, S. (2007) The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair. *RNA* **13**: 1235–1244.

Wang, P., Chan, C.M., Christensen, D., Zhang, C., Selvadurai, K., and Huang, R.H. (2012) Molecular basis of bacterial protein Hen1 activating the ligase activity of bacterial protein Pnkp for RNA repair. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 13248–53.

Wang, Z., and Englund, P.T. (2001) RNA interference of a trypanosome topoisomerase II causes progressive loss of mitochondrial DNA. *The EMBO journal* **20**: 4674–83.

Wastling, S.L., and Welburn, S.C. (2011) Diagnosis of human sleeping sickness: sense and sensitivity. *Trends in parasitology* **27**: 394–402.

Weelden, S.W. van, Fast, B., Vogt, A., der, M.P. van, Saas, J., Hellemond, J.J. Van, *et al.* (2003) Procyclic *Trypanosoma brucei* do not use Krebs cycle activity for energy generation. *Journal of Biological Chemistry* **278**: 12854–12863.

Welburn, S.C., Fèvre, E.M., Colemam, P.J., Odiit, M., and Maudlin, I. (2001) Sleeping sickness: a tale of two diseases. *Trends in Parasitology* **17**: 19–24.

Weng, J., Aphasizheva, I., Etheridge, R.D., Huang, L., Wang, X., Falick, A.M., and Aphasizhev, R. (2008) Guide RNA-binding complex from mitochondria of trypanosomatids. *Molecular Cell* **32**: 198–209.

Williamson, J., and Rollo, I. (1959) Drug resistance in trypanosomes; cross resistance analyses. *British Journal of Pharmacology and Chemotherapy* **14**: 423–430.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A.M. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **99**: 89–101.

Woo, P.T.K. (1970) The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis. *Acta Tropica* **27**: 384–6.

Worthey, E.A., Schnaufer, A., Mian, I.S., Stuart, K., and Salavati, R. (2003) Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Research* **31**: 6392–6408.

Xong, H. Van, Vanhamme, L., Chamekh, M., Chimfwembe, C.E., Abbeele, J. Van Den, Pays, A., *et al.* (1998) A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in *Trypanosoma rhodesiense*. *Cell* **95**: 839–846.

Yang, Z. (2000) Maximum likelihood estimation on large phylogenies and analysis

of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution* **51**: 423–432.

Yin, S., Ho, C.K., and Shuman, S. (2003) Structure-function analysis of t4 RNA ligase 2. *Journal of Biological Chemistry* **278**: 17601–17608.

Zhang, C., Chan, C.M., Wang, P., and Huang, R.H. (2012) Probing the substrate specificity of the bacterial Pnkp/Hen1 RNA repair system using synthetic RNAs. *RNA* **18**: 335–44.

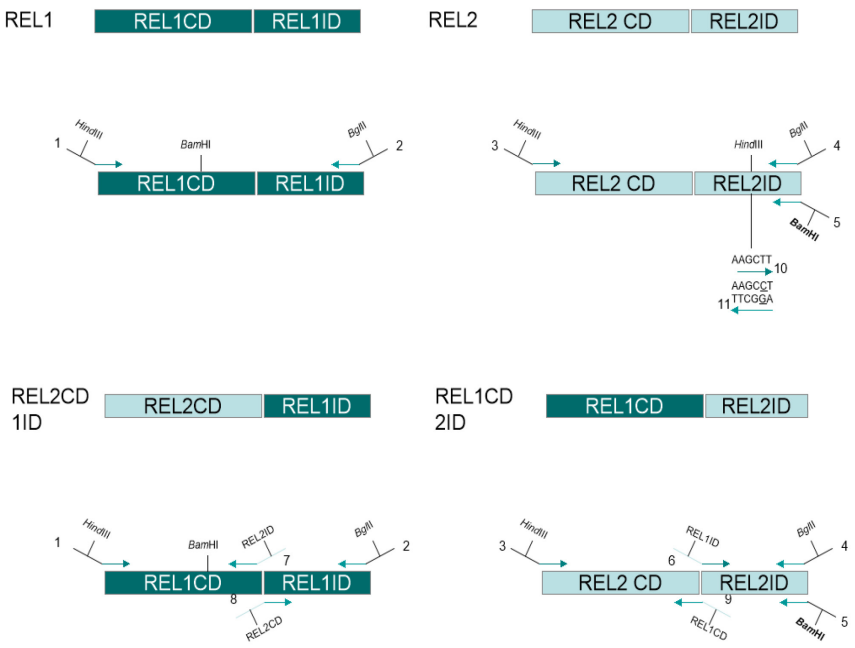
Zhang, L., and Li, W.-H. (2004) Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology & Medicine* **21**: 236–239.

Zhelonkina, A.G., O'Hearn, S.F., Law, J.A., Cruz-Reyes, J., Huang, C.E., Alatortsev, V.S., and Sollner-Webb, B. (2006) *T. brucei* RNA editing: action of the U-insertional TUTase within a U-deletion cycle. *RNA* **12**: 476–487.

Appendices

Appendix 1

Below is illustrated the PCR strategy, for cloning and expression via pHD1344-TAP in REL1 cKO cell lines, for the full length and chimeric ligases (Chapter 2).



Primers used

Number Schematic	Primer sequence
1	ATAAAGCTTATGCAACTCCAAAGGTTGGG
2	ATAAGATCTTTCGCCCTTTGTGGGGGCAG
3	ATAAAGCTTATGTTGCGTCGCCTCGGTGT
4	ATAGGATCCTTCGCTAAAGTCAGGAGACT
5	ATAAGATCTTTCGCTAAAGTCAGGAGACT
6	CCCGGTAAGCAGAAGGAACCTCGCGTAGATGAAATG
7	GACATTTATCGACACGGTACGAAGAGACTCCATCAATC or GTTGAAGCACCCCGGTAAGCAAGGGCCTCGCGTAGATG
8	CACAGACCGAGCGCAAGGGCTGAAGGAGACATTTATCG
9	GATTGATGGAGTCTCTTCGTACCGTGTCGATAAATGTC or CATCTACGCGAGGCCCTTGCTTACCGGGGTGCTTCAAC
10	ATCCAATTCAACTCGAAGCTTCGAAACTACTCCTTA
11	TAAGGAGTATTTCGAGGCTTCGAGTTGAATTGGAT

Appendix 2

Below is listed the solution, and their methods of preparation, used in the Northern Blotting protocol, which has been modified from Chapter 7 of Maniatis *et al* (1982) (Chapter 3).

[10x] MOPS

23.13g MOPS

10ml of 2.5 M Na-Acetate (pH 7.0)

10ml of 0.5 M EDTA

Make up to 500ml with water, Autoclave and store in the dark at 4°C

RNA Gel Loading Buffer

150µl of formamide

83µl of formaldehyde (37%)

50µl of 10 x MOPS

0.01% Bromophenol blue

50µl glycerol

167µl of water (to give a total volume of 500µl)

[20x] SSC	for 1 litre
3 M NaCl	175.2g
0.3 M Tri Sodium Citrate	88.2g

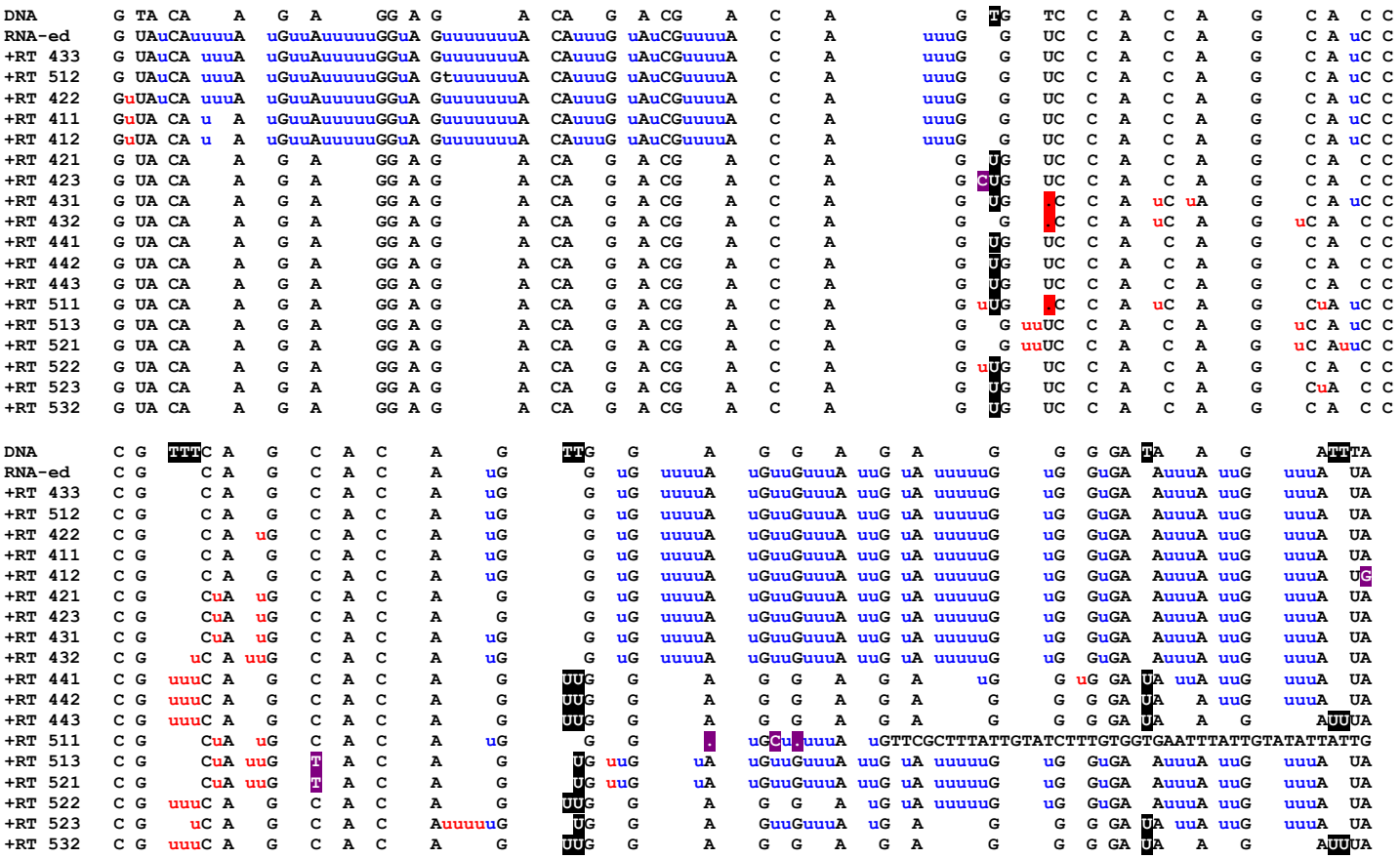
Maleic Acid Buffer	for 500ml	
Maleic Acid	5.8g	100mM Maleic acid
NaCl	4.38g	150mM NaCl

adjust to pH 7.5 with NaOH and Autoclave

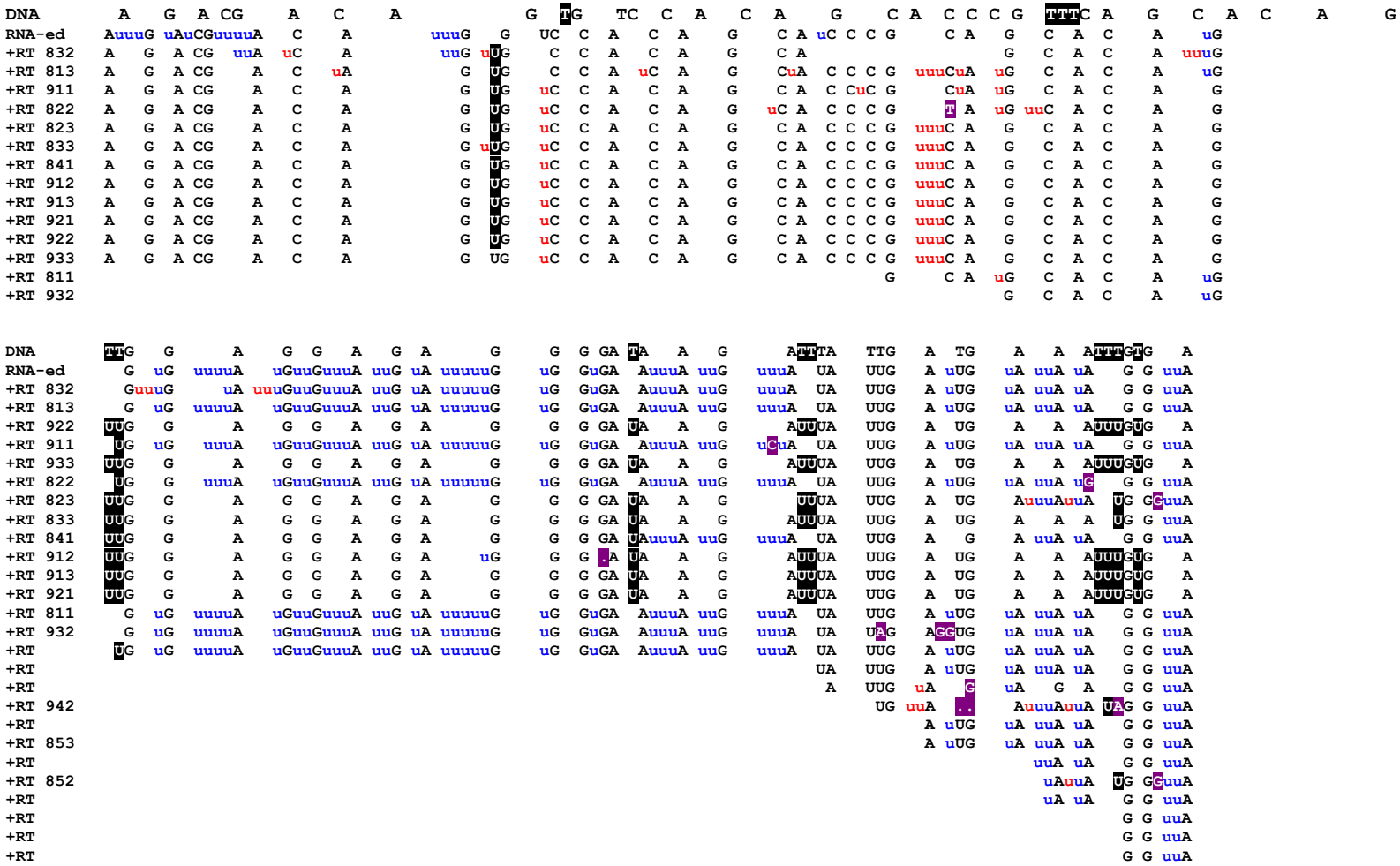
Appendix 3

Below is listed the full length sequences (over two paragraphs, 5' – 3') cloned from limited sequencing reactions (Chapter 3). Correctly inserted Us - blue, incorrectly inserted Us – red. Non-deleted Us – white.

ND7 + REL1



ND7 - REL1


DNA
RNA-ed
+RT 832
+RT 813
+RT 911
+RT 822
+RT 823
+RT 833
+RT 841
+RT 912
+RT 913
+RT 921
+RT 922
+RT 933
+RT 811
+RT 932
+RT 942
+RT 853
+RT 852

RPS12 + REL1

```
DNA      CCGG AACCGACG G A  G A GCTTCTTTTTG A    A  TA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GTTTC
RNA-ed   CCGGuAACCGACGuGuAuuGuAuGC  C    GuAuuuuA uUAuAuA  AuuuuGuuuG G        AuGuuGCGuuGuuuuuuuuGuuGuuuuAuuGGuuuAGuuAuG  UC
E (A)    CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A G  UC
PE1 (B)  CCGGuAACCGACGuGuAuuGuAuGC  C    GuAuuuuAuuUAuAuA  AuuuuGuuuG G        AuGuuGCGuuG    uuuuuGuuGuuuuAuuGGuuuAGuuAuG  UC
PE2 (B)  CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A AuuA   G   GuGuuuuuuuuAuGuuGCGuuG    uuuuG uG   uA  GGuuuAGuuAuG  UC
U1 (C)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC
U1 (C)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC
U3 (C)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC

DNA      A  A   A  A A G ATTTTGGG TGG G G   G  G A  AC C CTTTTGTTTTTG  GTT  A A  A  G A  A  A  CA TCGTTTA  G AAG  AGATTTTAGA AT
RNA-ed   AuuAuuuAuuAuA G A   GGG UGGuG GuuuuGuuG AuuuAC C C   G   G  uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
E(A)     A  A   A  AuAuGuA   GGGuUGG GuG  uuGuuGuA  uAC C C   G  UUGuuuGUUuuAuAuuAuuGuAuuA   AuuCA  CG   AuuGuAAGuuAGA UUUAGAuAU
PE1(B)   AuuAuuuAuuAuA G A   GGG UGGuG GuuuuGuuG AuuuAC C C   G   G  uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
PE2(B)   AuuAuuuAuuAuA G A   GGG UGGuG GuuuuGuuG AuuuAC C C   G   G  uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
U1 (C)   A  A   A  A A G AUUUGGG UGG G G   G  G A  AC C C   G   G  G U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
U1 (C)   A  A   A  A A G AUUUGGG UGG G GuuuuGuuG AuuuAC C C   G   G  uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
U3 (C)   A  A   A  A A G A   GGG UGG G G   G  G A  AC C CUUUGUUUUG  GUU  A A  A  G A  A  A  CA TCGUUUA  G AAG  AGAUUUUAGAuAU
```

RPS12 - REL1

```
DNA      CCGG AACCGACG G A  G A GCTTCTTTTTG A    A  TA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GTTTC
RNA      CCGGuAACCGACGuGuAuuGuAuGC  C    GuAuuuuA uUAuAuA  AuuuuGuuuG G        AuGuuGCGuuGuuuuuuuuGuuGuuuuAuuGGuuuAGuuAuG  UC
U1 (D)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC
U2 (D)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC
U3 (D)   CCGG AACCGACG G A  G A GCUUCUUUUG A    A  UA A A  A    G   G G        A G  GCG  G        G  G    A  GG   AG  A GUUUC

DNA      A  A   A  A A G ATTTTGGG TGG G G   G  G A  AC C CTTTTGTTTTTG  GTT  A A  A  G A  A  A  CA TCGTTTA  G AAG  AGATTTTAGA AT
RNA-ed   AuuAuuuAuuAuA G A   GGG UGGuG GuuuuGuuG AuuuAC C C   G   G  uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
U1 (C)   A  A   A  A A G AUUUGGG UGG G G   G  G A  AC C CUUUGUUUUG  GUU  A A  A  G A  A uuA  CA      UUAuuGuAAGuuAGA UUUAGAuAU
U1 (C)   A  A   A  A A G AUUUGGG UGG G G   G  G A  AC CuC UUG  UUG  GUU uAuA  A  G A uA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
U3 (C)   A  A   A  A A G AUUUGGG UGG G G   G  GuA  ACuC C   G  UUG  -UU uAuA uA  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F1 (E)                                                    uG U  A A  A  GuA uA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F1 (E)                                                ACUC uAuA uA  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F1 (E)                                                uG U  A A  A  GuAuuA  uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F1 (E)                                                                uAuuuA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F1 (E)                                                                   uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F2 (F)                                              G  UUG   GUU uAuA uA  GuA uA uuA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F2 (F)                                                                uGuuuA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F2 (F)                                                                   uA  CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
F2 (F)                                                                       CA  CG  UAuuGuAAGuuAGA UUUAGAuAU
```

Appendix 4

Below is collated the reads out for both Ion Torrent runs (Chapter 3) performed on +REL1 (Ion_Xpress 011) and –REL1 (Ion_Xpress 012) ND71 samples.

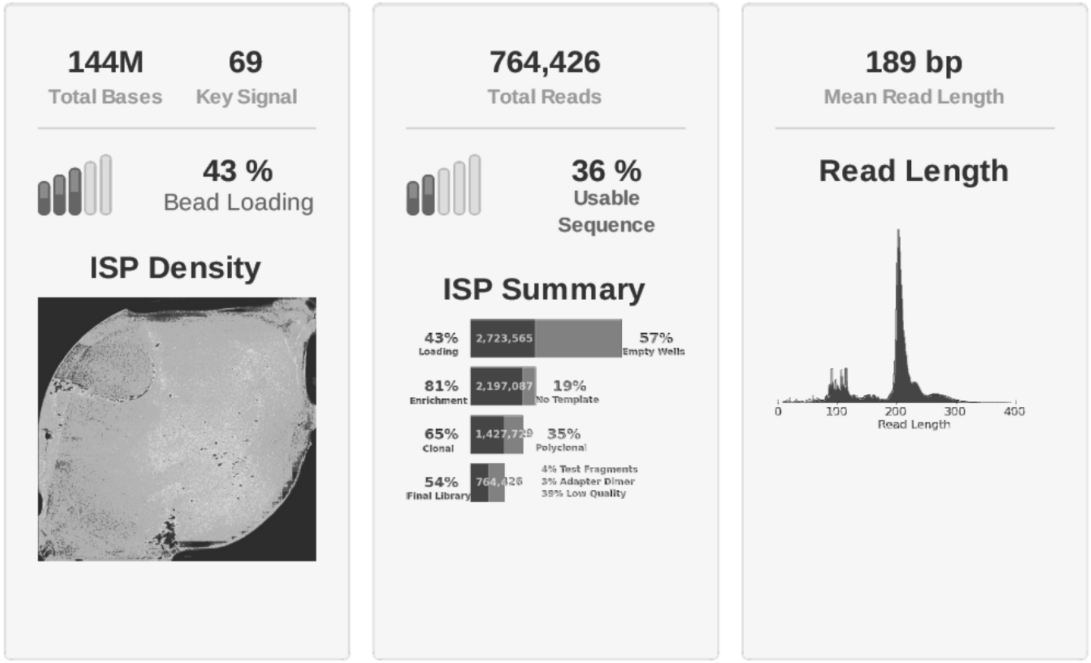
First run

Run: R_2012_11_16_18_52_34_user_ISA-49

Barcode Name	Sample	Output	%>=Q20	Reads	Mean Read Length	Read Length Histogram	BAM
No barcode	nd71-pool	6.6M	2.5M	35293	189 bp		
IonXpress_011	nd71-pool	66.1M	28.6M	334197	198 bp		
IonXpress_012	nd71-pool	71.6M	32.6M	394224	181 bp		

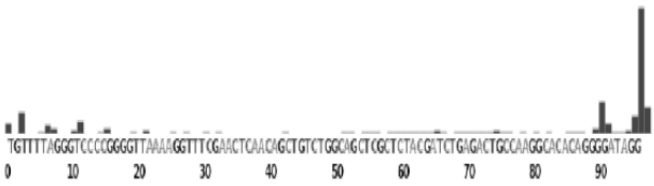
Run Summary

Unaligned



Plugin Output

Test Fragments

Test Fragment	Reads	Percent 50AQ17	Read length histogram
TF_A	52,076	75%	

Analysis Details

Run Name	R_2012_11_16_18_52_34_user_ISA-49
Run Date	2012/11/16 06:52 PM
Run Cycles	20
Run Flows	640
Project	E121179
Sample	nd71-pool
Library	none
PGM	Isaac
Flow Order	TACGTACGTCTGAGCATCGATCGATGTACAGC
Library Key	TCAG
TF Key	ATCG
Chip Check	Passed
Chip Type	316D
Chip Data	single
Notes	
Barcode Set	IonXpress

Analysis Name	ISA-49_beverly_filter_off
Analysis Date	2012-12-10
Analysis Flows	0
runID	MW6HD

Support

- Download the Customer Support Archive
- View the report log

Software Version

Torrent_Suite	**3.2.1**
Datacollect	210
Graphics	31
LiveView	395
OS	19
Script	20.1.4
host	FLDNNQ1
ion-alignment	3.2.1-1
ion-analysis	3.2.5-1
ion-dbreports	3.2.15-1
ion-gpu	3.0.0-1
ion-pipeline	3.2.10-1
ion-plugins	3.2.8-1

Second run

Run: R_2012_11_23_13_04_26_user_ISA-50

Barcode Name	Sample	Output	%>=Q20	Reads	Mean Read Length	Read Length Histogram	BAM
No barcode	nd71-pool	5.4M	1.7M	23693	231 bp		
IonXpress_011	nd71-pool	65.3M	28.7M	323294	202 bp		
IonXpress_012	nd71-pool	71.1M	32.9M	386214	184 bp		

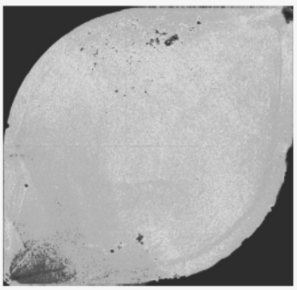
Run Summary

Unaligned

142M Total Bases | **69** Key Signal

52 % Bead Loading

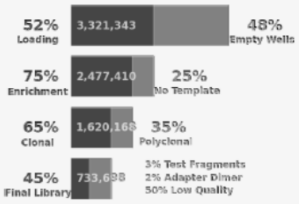
ISP Density



733,698 Total Reads

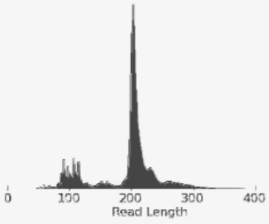
31 % Usable Sequence

ISP Summary

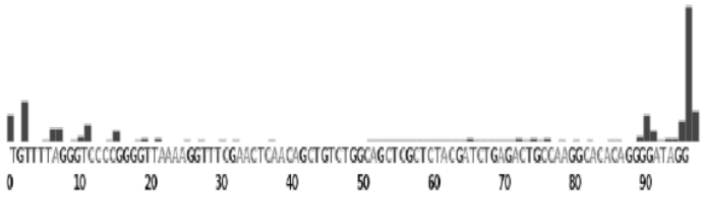


193 bp Mean Read Length

Read Length



Test Fragments

Test Fragment	Reads	Percent 50AQ17	Read length histogram
TF_A	48,498	65%	

Analysis Details

Run Name	R_2012_11_23_13_04_26_user_ISA-50
Run Date	2012/11/23 01:04 PM
Run Cycles	20
Run Flows	640
Project	E121179
Sample	nd71-pool
Library	none
PGM	Isaac
Flow Order	TACGTACGTCTGAGCATCGATCGATGTACAGC
Library Key	TCAG
TF Key	ATCG
Chip Check	Passed
Chip Type	316D
Chip Data	single
Notes	
Barcode Set	IonXpress

Analysis Name	ISA50-beverly_filter_off
Analysis Date	2012-12-10
Analysis Flows	0
runID	UC7GK

Support

- Download the Customer Support Archive
- View the report log

Software Version

Torrent_Suite	**3.2.1**
Datacollect	210
Graphics	31
LiveView	395
OS	19
Script	20.1.4
host	FLDNNQ1
ion-alignment	3.2.1-1
ion-analysis	3.2.5-1
ion-dbreports	3.2.15-1
ion-gpu	3.0.0-1
ion-pipeline	3.2.10-1
ion-plugins	3.2.8-1

Appendix 5

Tabulated overleaf is the data collated from Ion Torrent sequencing that has been subjected to the pipeline described in Chapter 3. The most 5' and 3' end of ND7 and the edited region are highlighted in yellow. The 3' primer sequences used in the sequencing reaction is marked in red. Deletion (ESd) and insertion (ESi) editing sites are indicated in blue and pink respectively. The expected number of uridylyls in fully/correctly editied ND7 at each edited site is highlighted in turquoise.

The tables are collated in the order: Category 1 +REL1, Category 1 –REL1, Category 2 +REL1 and Category 2 -REL1.

Category 1

ND7 – Ion Torrent

+REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0	2	0	5	0	0	0	0	0	0	0	0	0	0
	A	110	1	1	0	0	0	4	0	0	0	0	0	0	0	0	0	0
	A	112	2	2	0	3	0	3	4	0	0	0	0	0	0	0	0	0
	A	114	0	0	0	3	7	0	0	0	0	0	0	0	0	0	0	0
	A	116	3	3	0	170	2	1	32	15	2	0	0	0	0	0	0	0
	A	118	1	1	0	170	32	164	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	11017	9	11084	115	0	0	0	0	0	0	0	0	0
	A	122	0	0	0	8079	11366	181	3	0	0	0	0	0	0	0	0	0
	A	124	1	1	0	7406	60	20020	28	0	1	0	0	0	0	0	0	0

C	126	0	0	0	861	26834	22	2	0	0	0	0	0	0	0	0	0
A	128	0	0	0	97	27713	1	0	0	0	0	0	0	0	0	0	0
A	130	0	0	0	26	27804	7	0	0	0	0	0	0	0	0	0	0
A	132	0	0	0	9	27835	1	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	0	27845	0	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	0	27844	1	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	5	27845	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	10	27850	0	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	0	27840	20	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	0	27857	3	0	0	0	0	0	0	0	0	0	0
G	146	1	1	0	0	44	27739	77	0	0	0	0	0	0	0	0	0
A	148	0	0	0	0	27843	17	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	9	27839	21	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	0	32	27750	87	0	0	0	0	0	0	0	0	0
C	154	0	0	0	8	27866	3	0	0	0	0	0	0	0	0	0	0

	A	156	0	0	0	0	27874	3	0	0	0	0	0	0	0	0	0	0
	G	158	1	1	0	0	81	27634	162	0	0	0	0	0	0	0	0	0
	A	160	0	0	0	12	27867	7	0	0	0	0	0	0	0	0	0	0
	A	162	1	1	0	0	14	27788	83	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	0	27838	38	4	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	0	27325	249	285	9	0	0	0	0	0	0	0	0
	A	168	1	1	0	5	216	26850	646	15	6	1	0	0	0	0	0	0
ES39i	C	170	0	1	1	0	25674	1644	16	1	0	0	0	0	0	0	0	0
	A	172	0	0	0	6	27253	9	1	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	5	25041	508	132	1371	52	0	0	0	0	0	0	0
ES37i	G	176	0	1	1	27	24614	1798	459	12	3	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	8	24300	60	2202	47	0	0	1	1	0	0	0	0
ES35i	G	180	0	5	5	57	24267	40	449	39	105	1537	132	3	0	0	0	0
	G	182	0	0	0	1	26407	31	5	0	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	1	23998	2303	29	3	0	0	0	0	0	0	0	0

	G	186	0	0	0	0	25989	89	6	0	4	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	23	23546	53	12	35	19	29	75	1456	594	77	5	2
	C	190	0	0	0	5	25728	8	6	5	5	3	0	0	0	0	0	0
	A	192	0	0	0	3	25618	35	4	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	25	22956	154	204	2091	156	0	0	0	0	0	0	0
ES31i	A	196	0	1	1	0	22940	2350	119	5	0	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	0	22919	2293	46	0	0	0	0	0	0	0	0	0
	G	200	0	0	0	0	24979	9	0	0	1	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	17	21766	226	496	122	1912	228	0	1	0	0	2	3
	C	204	0	0	0	0	24193	103	28	37	2	0	0	0	0	0	0	0
	A	206	0	0	0	0	23880	86	12	4	37	12	0	0	0	0	0	0
ES28i	G	208	0	3	3	40	21351	74	247	1894	171	0	0	0	2	0	0	0
ES27d	G	210	1	0	-1	0	2749	20163	291	19	8	2	0	0	0	0	0	0
	C	212	1	1	0	0	2302	19944	42	6	8	2	0	0	1	0	1	3
	C	214	0	0	0	0	21899	99	0	6	0	0	0	0	0	0	0	0

	A	216	0	0	0	0	21657	88	5	0	0	0	0	0	0	0	0	0
	C	218	0	0	0	7	20091	1132	26	58	2	0	0	0	0	0	0	0
	A	220	0	0	0	0	20813	69	88	1	9	0	0	0	0	0	0	0
	G	222	0	0	0	3	19945	270	186	161	30	0	0	0	0	0	0	0
	C	224	0	0	0	20	19188	759	25	175	9	1	1	0	0	0	0	0
	A	226	0	0	0	4	19420	369	20	0	0	0	0	0	0	0	0	0
ES26i	C	228	0	1	1	0	16964	2695	57	7	2	4	0	0	0	0	0	0
	C	230	0	0	0	0	19694	18	3	0	0	0	0	0	0	0	0	0
	C	232	0	0	0	37	19112	424	2	0	0	0	0	0	0	0	0	0
	G	234	0	0	0	0	19373	41	13	2	0	0	0	0	0	0	0	0
ES25d	C	236	3	0	-3	96	3441	902	585	14184	244	29	6	0	0	0	0	0
	A	238	0	0	0	110	17904	874	36	3	1	0	2	0	0	0	0	0
	G	240	0	0	0	146	16000	2004	532	116	44	7	9	2	0	0	1	3
	C	242	0	0	0	65	18067	239	28	33	57	49	8	0	0	1	0	0
	A	244	0	0	0	24	18170	93	57	2	2	2	0	0	0	0	0	0

	C	246	0	0	0	72	17662	305	69	40	14	26	0	0	0	0	1	0
	A	248	0	0	0	117	17741	154	52	27	1	0	0	0	0	0	0	0
ES24i	G	250	0	1	1	44	13906	3479	221	93	19	107	11	1	4	1	0	0
ES23d	G	252	2	0	-2	308	3856	659	12904	394	71	22	3	4	0	0	0	0
Es22i	G	254	0	1	1	118	12895	4481	370	194	37	36	3	0	0	0	0	0
Es21i	A	256	0	4	4	186	12608	949	217	399	3579	197	35	4	0	0	0	0
ES20i	G	258	0	1	1	87	12537	4440	217	170	604	20	0	0	0	0	0	0
ES19i	G	260	0	2	2	68	12125	46	5282	157	204	47	3	0	0	0	0	0
ES18i	A	262	0	3	3	93	12006	87	579	4841	343	1	2	1	0	0	0	0
ES17i	G	264	0	2	2	100	11878	593	5170	98	1	14	2	0	0	0	0	0
ES16i	A	266	0	1	1	32	12330	5231	56	44	0	7	1	0	0	0	2	0
ES15i	G	268	0	5	5	105	11799	134	559	82	194	4628	312	5	2	0	0	0
ES14i	G	270	0	1	1	14	11939	4998	83	560	15	1	0	0	0	0	0	0
	G	272	0	0	0	30	17440	82	10	1	0	0	0	0	0	0	0	0
ES13i	G	274	0	1	1	164	11756	5563	123	0	0	1	0	0	0	0	0	0

	A	276	0	0	0	60	17251	215	2	1	0	0	0	0	0	0	0	0
ES12d	A	278	1	0	-1	158	5845	11535	71	41	0	0	0	0	0	0	0	0
ES11i	A	280	0	3	3	299	11315	899	393	4789	269	0	0	0	0	0	0	0
ES10i	G	282	0	2	2	32	11167	759	5514	211	7	1	0	0	0	0	0	0
ES9i	A	284	0	3	3	247	11656	106	252	5460	184	6	14	8	0	0	0	0
ES8d	A	286	3	1	-2	174	47	5738	457	11247	159	13	0	0	0	0	0	0
	G	288	2	2	0	86	22	235	16677	715	5	32	7	0	0	0	0	0
	A	290	0	0	0	95	17220	84	145	70	1	3	0	0	0	0	0	0
ES7i	G	292	1	2	1	98	338	11097	5634	412	4	11	1	0	0	0	0	0
ES6i	A	294	0	1	1	496	11326	5718	143	9	75	1	0	0	0	0	0	0
ES5i	A	296	0	2	2	267	11148	735	5795	121	14	0	0	0	0	0	0	0
ES4i	A	298	0	1	1	9	11038	6021	662	24	0	0	0	6	0	0	0	0
ES3d	G	300	3	0	-3	31	6129	704	248	9998	585	2	0	0	0	0	0	0
ES2d	G	302	1	0	-1	185	6873	10439	87	6	0	0	0	0	0	0	0	0
ES1i	A	304	0	2	2	693	9896	461	5917	1204	39	1	0	0	0	0	0	0

G	306	3	3	0	196	4	60	565	15482	1623	94	5	0	3	1	0	0
C	308	0	0	0	72	17287	174	3	1	0	0	0	0	0	0	0	0
A	310	0	0	0	7	17359	9	0	0	0	0	0	0	0	0	0	0
C	312	1	1	0	98	50	16961	261	0	0	0	0	0	0	0	0	0
G	314	0	0	0	14	16825	33	2	0	0	0	0	0	0	0	0	0
G	316	1	1	0	6	29	16210	359	0	0	0	0	0	0	0	0	0
G	318	0	0	0	173	16357	12	1	0	0	0	0	0	0	0	0	0
A	320	1	1	0	218	83	16373	142	0	0	0	0	0	0	0	0	0
C	322	0	0	0	64	16288	0	0	0	0	0	0	0	0	0	0	0
A	324	0	0	0	7	16184	12	0	0	0	0	0	0	0	0	0	0
G	326	0	0	0	0	15940	0	0	0	0	0	0	0	0	0	0	0
A	328	0	0	0	60	15925	11	0	0	0	0	0	0	0	0	0	0
A	330	0	0	0	209	15993	0	0	0	0	0	0	0	0	0	0	0
A	332	0	0	0	4	16186	1	0	0	0	0	0	0	0	0	0	0
A	334	0	0	0	199	16003	2	0	0	0	0	0	0	0	0	0	0

G	336	0	0	0	0	15902	276	0	0	0	0	0	0	0	0	0	0
A	338	2	2	0	1	11	116	15574	442	0	0	0	0	0	0	0	0
G	340	1	1	0	203	37	15858	100	0	0	0	0	0	0	0	0	0
G	342	1	1	0	1106	6	16537	48	0	0	0	0	0	0	0	0	0
A	344	0	0	0	1	16390	58	0	0	0	0	0	0	0	0	0	0
A	346	0	0	0	0	16328	29	2	0	0	0	0	0	0	0	0	0
A	348	1	1	0	0	50	15931	287	1	0	0	0	0	0	0	0	0
A	350	1	1	0	0	26	15782	83	0	0	0	0	0	0	0	0	0
A	352	0	0	0	0	15828	32	0	0	0	0	0	0	0	0	0	0
A	354	0	0	0	0	15852	1	0	0	0	0	0	0	0	0	0	0
A	356	0	0	0	0	15771	0	0	0	0	0	0	0	0	0	0	0
G	358	0	0	0	0	15713	33	1	0	0	0	0	0	0	0	0	0
G	360	1	1	0	0	4	15450	141	0	0	0	0	0	0	0	0	0
A	362	1	1	0	0	24	15193	190	0	0	0	0	0	0	0	0	0
G	364	0	0	0	0	15125	52	2	0	0	0	0	0	0	0	0	0

	A	366	0	0	0	0	15082	66	1	0	0	0	0	0	0	0	0	0
	A	368	0	0	0	0	15125	0	0	0	0	0	0	0	0	0	0	0
	C	370	0	0	0	0	15091	4	0	0	0	0	0	0	0	0	0	0
	A	372	0	0	0	0	15044	7	0	0	0	0	0	0	0	0	0	0
	A	374	0	0	0	0	15026	4	0	0	0	0	0	0	0	0	0	0
	G	376	1	1	0	0	21	14895	80	0	0	0	0	0	0	0	0	0
	C	378	1	1	0	0	88	14650	114	3	0	0	0	0	0	0	0	0
	C	380	2	2	0	0	4	488	14146	27	0	0	0	0	0	0	0	0
	C	382	0	0	0	0	14537	0	3	0	0	0	0	0	0	0	0	0
3' end	G	384	0	0	0	0	14349	15	0	0	0	0	0	0	0	0	0	0

Category 1

ND7 – Ion Torrent - REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0	0	0	4	0	0	0	0	0	0	0	0	0	0
	A	110	1	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0
	A	112	2	2	0	1	0	1	4	0	0	0	0	0	0	0	0	0
	A	114	0	0	0	3	5	0	0	0	0	0	0	0	0	0	0	0
	A	116	3	3	0	121	3	0	23	5	0	0	0	0	0	0	0	0
	A	118	1	1	0	232	33	107	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	10920	14	11044	119	1	0	0	0	0	0	0	0	0
	A	122	0	0	0	6791	11277	143	2	4	0	0	0	0	0	0	0	0
	A	124	1	1	0	5815	110	18529	42	1	0	0	0	0	0	0	0	0

C	126	0	0	0	600	23874	14	0	0	0	0	0	0	0	0	0	1
A	128	0	0	0	115	24482	3	0	0	0	0	0	0	0	0	0	1
A	130	0	0	0	29	24587	12	0	0	0	0	0	0	0	0	0	0
A	132	0	0	0	3	24625	3	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	1	24628	3	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	0	24631	1	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	4	24631	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	5	24634	1	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	0	24625	15	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	4	24637	3	0	0	0	0	0	0	0	0	0	0
G	146	1	1	0	1	44	24541	60	0	0	0	0	0	0	0	0	0
A	148	0	0	0	0	24633	12	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	4	24636	9	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	4	26	24531	92	0	0	0	0	0	0	0	0	0
C	154	0	0	0	7	24652	1	0	0	0	0	0	0	0	0	0	0

	A	156	0	0	0	9	24656	4	0	0	0	0	0	0	0	0	0	0
	G	158	1	1	0	0	72	24489	108	0	0	0	0	0	0	0	0	0
	A	160	0	0	0	6	24665	3	0	0	0	0	0	0	0	0	0	0
	A	162	1	1	0	2	5	24603	66	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	0	24672	2	0	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	0	24617	36	14	0	0	0	0	0	0	0	0	0
	A	168	1	1	0	5	170	24107	216	1	0	0	0	0	0	0	0	0
ES39i	C	170	0	1	1	0	24013	65	1	0	0	0	0	0	0	0	0	0
	A	172	0	0	0	3	24055	1	0	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	13	23862	17	0	59	1	0	0	0	0	0	0	0
ES37i	G	176	0	1	1	18	23699	75	17	0	0	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	0	23314	51	95	2	0	0	0	0	0	0	0	0
ES35i	G	180	0	5	5	0	23297	0	14	0	3	76	3	0	0	0	0	0
	G	182	0	0	0	0	23268	0	0	0	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	3	23008	118	1	0	0	0	0	0	0	0	0	0

	G	186	0	0	0	0	22888	1	0	0	0	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	0	22490	47	0	3	0	0	5	62	24	2	0	0
	C	190	0	0	0	0	22431	4	0	0	5	0	0	0	0	0	0	0
	A	192	0	0	0	17	22327	2	1	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	4	22079	56	10	87	6	0	0	0	0	0	0	0
ES31i	A	196	0	1	1	3	21986	112	31	0	0	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	8	21908	151	3	0	0	0	0	0	0	0	0	0
	G	200	0	0	0	0	22010	0	1	0	0	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	14	21795	20	19	13	99	12	0	0	0	0	0	0
	C	204	0	0	0	2	21689	56	3	14	1	0	0	0	0	0	0	0
	A	206	0	0	0	10	21547	25	0	0	12	1	0	0	0	0	0	0
ES28i	G	208	0	3	3	10	21195	37	26	102	5	0	0	0	0	0	0	0
ES27d	G	210	1	0	-1	26	170	20436	283	27	3	1	0	0	0	0	0	0
	C	212	1	1	0	8	1245	18763	28	1	5	0	0	0	0	0	0	1
	C	214	0	0	0	3	19746	13	0	0	0	0	0	0	0	0	0	0

	A	216	0	0	0	16	19513	54	1	0	1	0	0	0	0	0	0	0
	C	218	0	0	0	0	19099	78	0	4	3	0	0	0	0	0	0	0
	A	220	0	0	0	0	18841	9	9	0	2	0	0	0	0	0	0	0
	G	222	0	0	0	0	18423	45	38	14	0	2	0	0	0	0	0	0
	C	224	0	0	0	33	17948	167	6	20	2	0	0	0	0	0	0	0
	A	226	0	0	0	0	17806	41	0	1	0	0	0	0	0	0	0	0
ES26i	C	228	0	1	1	0	17587	215	11	0	0	3	1	0	0	0	0	0
	C	230	0	0	0	0	17802	4	0	0	0	0	0	0	0	0	0	0
	C	232	0	0	0	109	17488	145	4	0	0	0	0	0	0	0	0	0
	G	234	0	0	0	25	17605	22	13	1	0	0	0	0	0	0	0	0
ES25d	C	236	3	0	-3	408	395	390	413	16378	294	5	2	0	0	0	0	0
	A	238	0	0	0	245	16998	580	25	7	0	0	0	0	0	0	0	0
	G	240	0	0	0	370	16319	1065	228	50	7	0	5	0	0	0	1	0
	C	242	0	0	0	140	17385	123	9	30	16	3	1	1	0	0	0	0
	A	244	0	0	0	23	17332	30	29	1	1	0	0	0	0	0	0	0

	C	246	0	0	0	145	17059	87	19	22	12	4	1	0	0	0	0	0
	A	248	0	0	0	120	16986	57	40	8	7	0	0	0	0	0	0	0
ES24i	G	250	0	1	1	89	15695	1099	96	62	4	29	9	1	7	1	0	0
ES23d	G	252	2	0	-2	799	1449	781	14721	379	32	2	0	3	1	0	0	0
Es22i	G	254	0	1	1	481	15211	2401	143	148	19	45	2	0	0	0	0	0
Es21i	A	256	0	4	4	560	14728	1174	128	512	1791	46	23	2	0	0	0	0
ES20i	G	258	0	1	1	384	14599	2767	142	371	719	23	3	0	1	0	0	0
ES19i	G	260	0	2	2	78	14260	35	4187	50	91	5	4	0	0	0	0	0
ES18i	A	262	0	3	3	93	14113	104	308	3957	99	12	0	0	0	0	0	0
ES17i	G	264	0	2	2	56	13867	377	4135	58	9	42	0	0	0	0	0	0
ES16i	A	266	0	1	1	256	13999	4287	56	142	7	3	0	0	0	0	0	0
ES15i	G	268	0	5	5	108	13716	208	350	38	199	3922	132	0	1	0	0	1
ES14i	G	270	0	1	1	249	13969	4322	67	365	14	18	1	0	0	0	0	0
	G	272	0	0	0	119	18575	174	31	5	0	1	0	0	0	0	0	0
ES13i	G	274	0	1	1	134	13802	4924	55	7	1	0	0	0	0	0	0	0

	A	276	0	0	0	66	18575	236	23	5	0	0	0	0	0	0	0	0
ES12d	A	278	1	0	-1	303	5154	13401	104	140	15	1	0	0	0	0	0	0
ES11i	A	280	0	3	3	278	12987	1263	301	4451	66	0	0	0	0	0	0	0
ES10i	G	282	0	2	2	327	12906	857	5350	114	1	0	0	0	0	1	0	0
ES9i	A	284	0	3	3	733	13482	129	617	5361	134	10	10	1	0	0	0	0
ES8d	A	286	3	1	-2	3812	264	8735	533	13225	400	175	18	0	0	1	0	0
	G	288	2	2	0	711	86	802	22100	727	168	104	28	0	0	0	0	0
	A	290	0	0	0	1000	23026	109	478	497	7	31	1	0	0	0	0	0
ES7i	G	292	1	2	1	610	764	13602	10582	194	8	37	10	1	0	0	0	0
ES6i	A	294	0	1	1	1510	13846	11984	102	58	589	4	0	0	0	0	0	0
ES5i	A	296	0	2	2	1804	12949	2530	12584	202	75	1	0	0	0	0	0	0
ES4i	A	298	0	1	1	670	12755	13466	2607	60	7	0	0	0	0	0	0	0
ES3d	G	300	3	0	-3	3093	13549	3680	805	11229	654	40	7	6	2	2	0	1
ES2d	G	302	1	0	-1	1514	19159	12959	313	54	4	1	0	0	0	0	0	0
ES1i	A	304	0	2	2	1748	11327	1000	17825	3959	54	4	0	0	0	0	0	0

G	306	3	3	0	243	0	15	878	30456	3229	141	23	31	4	2	2	2
C	308	0	0	0	40	34144	168	0	0	0	0	0	0	0	0	0	0
A	310	0	0	0	69	34117	30	0	0	0	0	0	0	0	0	0	0
C	312	1	1	0	53	79	33706	261	0	0	0	0	0	0	0	0	0
G	314	0	0	0	18	33555	28	0	0	0	0	0	0	0	0	0	0
G	316	1	1	0	15	32	32816	412	0	0	0	0	0	0	0	0	0
G	318	0	0	0	11	32952	13	1	0	0	0	0	0	0	0	0	0
A	320	1	1	0	67	105	32607	139	0	0	0	0	0	0	0	0	0
C	322	0	0	0	19	32393	2	0	0	0	0	0	0	0	0	0	0
A	324	0	0	0	45	32166	19	0	0	0	0	0	0	0	0	0	0
G	326	0	0	0	16	31897	0	0	0	0	0	0	0	0	0	0	0
A	328	0	0	0	59	31893	10	0	0	0	0	0	0	0	0	0	0
A	330	0	0	0	85	31949	13	0	0	0	0	0	0	0	0	0	0
A	332	0	0	0	23	32031	0	0	0	0	0	0	0	0	0	0	0
A	334	0	0	0	18	31808	2	0	0	0	0	0	0	0	0	0	0

G	336	0	0	0	50	31482	309	2	0	0	0	0	0	0	0	0	0
A	338	2	2	0	46	12	192	30982	645	0	0	0	0	0	0	0	0
G	340	1	1	0	14	55	31080	104	0	0	0	0	0	0	0	0	0
G	342	1	1	0	68	3	30594	75	0	0	0	0	0	0	0	0	0
A	344	0	0	0	2	30174	71	0	0	0	0	0	0	0	0	0	0
A	346	0	0	0	0	30065	52	0	0	0	0	0	0	0	0	0	0
A	348	1	1	0	0	72	29530	318	0	0	0	0	0	0	0	0	0
A	350	1	1	0	0	63	29134	114	0	0	0	0	0	0	0	0	0
A	352	0	0	0	0	29195	53	0	0	0	0	0	0	0	0	0	0
A	354	0	0	0	0	29236	1	0	0	0	0	0	0	0	0	0	0
A	356	0	0	0	0	29091	0	0	0	0	0	0	0	0	0	0	0
G	358	0	0	0	0	29014	29	0	0	0	0	0	0	0	0	0	0
G	360	1	1	0	0	6	28583	227	1	0	0	0	0	0	0	0	0
A	362	1	1	0	0	30	28278	211	0	0	0	0	0	0	0	0	0
G	364	0	0	0	0	27947	99	0	0	0	0	0	0	0	0	0	0

	A	366	0	0	0	0	27897	114	0	0	0	0	0	0	0	0	0	0
	A	368	0	0	0	0	27963	2	0	0	0	0	0	0	0	0	0	0
	C	370	0	0	0	0	27926	0	0	0	0	0	0	0	0	0	0	0
	A	372	0	0	0	0	27865	13	0	0	0	0	0	0	0	0	0	0
	A	374	0	0	0	0	27832	5	0	0	0	0	0	0	0	0	0	0
	G	376	1	1	0	0	47	27626	114	0	0	0	0	0	0	0	0	0
	C	378	1	1	0	0	147	27309	116	1	0	0	0	0	0	0	0	0
	C	380	2	2	0	0	6	686	26560	51	0	0	0	0	0	0	0	0
	C	382	0	0	0	0	27045	1	3	0	0	0	0	0	0	0	0	0
3' end	G	384	0	0	0	0	26779	25	0	0	0	0	0	0	0	0	0	0

Category 2

ND7 – Ion Torrent

+REL1

			no of Ts present before that position				no. of reads with indicated no. of Ts present before that position											
	T-stripped sequence	Position	pre-edited	edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0													
	A	110	1	1	0													
	A	112	2	2	0													
	A	114	0	0	0													
	A	116	3	3	0													
	A	118	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	86	0	37	0	0	0	0	0	0	0	0	0	0
	A	122	0	0	0	43	87	2	0	0	0	0	0	0	0	0	0	0
	A	124	1	1	0	185	2	225	7	0	0	0	0	0	0	0	0	0

C	126	0	0	0	91	312	8	1	0	0	0	0	0	0	0	0	0
A	128	0	0	0	129	406	0	0	0	0	0	0	0	0	0	0	0
A	130	0	0	0	24	534	1	0	0	0	0	0	0	0	0	0	0
A	132	0	0	0	1	559	0	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	2	560	0	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	1	562	0	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	1	563	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	2	564	0	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	3	565	1	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	290	566	88	1	0	0	0	0	0	0	0	0	0
G	146	1	1	0	428	135	931	2	0	0	0	0	0	0	0	0	0
A	148	0	0	0	144	1285	4	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	209	1429	3	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	677	9	2095	15	0	0	0	0	0	0	0	0	0
C	154	0	0	0	255	2295	25	0	0	0	0	0	0	0	0	0	0

	A	156	0	0	0	416	2572	44	0	0	0	0	0	0	0	0	0	0
	G	158	1	1	0	1930	101	4402	21	0	0	0	0	0	0	0	0	0
	A	160	0	0	0	317	4890	59	1	0	0	0	0	0	0	0	0	0
	A	162	1	1	0	645	54	5583	19	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	5	5874	8	0	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	34	5871	12	3	0	0	0	0	0	0	0	0	0
	A	168	1	1	0	126	119	5814	47	2	0	0	0	0	0	0	0	0
ES39i	C	170	0	1	1	112	5972	70	10	1	0	0	0	0	0	0	0	0
	A	172	0	0	0	113	6148	10	0	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	10	6201	18	6	47	5	0	0	0	0	0	0	0
ES37i	G	176	0	1	1	24	6174	78	26	3	1	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	123	6201	10	93	6	0	1	0	0	0	0	0	0
ES35i	G	180	0	5	5	145	6302	24	15	1	37	54	0	0	0	0	0	0
	G	182	0	0	0	5	6571	1	0	0	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	74	6451	111	18	0	0	0	0	0	0	0	0	0

	G	186	0	0	0	174	6647	5	0	0	0	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	176	6695	2	1	0	2	5	12	82	35	2	0	0
	C	190	0	0	0	218	6990	8	1	0	3	0	0	0	0	0	0	0
	A	192	0	0	0	241	7170	49	1	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	190	7248	17	27	170	6	0	0	0	0	0	0	0
ES31i	A	196	0	1	1	210	7420	205	23	3	0	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	253	7639	126	97	1	0	0	0	0	0	0	0	0
	G	200	0	0	0	234	8104	9	0	0	0	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	188	7938	71	85	10	207	45	1	0	0	0	0	0
	C	204	0	0	0	274	8485	40	4	10	0	0	0	0	0	0	0	0
	A	206	0	0	0	342	8771	15	6	5	12	2	0	0	0	0	0	0
ES28i	G	208	0	3	3	276	8750	37	65	266	72	0	0	0	0	0	0	0
ES27d	G	210	1	0	-1	836	526	9296	29	4	0	0	0	0	0	0	0	0
	C	212	1	1	0	768	294	10167	93	1	2	0	0	0	1	1	0	0
	C	214	0	0	0	26	11006	26	0	4	0	0	0	0	0	0	0	0

	A	216	0	0	0	333	10880	178	0	0	0	0	0	0	0	0	0	0
	C	218	0	0	0	341	11083	284	11	18	0	0	0	0	0	0	0	0
	A	220	0	0	0	405	11663	39	28	2	3	0	0	0	0	0	0	0
	G	222	0	0	0	369	11925	108	59	44	5	1	0	0	0	0	0	0
	C	224	0	0	0	748	12096	348	15	65	2	0	0	0	0	0	0	0
	A	226	0	0	0	1280	13150	100	7	0	0	0	0	0	0	0	0	0
ES26i	C	228	0	1	1	459	13678	864	18	2	1	0	0	0	0	0	0	0
	C	230	0	0	0	18	14985	6	2	0	0	0	0	0	0	0	0	0
	C	232	0	0	0	19	14827	180	6	0	0	0	0	0	0	0	0	0
	G	234	0	0	0	105	14988	35	7	0	0	0	0	0	0	0	0	0
ES25d	C	236	3	0	-3	556	1206	326	192	13293	595	3	2	0	0	0	0	0
	A	238	0	0	0	731	15395	312	22	4	1	0	0	0	0	0	0	0
	G	240	0	0	0	423	15356	803	215	54	18	2	8	2	2	0	0	0
	C	242	0	0	0	863	16686	109	9	13	22	14	2	0	0	0	0	0
	A	244	0	0	0	1561	17406	333	38	4	1	0	0	0	0	0	0	0

	C	246	0	0	0	278	18995	205	38	25	9	6	3	0	0	0	0	0
	A	248	0	0	0	605	19397	111	42	5	1	2	0	0	0	0	0	0
ES24i	G	250	0	1	1	398	17810	2147	166	42	21	39	7	4	0	0	0	0
ES23d	G	252	2	0	-2	361	2625	434	17255	485	33	18	3	2	0	0	0	0
Es22i	G	254	0	1	1	292	17779	2727	294	121	29	5	0	1	0	0	0	0
Es21i	A	256	0	4	4	278	17634	500	151	142	1548	1345	33	5	1	1	0	0
ES20i	G	258	0	1	1	737	17556	3444	267	125	249	75	1	0	0	0	0	0
ES19i	G	260	0	2	2	190	17829	116	4173	103	128	23	2	0	0	0	0	0
ES18i	A	262	0	3	3	229	17796	44	344	3377	1018	2	3	1	0	0	0	0
ES17i	G	264	0	2	2	593	17820	457	4369	131	1	10	0	0	0	0	0	0
ES16i	A	266	0	1	1	456	18554	4582	103	15	1	4	0	0	0	0	0	0
ES15i	G	268	0	5	5	554	18584	97	387	32	441	3779	439	3	0	0	0	0
ES14i	G	270	0	1	1	429	19153	4710	83	369	9	4	2	0	0	0	0	0
	G	272	0	0	0	42	24604	63	5	3	0	0	0	0	0	0	0	0
ES13i	G	274	0	1	1	153	19238	5469	24	2	0	2	0	0	0	0	0	0

	A	276	0	0	0	175	24807	58	2	1	0	0	0	0	0	0	0	0
ES12d	A	278	1	0	-1	1080	5942	19351	278	20	1	0	0	0	0	0	0	0
ES11i	A	280	0	3	3	135	19901	649	271	5041	338	0	0	0	0	0	0	0
ES10i	G	282	0	2	2	414	19817	723	5814	87	3	0	0	0	0	0	0	0
ES9i	A	284	0	3	3	707	20393	110	152	5805	399	4	1	0	0	0	0	0
ES8d	A	286	3	1	-2	1163	90	6533	490	20660	407	9	0	0	0	0	0	0
	G	288	2	2	0	1061	43	249	28833	310	11	27	3	0	0	0	0	0
	A	290	0	0	0	700	29372	92	82	73	2	10	1	0	0	0	0	0
ES7i	G	292	2	1	-1	1170	177	22903	7808	37	3	16	0	0	0	0	0	0
ES6i	A	294	1	2	1	817	23490	8018	128	8	44	7	0	0	0	0	0	0
ES5i	A	296	0	2	2	564	23264	729	8465	206	29	2	0	0	0	0	0	0
ES4i	A	298	0	1	1	736	23009	9504	709	93	0	0	2	3	0	0	0	0
ES3d	G	300	3	0	-3	691	10011	1143	170	21681	813	2	0	0	0	0	0	0
ES2d	G	302	1	0	-1	1433	11575	23175	131	8	0	0	0	0	0	0	0	0
ES1i	A	304	0	2	2	1432	22978	1464	10382	1172	40	2	0	0	0	0	0	0

G	306	3	3	0	1238	0	6	430	34820	2995	110	2	7	5	0	0	0
C	308	0	0	0	1067	38254	135	0	0	0	0	0	0	0	0	0	0
A	310	0	0	0	2041	39446	1	0	0	0	0	0	0	0	0	0	0
C	312	1	1	0	1710	421	41661	132	0	0	0	0	0	0	0	0	0
G	314	0	0	0	1820	43105	134	4	0	0	0	0	0	0	0	0	0
G	316	1	1	0	3364	61	45954	420	0	0	0	0	0	0	0	0	1
G	318	0	0	0	106	48353	43	2	0	0	0	0	0	0	0	0	0
A	320	1	1	0	1640	267	48469	225	0	0	0	0	0	0	0	0	0
C	322	0	0	0	1412	50127	12	0	0	0	0	0	0	0	0	0	0
A	324	0	0	0	2331	51337	218	0	0	0	0	0	0	0	0	0	0
G	326	0	0	0	2088	53730	156	0	0	0	0	0	0	0	0	0	0
A	328	0	0	0	967	55805	158	0	0	0	0	0	0	0	0	0	0
A	330	0	0	0	105	56924	2	1	0	0	0	0	0	0	0	0	0
A	332	0	0	0	98	57029	3	0	0	0	0	0	0	0	0	0	0
A	334	0	0	0	54	57128	0	0	0	0	0	0	0	0	0	0	0

G	336	0	0	0	178	56753	430	5	0	0	0	0	0	0	0	0	0
A	338	2	2	0	321	10	1067	56228	274	0	0	0	0	0	0	0	0
G	340	1	1	0	603	544	57044	422	1	0	0	0	0	0	0	0	0
G	342	1	1	0	957	15	58416	249	2	0	0	0	0	0	0	0	0
A	344	0	0	0	623	59191	68	8	0	0	0	0	0	0	0	0	0
A	346	0	0	0	207	59834	71	2	0	0	0	0	0	0	0	0	0
A	348	1	1	0	1058	182	59575	581	0	0	0	0	0	0	0	0	0
A	350	1	1	0	2130	320	61601	297	0	0	0	0	0	0	0	0	0
A	352	0	0	0	274	62958	430	13	0	0	0	0	0	0	0	0	0
A	354	0	0	0	186	63427	153	34	0	0	0	0	0	0	0	0	0
A	356	0	0	0	108	63664	92	11	0	0	0	0	0	0	0	0	0
G	358	0	0	0	181	63742	111	13	1	0	2	0	0	1	0	0	0
G	360	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	362	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	364	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0

	A	366	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
	A	368	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
	C	370	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
	A	372	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
	A	374	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
	G	376	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	378	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	380	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	382	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0
3' end	G	384	0	0	0	0	64008	0	0	0	0	0	0	0	0	0	0	0

Category 2

ND7 – Ion Torrent - REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0													
	A	110	1	1	0													
	A	112	2	2	0													
	A	114	0	0	0													
	A	116	3	3	0													
	A	118	1	1	0	6	6	0	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	69	1	68	0	0	0	0	0	0	0	0	0	0
	A	122	0	0	0	140	138	2	0	0	0	0	0	0	0	0	0	0
	A	124	1	1	0	386	9	367	10	0	0	0	0	0	0	0	0	0

C	126	0	0	0	508	495	13	0	0	0	0	0	0	0	0	0	0
A	128	0	0	0	649	640	9	0	0	0	0	0	0	0	0	0	0
A	130	0	0	0	801	800	1	0	0	0	0	0	0	0	0	0	0
A	132	0	0	0	825	824	1	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	828	827	1	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	830	829	1	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	831	831	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	835	835	0	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	835	834	1	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	908	836	71	1	0	0	0	0	0	0	0	0	0
G	146	1	1	0	1443	112	1326	5	0	0	0	0	0	0	0	0	0
A	148	0	0	0	1658	1653	5	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	1863	1850	13	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	2688	18	2654	16	0	0	0	0	0	0	0	0	0
C	154	0	0	0	2992	2960	32	0	0	0	0	0	0	0	0	0	0

	A	156	0	0	0	3390	3338	52	0	0	0	0	0	0	0	0	0	0
	G	158	1	1	0	5607	108	5471	28	0	0	0	0	0	0	0	0	0
	A	160	0	0	0	6073	6018	54	1	0	0	0	0	0	0	0	0	0
	A	162	1	1	0	6442	10	6402	30	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	6503	6494	9	0	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	6506	6493	13	0	0	0	0	0	0	0	0	0	0
	A	168	1	1	0	6560	118	6382	59	1	0	0	0	0	0	0	0	0
ES39i	C	170	0	1	1	6632	6566	66	0	0	0	0	0	0	0	0	0	0
	A	172	0	0	0	6791	6789	2	0	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	6977	6974	1	1	1	0	0	0	0	0	0	0	0
ES37i	G	176	0	1	1	6982	6926	55	1	0	0	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	7021	7010	5	6	0	0	0	0	0	0	0	0	0
ES35i	G	180	0	5	5	7195	7148	41	0	0	2	4	0	0	0	0	0	0
	G	182	0	0	0	7371	7371	0	0	0	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	7376	7346	28	2	0	0	0	0	0	0	0	0	0

	G	186	0	0	0	7447	7444	3	0	0	0	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	7715	7707	1	0	1	0	0	2	3	1	0	0	0
	C	190	0	0	0	7931	7929	2	0	0	0	0	0	0	0	0	0	0
	A	192	0	0	0	8202	8130	72	0	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	8546	8529	6	1	10	0	0	0	0	0	0	0	0
ES31i	A	196	0	1	1	8823	8796	21	6	0	0	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	9113	9098	11	4	0	0	0	0	0	0	0	0	0
	G	200	0	0	0	9445	9434	11	0	0	0	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	9758	9693	46	3	2	13	1	0	0	0	0	0	0
	C	204	0	0	0	10048	10000	44	0	4	0	0	0	0	0	0	0	0
	A	206	0	0	0	10397	10387	7	0	1	1	1	0	0	0	0	0	0
ES28i	G	208	0	3	3	10945	10886	34	6	15	4	0	0	0	0	0	0	0
ES27d	G	210	1	0	-1	11885	51	11795	32	6	0	1	0	0	0	0	0	0
	C	212	1	1	0	12741	51	12522	166	0	2	0	0	0	0	0	0	0
	C	214	0	0	0	13348	13326	21	1	0	0	0	0	0	0	0	0	0

	A	216	0	0	0	13380	13218	162	0	0	0	0	0	0	0	0	0	0
	C	218	0	0	0	13835	13815	19	0	1	0	0	0	0	0	0	0	0
	A	220	0	0	0	14254	14210	40	4	0	0	0	0	0	0	0	0	0
	G	222	0	0	0	14836	14753	66	14	2	1	0	0	0	0	0	0	0
	C	224	0	0	0	15327	15065	247	7	7	1	0	0	0	0	0	0	0
	A	226	0	0	0	16337	16327	10	0	0	0	0	0	0	0	0	0	0
ES26i	C	228	0	1	1	17922	17819	100	2	0	0	1	0	0	0	0	0	0
	C	230	0	0	0	18443	18441	2	0	0	0	0	0	0	0	0	0	0
	C	232	0	0	0	18468	18420	42	6	0	0	0	0	0	0	0	0	0
	G	234	0	0	0	18477	18440	37	0	0	0	0	0	0	0	0	0	0
ES25d	C	236	3	0	-3	19223	132	54	144	17997	894	2	0	0	0	0	0	0
	A	238	0	0	0	19273	19178	79	4	11	1	0	0	0	0	0	0	0
	G	240	0	0	0	19893	19688	142	43	17	2	0	1	0	0	0	0	0
	C	242	0	0	0	20401	20359	18	9	4	6	3	2	0	0	0	0	0
	A	244	0	0	0	20619	20226	389	3	1	0	0	0	0	0	0	0	0

	C	246	0	0	0	20981	20921	42	11	6	1	0	0	0	0	0	0	0
	A	248	0	0	0	21262	21191	64	6	1	0	0	0	0	0	0	0	0
ES24i	G	250	0	1	1	21863	21413	380	47	9	2	6	3	3	0	0	0	0
ES23d	G	252	2	0	-2	22817	427	201	21448	726	12	2	0	1	0	0	0	0
Es22i	G	254	0	1	1	22917	22346	478	58	24	8	2	1	0	0	0	0	0
Es21i	A	256	0	4	4	23041	22263	124	31	69	338	213	1	2	0	0	0	0
ES20i	G	258	0	1	1	23241	22156	873	63	43	85	15	5	1	0	0	0	0
ES19i	G	260	0	2	2	23854	22782	23	990	20	28	10	1	0	0	0	0	0
ES18i	A	262	0	3	3	23894	22757	25	131	843	138	0	0	0	0	0	0	0
ES17i	G	264	0	2	2	23966	22770	182	959	38	7	8	2	0	0	0	0	0
ES16i	A	266	0	1	1	24647	23496	1091	46	8	2	4	0	0	0	0	0	0
ES15i	G	268	0	5	5	25335	24058	51	144	14	123	875	69	0	0	0	1	0
ES14i	G	270	0	1	1	26138	24843	1125	19	138	5	4	4	0	0	0	0	0
	G	272	0	0	0	26366	26300	59	5	0	1	0	0	1	0	0	0	0
ES13i	G	274	0	1	1	26394	24888	1489	17	0	0	0	0	0	0	0	0	0

	A	276	0	0	0	26493	26446	40	7	0	0	0	0	0	0	0	0	0
ES12d	A	278	1	0	-1	27442	1788	25285	357	9	3	0	0	0	0	0	0	0
ES11i	A	280	0	3	3	28281	26244	412	112	1429	83	1	0	0	0	0	0	0
ES10i	G	282	0	2	2	28435	26149	422	1820	39	5	0	0	0	0	0	0	0
ES9i	A	284	0	3	3	28896	26564	96	221	1900	107	7	1	0	0	0	0	0
ES8d	A	286	3	1	-2	30747	99	2256	545	26967	820	54	6	0	0	0	0	0
	G	288	2	2	0	32227	52	291	31498	324	24	30	8	0	0	0	0	0
	A	290	0	0	0	32509	32129	174	103	88	6	9	0	0	0	0	0	0
ES7i	G	292	2	1	-1	34235	175	30374	3635	28	8	11	4	0	0	0	0	0
ES6i	A	294	1	2	1	35264	31328	3731	89	30	82	4	0	0	0	0	0	0
ES5i	A	296	0	2	2	36524	31382	649	4291	179	21	2	0	0	0	0	0	0
ES4i	A	298	0	1	1	37303	31159	5236	765	135	8	0	0	0	0	0	0	0
ES3d	G	300	3	0	-3	37981	5802	1745	677	28604	1074	40	28	9	2	0	0	0
ES2d	G	302	1	0	-1	39484	8470	30562	411	27	7	7	0	0	0	0	0	0
ES1i	A	304	0	2	2	41205	30147	1829	7846	1340	41	2	0	0	0	0	0	0

G	306	3	3	0	44534	0	12	515	41622	2298	62	14	5	1	4	1	0
C	308	0	0	0	44575	44360	214	0	1	0	0	0	0	0	0	0	0
A	310	0	0	0	46130	46126	3	1	0	0	0	0	0	0	0	0	0
C	312	1	1	0	49747	486	49052	209	0	0	0	0	0	0	0	0	0
G	314	0	0	0	51147	50869	274	4	0	0	0	0	0	0	0	0	0
G	316	1	1	0	55504	77	54752	675	0	0	0	0	0	0	0	0	0
G	318	0	0	0	58383	58280	102	1	0	0	0	0	0	0	0	0	0
A	320	1	1	0	59270	283	58551	436	0	0	0	0	0	0	0	0	0
C	322	0	0	0	60957	60916	41	0	0	0	0	0	0	0	0	0	0
A	324	0	0	0	63266	62949	317	0	0	0	0	0	0	0	0	0	0
G	326	0	0	0	65923	65656	267	0	0	0	0	0	0	0	0	0	0
A	328	0	0	0	66269	65995	274	0	0	0	0	0	0	0	0	0	0
A	330	0	0	0	66493	66492	1	0	0	0	0	0	0	0	0	0	0
A	332	0	0	0	66547	66546	1	0	0	0	0	0	0	0	0	0	0
A	334	0	0	0	66729	66728	1	0	0	0	0	0	0	0	0	0	0

G	336	0	0	0	66807	66155	649	3	0	0	0	0	0	0	0	0	0
A	338	2	2	0	67388	2	1203	65811	372	0	0	0	0	0	0	0	0
G	340	1	1	0	68469	479	67354	635	1	0	0	0	0	0	0	0	0
G	342	1	1	0	70008	25	69662	321	0	0	0	0	0	0	0	0	0
A	344	0	0	0	71159	71063	94	2	0	0	0	0	0	0	0	0	0
A	346	0	0	0	72227	72165	56	6	0	0	0	0	0	0	0	0	0
A	348	1	1	0	73118	219	72156	743	0	0	0	0	0	0	0	0	0
A	350	1	1	0	76503	339	75835	329	0	0	0	0	0	0	0	0	0
A	352	0	0	0	78525	77736	769	19	1	0	0	0	0	0	0	0	0
A	354	0	0	0	78831	78701	110	18	1	0	1	0	0	0	0	0	0
A	356	0	0	0	78977	78897	71	8	1	0	0	0	0	0	0	0	0
G	358	0	0	0	79098	78981	88	29	0	0	0	0	0	0	0	0	0
G	360	1	1	0	79251	0	0	0	0	0	0	0	0	0	0	0	0
A	362	1	1	0	79251	0	0	0	0	0	0	0	0	0	0	0	0
G	364	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0

	A	366	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
	A	368	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
	C	370	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
	A	372	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
	A	374	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
	G	376	1	1	0	79251	0	0	0	0	0	0	0	0	0	0	0	0
	C	378	1	1	0	79251	0	0	0	0	0	0	0	0	0	0	0	0
	C	380	2	2	0	79251	0	0	0	0	0	0	0	0	0	0	0	0
	C	382	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0
3' end	G	384	0	0	0	79251	79251	0	0	0	0	0	0	0	0	0	0	0

Appendix 6

The following tables tabulated overleaf contain MiSeq data that has been subjected to the pipeline described in Chapter 3.

The most 5' and 3' end of ND7 and RPS12 and the edited region are highlighted in yellow. The 3' primer sequences used in the sequencing reaction is marked in red. Deletion (ESd) and insertion (ESi) editing sites are indicated in blue and pink respectively. The expected number of uridylyls in fully/correctly editied ND7 at each edited site is highlighted in turquoise.

There are no Category 2 reads for this data set due to the length of the fully edited sequences, so the following two tables are for +REL1 and –REL1 samples.

ND7 MiSeq +REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	A	110	1	1	0	8	1	0	0	0	0	0	0	0	0	0	0	0
	A	112	2	2	0	19	9	0	19	0	0	0	0	0	0	0	0	0
	A	114	0	0	0	50	28	0	0	0	0	0	0	0	0	0	0	0
	A	116	3	3	0	1021	59	0	316	77	0	0	0	0	0	0	0	0
	A	118	1	1	0	908	139	1015	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	47433	7	48864	0	0	0	0	0	0	0	0	0	0
	A	122	0	0	0	32761	49440	0	8	0	0	0	0	0	0	0	0	0
	A	124	1	1	0	28359	79	84458	14	0	0	0	0	0	0	0	0	0
	C	126	0	0	0	2439	110535	5	27	0	0	0	0	0	0	0	0	0
	A	128	0	0	0	317	112999	6	0	0	0	0	0	0	0	0	0	0

A	130	0	0	0	0	113316	0	0	0	0	0	0	0	0	0	0	0
A	132	0	0	0	17	113316	0	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	4	113333	0	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	0	113337	0	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	21	113337	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	0	113358	0	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	0	113358	0	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	65	113358	0	0	0	0	0	0	0	0	0	0	0
G	146	1	1	0	0	0	113423	0	0	0	0	0	0	0	0	0	0
A	148	0	0	0	0	113423	0	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	33	113423	0	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	0	1	113455	0	0	0	0	0	0	0	0	0	0
C	154	0	0	0	0	113456	0	0	0	0	0	0	0	0	0	0	0
A	156	0	0	0	0	113456	0	0	0	0	0	0	0	0	0	0	0
G	158	1	1	0	0	0	113456	0	0	0	0	0	0	0	0	0	0

	A	160	0	0	0	0	113456	0	0	0	0	0	0	0	0	0	0	0
	A	162	1	1	0	0	154	113261	41	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	0	112511	621	324	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	0	104895	3731	4825	5	0	0	0	0	0	0	0	0
	A	168	1	1	0	127	284	106812	6194	155	117	21	0	0	0	0	0	0
ES39i	C	170	0	1	1	18	87993	25374	108	108	0	0	0	0	0	0	0	0
	A	172	0	0	0	0	113557	44	0	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	0	81331	7233	1790	23125	109	1	12	0	0	0	0	0
ES37i	G	176	0	1	1	63	77867	28996	6794	7	0	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	111	77026	97	36569	19	2	0	0	1	0	0	0	0
ES35i	G	180	0	5	5	157	77024	132	7606	295	720	27762	337	0	0	0	0	0
	G	182	0	0	0	1	113832	89	11	1	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	0	76807	37102	24	0	0	0	0	0	0	0	0	0
	G	186	0	0	0	0	113411	439	30	46	7	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	112	76254	258	67	370	148	80	202	32649	3660	341	5	9

	C	190	0	0	0	1	113685	233	23	46	58	0	0	0	0	0	0	0
	A	192	0	0	0	14	113933	70	43	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	0	73478	740	1526	38275	0	0	41	0	0	0	0	0
ES31i	A	196	0	1	1	0	73012	40495	520	0	33	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	12	73330	40428	264	0	11	34	0	0	0	0	0	0
	G	200	0	0	0	0	114044	17	0	11	0	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	119	63289	2692	6847	667	40427	18	39	0	6	94	50	1
	C	204	0	0	0	0	112723	772	238	332	126	0	0	0	0	0	0	0
	A	206	0	0	0	30	113217	613	104	15	219	23	0	0	0	0	0	0
ES28i	G	208	0	3	3	23	69674	851	3159	40394	127	19	0	2	9	4	5	0
ES27d	G	210	1	0	-1	11	49609	63457	694	269	193	0	22	0	0	0	0	0
	C	212	1	1	0	0	18342	95035	244	127	319	17	0	41	14	25	53	20
	C	214	0	0	0	0	113049	1107	56	43	0	0	0	0	0	0	0	0
	A	216	0	0	0	0	113732	400	99	24	0	0	0	0	0	0	0	0
	C	218	0	0	0	0	94394	18085	432	1185	159	0	0	0	0	0	0	0

	A	220	0	0	0	6	110354	1216	2662	23	0	0	0	0	0	0	0	0
	G	222	0	0	0	0	103555	3914	4085	2282	354	17	0	54	0	0	0	0
	C	224	0	0	0	36	96781	12652	389	4372	59	15	0	0	0	0	0	0
	A	226	0	0	0	28	106812	7093	309	83	0	0	0	0	0	0	0	0
ES26i	C	228	0	1	1	0	56849	56880	241	251	104	0	0	0	0	0	0	0
	C	230	0	0	0	1	113798	512	15	0	0	0	0	0	0	0	0	0
	C	232	0	0	0	9	106792	7416	127	0	0	0	0	0	0	0	0	0
	G	234	0	0	0	45	113637	530	161	7	0	0	0	0	0	0	0	0
ES25d	C	236	3	0	-3	94	71233	15211	4886	22593	56	457	0	0	0	0	0	0
	A	238	0	0	0	89	100344	13587	382	159	2	0	0	0	0	0	0	0
	G	240	0	0	0	203	68481	33559	9497	1214	1029	212	559	0	5	28	7	28
	C	242	0	0	0	124	108872	2222	494	1270	863	1045	61	0	0	2	9	0
	A	244	0	0	0	18	112874	910	840	226	4	0	36	0	0	0	0	0
	C	246	0	0	0	101	107464	4245	1236	264	562	1045	31	0	40	15	5	8
	A	248	0	0	0	87	111695	2517	603	61	108	0	25	0	0	0	0	0

ES24i	G	250	0	1	1	51	32609	75627	2676	2203	304	1340	360	0	0	0	0	0
ES23d	G	252	2	0	-2	898	81320	8406	19981	2559	2627	216	56	273	4	0	0	0
Es22i	G	254	0	1	1	388	20911	89331	2788	2797	274	149	0	0	5	25	0	0
Es21i	A	256	0	4	4	422	17621	14016	2659	3094	77476	719	1072	51	0	0	0	2
ES20i	G	258	0	1	1	262	15922	86259	3195	2606	8966	36	108	0	0	0	0	0
ES19i	G	260	0	2	2	106	13655	199	102808	100	313	62	0	0	0	0	0	0
ES18i	A	262	0	3	3	90	13309	537	1041	102282	0	67	6	0	0	0	0	0
ES17i	G	264	0	2	2	40	12654	1511	102538	255	4	364	0	0	0	0	0	0
ES16i	A	266	0	1	1	87	13616	102403	503	758	24	106	0	0	0	0	0	0
ES15i	G	268	0	5	5	40	12862	1281	943	284	129	101721	235	16	1	4	0	0
ES14i	G	270	0	1	1	16	13989	102464	155	793	9	70	0	0	0	0	0	0
	G	272	0	0	0	1	116167	1070	223	0	0	20	16	0	0	0	0	0
ES13i	G	274	0	1	1	104	13815	103611	72	0	0	0	0	0	0	0	0	0
	A	276	0	0	0	15	117225	352	24	0	0	0	0	0	0	0	0	0
ES12d	A	278	1	0	-1	58	105264	11335	289	539	194	0	0	0	0	0	0	0

ES11i	A	280	0	3	3	133	7596	3675	1965	104496	40	0	0	0	0	0	0	0
ES10i	G	282	0	2	2	111	7050	1025	109405	313	0	2	32	0	0	0	0	0
ES9i	A	284	0	3	3	188	7189	125	1834	108660	48	50	94	0	0	0	0	0
ES8d	A	286	3	1	-2	146	638	108612	32	8433	48	394	0	0	0	0	0	0
	G	288	2	2	0	140	356	1591	115022	133	74	1153	26	0	0	0	0	0
	A	290	0	0	0	31	115260	203	1591	1318	15	5	0	12	0	0	0	0
ES7i	G	292	1	2	1	58	1817	6969	109621	46	2	6	0	0	0	0	0	0
ES6i	A	294	0	1	1	91	6501	110040	169	308	1494	6	0	17	0	0	0	0
ES5i	A	296	0	2	2	184	4614	2045	111981	17	6	0	0	0	0	0	0	0
ES4i	A	298	0	1	1	165	4058	114007	726	0	17	0	0	0	0	0	0	0
ES3d	G	300	3	0	-3	38533	115126	38742	1847	1277	1	1	0	0	0	0	0	0
ES2d	G	302	1	0	-1	0	157454	0	0	0	0	0	0	0	0	0	0	0
ES1i	A	304	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	306	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	308	0	0	0	0	157454	0	0	0	0	0	0	0	0	0	0	0

A	310	0	0	0	0	157454	0	0	0	0	0	0	0	0	0	0	0
C	312	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	314	0	0	0	0	157454	0	0	0	0	0	0	0	0	0	0	0
G	316	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	318	0	0	0	0	157454	0	0	0	0	0	0	0	0	0	0	0
A	320	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ND7 MiSeq -REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	108	1	1	0													
	A	110	1	1	0													
	A	112	2	2	0	15	0	0	15	0	0	0	0	0	0	0	0	0
	A	114	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
	A	116	3	3	0	189	0	0	30	15	0	0	0	0	0	0	0	0
	A	118	1	1	0	467	0	204	0	0	0	0	0	0	0	0	0	0
5' end	G	120	1	1	0	17849	1	18333	0	0	0	0	0	0	0	0	0	0
	A	122	0	0	0	14755	18520	1	4	0	0	0	0	0	0	0	0	0
	A	124	1	1	0	15327	127	33670	10	0	0	0	0	0	0	0	0	0
	C	126	0	0	0	762	48597	8	5	0	0	0	0	0	0	0	0	0
	A	128	0	0	0	114	49362	3	0	0	0	0	0	0	0	0	0	0
	A	130	0	0	0	1	49478	0	0	0	0	0	0	0	0	0	0	0

A	132	0	0	0	15	49479	0	0	0	0	0	0	0	0	0	0	0
A	134	0	0	0	0	49494	0	0	0	0	0	0	0	0	0	0	0
A	136	0	0	0	0	49494	0	0	0	0	0	0	0	0	0	0	0
A	138	0	0	0	0	49494	0	0	0	0	0	0	0	0	0	0	0
A	140	0	0	0	2	49494	0	0	0	0	0	0	0	0	0	0	0
C	142	0	0	0	12	49495	1	0	0	0	0	0	0	0	0	0	0
A	144	0	0	0	0	49508	0	0	0	0	0	0	0	0	0	0	0
G	146	1	1	0	0	0	49508	0	0	0	0	0	0	0	0	0	0
A	148	0	0	0	0	49508	0	0	0	0	0	0	0	0	0	0	0
C	150	0	0	0	41	49508	0	0	0	0	0	0	0	0	0	0	0
A	152	1	1	0	1	3	49546	1	0	0	0	0	0	0	0	0	0
C	154	0	0	0	0	49550	0	0	0	0	0	0	0	0	0	0	0
A	156	0	0	0	78	49550	0	0	0	0	0	0	0	0	0	0	0
G	158	1	1	0	0	0	49628	0	0	0	0	0	0	0	0	0	0
A	160	0	0	0	0	49628	0	0	0	0	0	0	0	0	0	0	0

	A	162	1	1	0	0	0	49628	0	0	0	0	0	0	0	0	0	0
	A	164	0	0	0	0	49321	263	44	0	0	0	0	0	0	0	0	0
	G	166	0	0	0	119	48081	824	780	0	0	0	59	3	0	0	0	0
	A	168	1	1	0	0	42	48413	1214	0	16	62	0	0	0	0	0	0
ES39i	C	170	0	1	1	0	45028	4679	0	40	0	0	0	0	0	0	0	0
	A	172	0	0	0	0	49747	0	0	0	0	0	0	0	0	0	0	0
ES38i	A	174	0	4	4	0	43714	1523	247	4259	4	0	0	0	0	0	0	0
ES37i	G	176	0	1	1	42	42804	5528	1366	91	0	0	0	0	0	0	0	0
ES36i	A	178	0	2	2	0	42520	97	7155	0	0	0	0	17	0	0	0	0
ES35i	G	180	0	5	5	26	42460	429	1757	0	158	5010	1	0	0	0	0	0
	G	182	0	0	0	0	49798	0	0	17	0	0	0	0	0	0	0	0
ES34i	A	184	0	1	1	0	42460	7355	0	0	0	0	0	0	0	0	0	0
	G	186	0	0	0	11	49596	213	17	0	0	0	0	0	0	0	0	0
ES33i	A	188	0	7	7	0	42276	0	9	170	50	29	38	6749	440	63	2	0
	C	190	0	0	0	74	49801	13	12	0	13	0	0	0	0	0	0	0

	A	192	0	0	0	24	49867	0	33	0	0	0	0	0	0	0	0	0
ES32i	G	194	0	3	3	9	41675	155	484	7619	0	0	0	0	0	0	0	0
ES31i	A	196	0	1	1	0	41680	8161	92	0	0	0	0	0	0	0	0	0
ES30i	C	198	0	1	1	0	41791	8082	16	44	0	0	0	0	0	0	0	0
	G	200	0	0	0	0	49917	16	0	0	0	0	0	0	0	0	0	0
ES29i	A	202	0	4	4	0	40391	376	845	130	8178	0	13	0	0	0	0	0
	C	204	0	0	0	0	49734	3	128	68	0	0	0	0	0	0	0	0
	A	206	0	0	0	0	49695	226	0	0	12	0	0	0	0	0	0	0
ES28i	G	208	0	3	3	0	40397	219	1043	8194	77	0	0	1	2	0	0	0
ES27d	G	210	1	0	-1	41	10113	39393	320	134	0	0	0	0	0	0	0	0
	C	212	1	1	0	50	3073	46674	71	75	62	19	0	0	0	0	0	
	C	214	0	0	0	23	49562	438	24	0	0	0	0	0	0	0	0	
	A	216	0	0	0	8	49849	162	44	0	0	0	0	0	0	0	0	
	C	218	0	0	0	0	47055	2608	87	305	0	0	0	0	0	0	0	
	A	220	0	0	0	0	48984	606	402	54	9	0	0	0	0	0	0	
	G	222	0	0	0	0	48063	719	666	442	165	0	0	0	0	0	0	

	C	224	0	0	0	41	45843	3021	88	882	48	98	22	61	0	0	0
	A	226	0	0	0	32	48874	1092	66	18	61	0	0	0	0	0	0
ES																	
26i	C	228	0	1	1	0	39225	10604	216	73	0	10	0	0	0	0	0
	C	230	0	0	0	14	49942	178	8	0	0	0	0	0	0	0	0
	C	232	0	0	0	163	46381	3721	73	14	0	0	0	0	0	0	0
	G	234	0	0	0	29	50162	139	22	0	0	0	0	0	0	0	0
ES																	
25										2802							
d	C	236	3	0	-3	835	16487	4199	1683	0	199	133	0	0	0	0	1
	A	238	0	0	0	515	45378	5426	282	286	10	0	0	0	0	0	0
	G	240	0	0	0	741	35871	11612	3101	1139	326	52	145	0	0	0	0
	C	242	0	0	0	231	49602	1216	265	556	566	204	120	0	0	8	0
	A	244	0	0	0	62	51880	430	308	41	0	11	0	0	0	0	0
	C	246	0	0	0	415	49309	1818	549	242	453	415	10	0	1	31	15
	A	248	0	0	0	78	51536	1211	305	24	68	0	0	0	0	0	0
ES																	
24i	G	250	0	1	1	231	32066	18035	1131	1098	283	565	78	0	0	0	0
ES																	
23																	
d	G	252	2	0	-2	1857	21612	4368	25045	1440	1191	80	75	237	44	19	27
Es2																	
2i	G	254	0	1	1	1429	26215	26728	1545	1702	84	136	3	0	0	0	0

Es2																	
1i	A	256	0	4	4	1085	23870	8839	1164	1936	21119	438	385	0	0	0	0
ES																	
20i	G	258	0	1	1	852	23296	25665	1079	2381	5835	149	22	163	1	0	0
ES																	
19i	G	260	0	2	2	88	21622	162	36543	112	191	0	4	1	96	0	0
ES										3627							
18i	A	262	0	3	3	118	21129	329	1011	8	15	0	48	14	0	0	0
ES																	
17i	G	264	0	2	2	101	20379	1242	36942	219	0	151	22	8	0	0	0
ES																	
16i	A	266	0	1	1	388	21008	37105	251	794	62	98	0	31	0	0	0
ES																	
15i	G	268	0	5	5	196	20012	1192	852	226	57	####	132	43	0	0	0
ES																	
14i	G	270	0	1	1	92	21036	37453	64	675	148	221	19	19	0	0	0
	G	272	0	0	0	56	58556	821	217	9	6	0	44	0	0	0	0
ES																	
13i	G	274	0	1	1	37	20399	39253	24	7	23	0	0	0	0	0	0
	A	276	0	0	0	5	59510	186	45	0	0	0	0	0	0	0	0
ES																	
12																	
d	A	278	1	0	-1	107	40974	18007	28	666	82	0	0	0	0	0	0
ES										4007							
11i	A	280	0	3	3	224	14824	3829	1224	0	29	0	0	0	0	0	0
ES																	
10i	G	282	0	2	2	218	14064	1360	44545	182	28	0	14	0	0	0	0

ES										4399							
9i	A	284	0	3	3	403	14049	131	2146	6	140	97	11	19	0	0	0
ES										1497							
8d	A	286	3	1	-2	1975	652	45236	344	2	400	675	63	15	0	0	0
	G	288	2	2	0	282	347	1682	59797	67	242	605	130	1	15	0	0
	A	290	0	0	0	410	60113	127	1167	1391	19	87	36	11	65	7	0
ES																	
7i	G	292	1	2	1	100	1239	13020	49011	94	1	39	0	0	0	0	0
ES																	
6i	A	294	0	1	1	357	12638	49894	69	95	1036	26	0	0	0	0	0
ES																	
5i	A	296	0	2	2	134	10667	1959	51190	49	68	0	0	0	0	0	0
ES																	
4i	A	298	0	1	1	158	9959	52815	1235	26	10	0	0	0	0	0	0
ES																	
3d	G	300	3	0	-3	6296	53654	8835	5962	1546	5	0	0	0	0	0	0
ES																	
2d	G	302	1	0	-1	0	70410	0	0	0	0	0	0	0	0	0	0
ES																	
1i	A	304	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0
	G	306	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	308	0	0	0	0	70410	0	0	0	0	0	0	0	0	0	0
	A	310	0	0	0	0	70410	0	0	0	0	0	0	0	0	0	0
	C	312	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	314	0	0	0	0	70410	0	0	0	0	0	0	0	0	0	0

G	316	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G	318	0	0	0	0	70410	0	0	0	0	0	0	0	0	0	0
A	320	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

RPS12 MiSeq +REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	130	0	0	0	10	0	0	0	2	26	0	0	0	0	0	0	0
	A	132	0	0	0	133	28	0	0	0	0	0	0	0	0	0	0	0
	A	134	0	0	0	196	161	0	0	0	0	0	0	0	0	0	0	0
	A	136	0	0	0	1023	357	0	0	0	0	0	0	0	0	0	0	0
	C	138	1	1	0	1807	1	3031	0	0	0	0	0	0	0	0	0	0
	C	140	0	0	0	294	3187	9	4	0	0	0	0	0	0	0	0	0
5' end	C	142	0	0	0	22408	3481	4	0	0	0	0	0	0	0	0	0	0
	A	144	1	1	0	2291	22	28085	54	0	0	0	0	0	0	0	0	0
	A	146	0	0	0	778	28177	7	1	0	0	0	0	0	0	0	0	0
	A	148	1	1	0	18	1	28972	1	0	0	0	0	0	0	0	0	0

C	150	0	0	0	0	28975	1	0	0	0	0	0	0	0	0	0	0
A	152	0	0	0	2	28976	0	0	0	0	0	0	0	0	0	0	0
C	154	0	0	0	14	28978	0	0	0	0	0	0	0	0	0	0	0
G	156	4	4	0	19	0	7	3	7	28987	7	0	0	0	0	0	0
A	158	0	0	0	0	29011	0	0	0	0	0	0	0	0	0	0	0
A	160	1	1	0	9	1	29015	0	0	0	0	0	0	0	0	0	0
A	162	0	0	0	8	29020	0	0	0	0	0	0	0	0	0	0	0
C	164	0	0	0	1	29028	0	0	0	0	0	0	0	0	0	0	0
A	166	0	0	0	4	29029	0	0	0	0	0	0	0	0	0	0	0
A	168	0	0	0	0	29033	0	0	0	0	0	0	0	0	0	0	0
A	170	0	0	0	3	29033	0	0	0	0	0	0	0	0	0	0	0
C	172	0	0	0	0	29036	0	0	0	0	0	0	0	0	0	0	0
A	174	1	1	0	20	81	28955	0	0	0	0	0	0	0	0	0	0
A	176	0	0	0	1	29027	9	11	9	0	0	0	0	0	0	0	0
A	178	0	0	0	0	29002	33	6	15	1	0	0	0	0	0	0	0

	G	180	0	0	0	4	28928	52	76	0	1	0	0	0	0	0	0	0
	A	182	1	1	0	2	3210	25664	154	28	3	0	2	0	0	0	0	0
	A	184	0	0	0	1	26059	731	2174	65	12	5	15	2	0	0	0	0
	A	186	0	0	0	3	27333	1320	284	74	31	3	19	0	0	0	0	0
ES77i	A	188	0	1	1	3	28099	263	300	301	45	31	23	4	1	0	0	0
ES76i	A	190	0	1	1	19	26022	353	250	168	2069	59	120	3	6	21	1	0
ES75i	G	192	0	4	4	5	24415	481	2196	192	1722	54	26	3	1	0	0	0
ES74i	G	194	0	8	8	51	24078	330	310	240	65	1945	172	111	1498	142	27	43
	C	196	0	0	0	3	28276	499	227	106	29	4	4	0	0	0	0	0
	G	198	0	0	0	7	28907	159	65	20	0	0	0	0	0	0	0	0
ES73i	A	200	0	1	1	20	23965	4562	261	278	38	8	44	3	1	0	0	0
ES72i	G	202	0	1	1	13	24058	4591	334	127	35	19	14	1	0	0	0	0
ES71i	G	204	0	1	1	18	24189	4383	364	151	97	3	10	0	0	0	0	0
	A	206	0	0	0	30	28033	410	365	139	48	163	44	0	15	0	0	0
ES70d	G	208	6	5	-1	49	287	455	742	250	202	4084	21739	1492	21	0	0	0

ES69i	A	210	0	1	1	8	23669	5001	212	94	80	150	70	2	3	1	2	1
ES68i	G	212	0	1	1	2	23493	4988	530	166	78	21	3	0	2	0	6	4
ES67d	G	214	1	0	-1	10	5212	23556	348	72	70	10	17	0	3	0	9	1
ES66i	G	216	0	2	2	7	23969	207	4951	139	13	11	4	1	12	0	0	0
ES65i	G	218	0	2	2	21	23577	309	5083	154	148	38	0	12	0	0	0	0
ES64i	A	220	0	3	3	12	23287	489	347	4957	33	10	208	8	2	0	0	0
	C	222	0	0	0	0	28385	739	96	78	40	7	0	0	0	0	0	0
ES63d	G	224	1	0	-1	5	8325	20924	78	15	3	0	0	0	0	0	0	0
ES62i	G	226	0	4	4	14	22889	797	455	264	4894	50	0	1	0	0	0	0
ES61i	A	228	0	4	4	34	20799	922	2565	211	4865	6	0	0	0	2	0	2
ES60i	G	230	0	3	3	7	20631	263	250	8145	106	8	0	0	0	0	0	0
ES59i	A	232	0	4	4	17	20645	61	24	453	8231	0	1	0	0	0	0	0
ES58i	G	234	0	1	1	15	20536	8334	61	90	399	4	1	0	0	0	0	0
ES57i	A	236	0	2	2	4	20504	555	8368	7	3	0	0	0	0	0	0	0
ES56i	A	238	0	2	2	6	20451	143	8840	6	5	0	0	0	0	0	0	0

ES55i	A	240	0	1	1	7	20549	8454	428	11	3	3	0	0	0	0	0	0
ES54i	G	242	0	1	1	4	20623	8732	88	7	2	0	0	2	0	0	0	0
	A	244	0	0	0	1	28726	85	616	12	2	1	4	0	12	0	0	0
	G	246	0	0	0	0	29193	20	240	4	0	2	0	0	0	0	0	0
ES53i	C	248	0	1	1	0	20612	8811	27	4	4	0	1	0	0	0	0	0
	C	250	0	0	0	3	29432	27	0	0	0	0	0	0	0	0	0	0
	G	252	0	0	0	3	29441	7	7	7	0	0	0	0	0	0	0	0
ES52d	C	254	2	0	-2	17	9104	244	20092	13	11	0	1	0	0	1	0	0
	G	256	0	0	0	0	29458	8	16	0	0	0	0	0	0	0	0	0
	A	258	0	0	0	0	28356	1041	45	27	7	0	4	2	0	0	0	0
ES51i	G	260	0	2	2	5	18252	2510	8619	90	4	0	3	5	1	0	0	0
	C	262	0	0	0	3	29444	15	9	5	10	2	2	0	0	0	0	0
	C	264	0	0	0	0	29485	5	0	0	0	0	0	0	0	0	0	0
	C	266	0	0	0	2	29478	3	9	0	0	0	0	0	0	0	0	0
	A	268	0	0	0	2	29474	15	3	0	0	0	0	0	0	0	0	0

	G	270	0	0	0	2	29009	198	104	175	0	6	2	0	0	0	0	0
ES50i	C	272	0	2	2	4	22564	1230	5518	182	2	1	0	0	0	0	0	0
	C	274	0	0	0	12	26040	3444	18	0	0	0	0	0	0	0	0	0
	G	276	0	0	0	9	27209	1767	260	276	0	0	0	0	0	0	0	0
	G	278	0	0	0	61	28258	310	314	283	377	8	0	3	0	0	0	0
ES49i	A	280	0	1	1	9	18669	10604	267	24	17	0	0	3	0	0	0	0
	A	282	0	0	0	1	28270	1264	23	12	4	18	0	0	0	0	0	0
	C	284	0	0	0	6	28085	1499	6	2	0	0	0	0	0	0	0	0
	C	286	0	0	0	1	29368	221	9	0	0	0	0	0	0	0	0	0
	G	288	0	0	0	2	29566	24	9	0	0	0	0	0	0	0	0	0
	A	290	0	0	0	24	29254	85	218	5	25	15	2	0	0	0	0	0
	C	292	0	0	0	1	29242	224	149	4	6	0	0	0	0	0	0	0
	G	294	0	0	0	0	29601	20	3	2	0	0	0	0	0	0	0	0
ES48i	G	296	0	1	1	24	17339	12062	53	35	151	3	0	0	1	0	0	0
ES47i	A	298	0	1	1	20	17283	12092	61	205	3	5	7	1	0	0	0	0

ES46i	G	300	0	2	2	9	17286	527	11810	14	12	10	0	13	0	0	0	0
ES45i	A	302	0	1	1	5	17463	12113	56	20	11	4	0	14	1	0	0	0
ES44i	G	304	0	1	1	3	17117	11354	1104	55	18	33	5	0	0	0	0	0
	C	306	0	0	0	0	29155	343	169	20	0	0	0	0	0	0	0	0
ES43d	C	308	2	0	-2	14	13418	456	15817	6	0	0	0	0	0	0	0	0
ES42d	G	310	4	0	-4	25	15988	343	187	135	13025	35	0	0	0	0	0	0
ES41i	A	312	0	1	1	19	16864	8921	3817	32	26	59	14	2	0	0	0	0
ES40i	A	314	0	4	4	75	16183	1084	779	584	9012	2108	45	0	0	0	0	0
ES39i	A	316	1	3	2	63	636	15715	864	11809	571	263	1	0	1	0	0	0
ES38i	A	318	0	1	1	9	12202	12991	3680	946	36	8	13	8	2	0	0	0
ES37i	A	320	0	1	1	34	12071	16447	832	197	83	203	46	7	5	0	1	0
	A	322	0	0	0	34	23596	1420	1063	550	265	165	2597	99	17	33	125	10
ES36i	G	324	0	4	4	43	17003	580	312	110	11888	66	21	5	0	1	1	0
ES35i	G	326	0	3	3	68	14803	1504	417	12402	864	40	5	4	0	1	0	1
	G	328	0	0	0	62	23768	4721	235	662	638	40	7	25	0	0	0	0

	A	330	0	0	0	133	24409	546	316	225	259	115	74	61	190	3484	418	89
ES34i	G	332	0	1	1	37	11343	14032	4694	113	16	55	11	4	21	0	0	2
ES33i	G	334	0	2	2	42	10602	5275	13228	243	853	44	66	20	3	1	0	0
	C	336	0	0	0	10	23640	6307	104	170	109	4	0	10	3	0	0	0
	G	338	0	0	0	7	29914	252	35	153	1	0	0	0	0	0	0	0
ES32i	G	340	0	2	2	62	9497	752	19470	137	283	113	14	14	94	21	6	0
ES31i	G	342	0	8	8	89	9461	504	868	165	246	2191	160	478	15834	573	9	3
ES30i	G	344	0	2	2	36	9317	330	20099	223	525	28	7	4	2	1	1	0
ES29i	A	346	0	4	4	171	9258	943	599	204	19492	28	44	3	0	1	0	0
ES28i	G	348	0	2	2	26	9885	805	19948	83	8	0	0	0	0	0	0	0
	G	350	0	0	0	5	29780	511	118	279	8	51	0	0	0	0	0	0
ES27i	A	352	0	3	3	40	9117	458	278	20905	11	0	2	0	0	0	0	0
	G	354	0	0	0	9	28535	442	1734	57	16	7	0	1	0	0	0	0
ES26i	A	356	0	2	2	8	6989	1710	22030	34	14	4	12	10	0	0	0	0
ES25i	G	358	0	1	1	25	6485	22656	1635	9	23	6	0	1	0	0	0	0

ES24d	C	360	3	1	-2	14	1897	23044	121	5765	2	7	0	0	0	0	0	0
	A	362	0	0	0	8	30496	85	208	58	0	0	0	0	0	0	0	0
ES23i	A	364	0	2	2	11	9286	3487	17932	44	47	60	3	0	0	1	0	0
ES22i	A	366	0	3	3	36	6820	3391	1042	19284	73	141	122	8	0	0	0	0
ES21i	A	368	0	2	2	33	5863	819	20184	3995	34	3	15	3	4	0	0	0
ES20i	A	370	0	1	1	35	5704	19656	3911	1523	114	18	18	1	0	0	0	0
	G	372	0	0	0	11	25048	3874	1695	154	178	14	7	4	0	0	0	0
	A	374	0	0	0	14	28696	1498	450	137	68	96	33	0	3	2	3	0
ES19d	G	376	3	0	-3	15	24947	522	236	5136	110	20	19	9	0	0	3	0
	G	378	0	0	0	1	30003	393	375	102	52	62	6	13	0	1	1	2
	G	380	0	0	0	14	28901	1109	616	343	21	14	1	3	0	2	0	1
	G	382	1	1	0	24	2629	27230	436	435	119	179	3	9	0	0	0	1
	G	384	0	0	0	33	28073	1800	990	112	28	48	3	4	2	0	0	0
ES18i	G	386	0	1	1	25	7385	23114	265	49	39	167	20	42	0	0	1	1
	G	388	0	0	0	137	27404	1061	2027	171	313	141	60	1	1	0	1	0

ES17i	G	390	0	4	4	77	4633	230	1469	609	24099	149	66	6	6	2	10	2
ES16i	G	392	0	2	2	91	5226	433	25332	116	212	20	14	16	2	0	2	1
	A	394	0	0	0	36	29323	1711	367	20	7	0	10	4	1	2	0	0
ES15i	A	396	0	3	3	18	5595	647	269	24915	17	0	9	0	5	2	0	0
	C	398	0	0	0	8	28933	2297	157	66	13	0	0	1	0	0	0	0
	C	400	0	0	0	33	30816	585	80	4	1	0	0	0	0	0	0	0
	C	402	0	0	0	138	30164	1350	37	11	3	6	0	0	0	0	0	0
ES14d	G	404	3	0	-3	677	27735	672	1105	2339	24	7	0	1	0	0	0	0
ES13d	G	406	4	0	-4	171	26257	439	3325	106	2171	19	30	8	1	2	1	2
ES12i	G	408	0	1	1	42	5868	26087	209	319	3	8	0	0	0	0	0	0
ES11d	A	410	2	1	-1	170	404	26678	2072	2829	574	1	14	5	0	0	0	0
	A	412	0	0	0	107	28795	3536	357	12	2	3	1	0	0	0	0	0
	A	414	0	0	0	210	28971	3485	347	5	1	0	5	0	0	4	1	0
	G	416	0	0	0	60	32563	64	397	0	2	0	0	0	0	0	0	0
ES10i	A	418	0	1	1	73	3698	29364	11	0	5	0	0	14	0	0	0	0

ES9i	A	420	0	2	2	610	4550	6687	22029	220	1	12	0	0	0	0	0	0
ES8i	A	422	0	1	1	5659	2748	30533	593	3374	385	0	1	0	0	0	0	0
	C	424	0	0	0	433	38743	74	636	3	2	1	0	0	0	0	0	0
	A	426	0	0	0	119	39852	4	0	2	0	0	0	0	0	0	0	0
ES7d	C	428	1	0	-1	80	36068	3887	35	7	1	0	0	0	0	0	0	0
	G	430	0	0	0	345	40014	35	8	0	0	0	0	0	0	0	0	0
ES6d	A	432	3	1	-2	145	418	35542	1049	3422	20	0	0	0	0	0	0	0
ES51	G	434	0	2	2	119	4058	66	36450	10	2	0	0	0	0	0	0	0
ES4i	A	436	0	1	1	80	3882	36757	26	7	7	0	4	0	1	0	0	0
	A	438	0	0	0	2	40732	4	8	1	0	0	0	0	0	0	0	0
	G	440	0	0	0	8	40735	1	11	0	0	0	0	0	0	0	0	0
ES3i	A	442	0	2	2	26	2501	40	38219	4	0	0	0	0	0	0	0	0
	G	444	0	0	0	15	40781	2	0	0	0	0	0	0	0	0	0	0
	A	446	0	0	0	16	40796	0	0	0	0	0	0	0	0	0	0	0
ES12d	A	448	4	3	-1	36	0	0	0	38442	2399	2	1	4	0	0	0	0

	G	450	0	0	0	0	40848	0	0	0	0	0	0	0	0	0	0	0
	A	452	0	0	0	0	40848	0	0	0	0	0	0	0	0	0	0	0
ES1i	A	454	0	1	1	0	0	40848	0	0	0	0	0	0	0	0	0	0
	A	456	1	1	0	0	0	40848	0	0	0	0	0	0	0	0	0	0
	A	458	0	0	0	0	40848	0	0	0	0	0	0	0	0	0	0	0
	G	460	0	0	0	0	40848	0	0	0	0	0	0	0	0	0	0	0
	A	462	0	0	0	0	40848	0	0	0	0	0	0	0	0	0	0	0
	A	464	1	1	0	0	0	40848	0	0	0	0	0	0	0	0	0	0
3' end	G	466	1	1	0	0	0	40848	0	0	0	0	0	0	0	0	0	0

RPS12 MiSeq +REL1

	T-stripped sequence	Position	no of Ts present before that position: pre-edited	no of Ts present before that position: edited	editing events (- = del; + = ins)	5' end in no. of cat1 reads	no. of reads with indicated no. of Ts present before that position: 0	1	2	3	4	5	6	7	8	9	10	11
	A	130	0	0	0	1.79748	0	1	0	1	4	0	0	0	0	0	0	0
	A	132	0	0	0	5.69203	6	1	0	0	0	0	0	0	0	0	0	0
	A	134	0	0	0	14.979	26	0	0	0	0	0	0	0	0	0	0	0
	A	136	0	0	0	115.638	76	1	0	0	0	0	0	0	0	0	0	0
	C	138	1	1	0	272.019	1	1245	0	0	0	0	0	0	0	0	0	0
	C	140	0	0	0	36.2493	1370	7	0	0	0	0	0	0	0	0	0	0
5' end	C	142	0	0	0	2216	1490	1	1	0	0	0	0	0	0	0	0	0
	A	144	1	1	0	153.385	5	9362	7	0	0	0	0	0	0	0	0	0
	A	146	0	0	0	59.0174	9399	4	0	0	0	0	0	0	0	0	0	0
	A	148	1	1	0	3.59497	3	9597	1	0	0	0	0	0	0	0	0	0

C	150	0	0	0	0	9605	0	4	0	0	0	0	0	0	0	0	0
A	152	0	0	0	0.59916	9609	0	0	0	0	0	0	0	0	0	0	0
C	154	0	0	0	2.69623	9611	0	0	0	0	0	0	0	0	0	0	0
G	156	4	4	0	3.29539	0	0	6	10	9615	0	0	0	0	0	0	0
A	158	0	0	0	0	9631	0	0	0	0	0	0	0	0	0	0	0
A	160	1	1	0	0.29958	0	9631	0	0	0	0	0	0	0	0	0	0
A	162	0	0	0	1.79748	9632	0	0	0	0	0	0	0	0	0	0	0
C	164	0	0	0	0.89874	9638	0	0	0	0	0	0	0	0	0	0	0
A	166	0	0	0	0.29958	9641	0	0	0	0	0	0	0	0	0	0	0
A	168	0	0	0	0	9642	0	0	0	0	0	0	0	0	0	0	0
A	170	0	0	0	0	9642	0	0	0	0	0	0	0	0	0	0	0
C	172	0	0	0	1.4979	9642	0	0	0	0	0	0	0	0	0	0	0
A	174	1	1	0	1.19832	9	9636	2	0	0	0	0	0	0	0	0	0
A	176	0	0	0	0	9643	7	1	0	0	0	0	0	0	0	0	0
A	178	0	0	0	1.4979	9650	0	1	0	0	0	0	0	0	0	0	0

	G	180	0	0	0	0	9642	9	5	0	0	0	0	0	0	0	0	0
	A	182	1	1	0	0.59916	288	9358	11	1	0	0	0	0	0	0	0	0
	A	184	0	0	0	0	9401	46	208	2	1	0	0	0	0	0	0	0
	A	186	0	0	0	0	9530	92	26	8	2	0	0	0	0	0	0	0
ES77i	A	188	0	1	1	0.89874	9605	7	16	18	9	1	2	0	0	0	0	0
ES76i	A	190	0	1	1	0.59916	9392	22	18	15	202	1	9	2	0	0	0	0
ES75i	G	192	0	4	4	0.29958	9280	29	217	10	121	1	5	0	0	0	0	0
ES74i	G	194	0	8	8	2.99581	9256	33	10	9	0	185	24	11	117	7	1	5
	C	196	0	0	0	0.29958	9616	32	22	4	0	0	0	0	0	0	0	0
	G	198	0	0	0	2.69623	9663	9	3	0	0	0	0	0	0	0	0	0
ES73i	A	200	0	1	1	0	9256	385	13	21	2	0	7	0	0	0	0	0
ES72i	G	202	0	1	1	2.69623	9250	408	17	5	4	0	0	0	0	0	0	0
ES71i	G	204	0	1	1	3.59497	9257	377	29	23	7	0	0	0	0	0	0	0
	A	206	0	0	0	0.29958	9601	44	37	7	1	7	3	0	5	0	0	0
ES70d	G	208	6	5	-1	9.88616	13	48	70	25	33	389	8656	484	4	1	0	0

ES69i	A	210	0	1	1	0.89874	9211	478	15	8	12	9	5	0	0	0	0	1
ES68i	G	212	0	1	1	0	9201	473	50	7	8	0	3	0	0	0	0	0
ES67d	G	214	1	0	-1	1.4979	505	9208	26	3	0	0	0	0	0	0	0	0
ES66i	G	216	0	2	2	2.69623	9257	7	471	8	3	0	0	0	1	0	0	0
ES65i	G	218	0	2	2	1.19832	9203	24	492	10	29	0	0	0	0	0	0	0
ES64i	A	220	0	3	3	0.29958	9181	35	25	487	2	0	29	1	0	0	0	0

ES64i	A	220	0	3	3	0.29958	9181	35	25	487	2	0	29	1	0	0	0	0
	C	222	0	0	0	0	9706	43	10	1	1	0	0	0	0	0	0	0
ES63d	G	224	1	0	-1	1.19832	777	8978	6	0	0	0	0	0	0	0	0	0
ES62i	G	226	0	4	4	0.29958	9143	70	31	24	484	14	0	0	0	0	0	0
ES61i	A	228	0	4	4	1.4979	8965	75	234	11	483	0	0	0	0	0	0	0
ES60i	G	230	0	3	3	1.19832	8936	26	17	791	4	0	0	0	0	0	0	0
ES59i	A	232	0	4	4	3.29539	8934	12	0	41	788	0	4	0	0	0	0	0
ES58i	G	234	0	1	1	1.4979	8925	816	10	0	35	0	0	0	0	0	0	0
ES57i	A	236	0	2	2	0.29958	8932	37	818	4	0	0	0	0	0	0	0	0
ES56i	A	238	0	2	2	0	8928	6	853	5	0	0	0	0	0	0	0	0

ES55i	A	240	0	1	1	0	8922	824	45	1	0	0	0	0	0	0	0	0
ES54i	G	242	0	1	1	0	8925	862	5	0	0	0	0	0	0	0	0	0
	A	244	0	0	0	0	9738	0	54	0	0	0	0	0	0	0	0	0
	G	246	0	0	0	0	9779	1	12	0	0	0	0	0	0	0	0	0
ES53i	C	248	0	1	1	0	8919	867	1	0	5	0	0	0	0	0	0	0
	C	250	0	0	0	0.89874	9791	1	0	0	0	0	0	0	0	0	0	0
	G	252	0	0	0	0.29958	9795	0	0	0	0	0	0	0	0	0	0	0
ES52d	C	254	2	0	-2	0	885	8	8902	0	1	0	0	0	0	0	0	0
	G	256	0	0	0	0.29958	9792	4	0	0	0	0	0	0	0	0	0	0
	A	258	0	0	0	0	9714	79	4	0	0	0	0	0	0	0	0	0
ES51i	G	260	0	2	2	2.39664	8749	214	833	2	0	0	0	0	0	0	0	0
	C	262	0	0	0	0	9803	2	0	0	0	0	0	0	0	0	0	0
	C	264	0	0	0	0	9805	0	0	0	0	0	0	0	0	0	0	0
	C	266	0	0	0	0	9805	0	0	0	0	0	0	0	0	0	0	0
	A	268	0	0	0	0.59916	9799	4	2	0	0	0	0	0	0	0	0	0
	G	270	0	0	0	0	9748	29	2	26	2	0	0	0	0	0	0	0
ES50i	C	272	0	2	2	0.89874	9149	109	528	16	5	0	0	0	0	0	0	0
	C	274	0	0	0	3.59497	9499	311	0	0	0	0	0	0	0	0	0	0
	G	276	0	0	0	2.39664	9634	126	29	33	0	0	0	0	0	0	0	0
	G	278	0	0	0	2.09706	9713	16	27	31	46	0	0	0	0	0	0	0

ES49i	A	280	0	1	1	0.89874	8814	1004	18	1	0	0	0	0	0	0	0	0
	A	282	0	0	0	0.89874	9690	144	5	1	0	0	0	0	0	0	0	0
	C	284	0	0	0	1.19832	9690	153	0	0	0	0	0	0	0	0	0	0
	C	286	0	0	0	0.29958	9830	16	0	1	0	0	0	0	0	0	0	0
	G	288	0	0	0	1.79748	9846	1	1	0	0	0	0	0	0	0	0	0
	A	290	0	0	0	0	9827	1	24	0	2	0	0	0	0	0	0	0
	C	292	0	0	0	0.29958	9824	17	9	0	4	0	0	0	0	0	0	0
	G	294	0	0	0	0.29958	9853	2	0	0	0	0	0	0	0	0	0	0
ES48i	G	296	0	1	1	0.89874	8672	1171	4	1	8	0	0	0	0	0	0	0
ES47i	A	298	0	1	1	0	8668	1172	10	9	0	0	0	0	0	0	0	0
ES46i	G	300	0	2	2	3.29539	8674	39	1145	0	3	0	0	0	0	0	0	0
ES45i	A	302	0	1	1	0	8669	1188	4	9	0	0	0	0	0	0	0	0
ES44i	G	304	0	1	1	0.89874	8651	1120	86	10	5	0	0	0	0	0	0	0
	C	306	0	0	0	0.59916	9808	46	19	0	0	0	0	0	0	0	0	0
ES43d	C	308	2	0	-2	0.89874	1307	32	8539	0	0	0	0	0	0	0	0	0
ES42d	G	310	4	0	-4	4.19413	1499	61	16	13	8293	2	1	0	0	0	0	0
ES41i	A	312	0	1	1	2.09706	8630	891	364	4	5	0	0	0	0	0	0	0
ES40i	A	314	0	4	4	8.38826	8555	101	73	44	946	203	0	0	0	0	0	0
ES39i	A	316	1	3	2	2.99581	50	8513	58	1198	74	39	3	0	0	0	0	0
ES38i	A	318	0	1	1	0.59916	8223	1302	285	121	5	0	0	1	0	0	0	0

ES37i	A	320	0	1	1	0.89874	8179	1626	88	27	5	6	4	0	0	4	0	0
	A	322	0	0	0	7.48951	9388	178	106	57	11	8	173	17	0	0	10	0
ES36i	G	324	0	4	4	6.59077	8657	47	28	9	1211	4	11	0	0	0	0	0
ES35i	G	326	0	3	3	24.266	8461	102	45	1263	118	3	7	0	1	0	0	0
	G	328	0	0	0	8.68784	9427	413	25	93	120	1	2	1	0	0	0	0
	A	330	0	0	0	21.5698	9598	61	58	26	16	10	17	10	13	280	36	17

Appendix 7

Below is listed the fasta formatted sequences used in MEGA5 and Datamonkey analysis. These are listed without stop and start codons. (Chapter 4).

REL1

>T b brucei 927 Tb09.1.60.2970

CAACTCCAAAGGTTGGGTGCTCCACTACTTAAAAGGCTTGTGGGGGGATGCATACGCCAATCAACGGCGCCGATTATGCCATGTGTTGTTGTTAGTGGCTCGGGTGGTT
TTTTGACTCCCGTGCGTACTTACATGCCGCTCCCCAATGATCAAAGTGACTTCTCACCGTACATCGAGATTGATTTGCCAAGTGAAAGCCGCATCCAATCACTGCATAA
GAGCGGACTTGCGGCACAGGAGTGGGTTGCATGTGAAAAAGTGCATGGGACAAACTTTGGTATCTACTTAATTAACCAGGGAGACCACGAGGTTGTGAGGTTTGCAAAG
CGTAGTGGCATCATGGACCCAAATGAGAATTTCTTTGGTTACCACATCCTTATCGACGAGTTCACAGCACAAATTCGTATTCTAAATGACCTATTGAAACAAAAATATG
GACTGAGCCGTGTTGGGCGTTTGGTGCTCAATGGAGAACTGTTTGGTGCCAAGTATAAGCACCCACTCGTTCCAAAGAGTGAGAAGTGGTGCACGCTACCAAACGGGAA
GAAGTTCCCTATCGCCGGTGTTCAAATACAGAGGGAACCCTTTCCACAATATAGTCCAGAGCTCCATTTCTTTGCATTTGACATTAAGTACAGTGTGAGTGGCGCGGAA
GAGGACTTCGTGCTGCTTGGTTACGACGAGTTCGTCGAGTTTTCTTCCAAAGTGCCCAACCTGTTATACGCGCGGGCTCTTGTTCGTGGGACGTTGGATGAGTGTTTAG
CATTTGATGTGGAGAATTTCATGACACCGCTGCCCGCGTTGCTTGGCTTAGGTAATTACCCCCTTGAGGGGAATTTGGCAGAGGGTGTGGTCATTCGTCACGTACGTCG
CGGGGATCCGGCGGTAGAGAAACACAACGTCTCGACTATAATCAAACTCCGCTGCTCCAGTTTCATGGAGTTGAAGCACCCCGGTAAGCAGAAGGAACTGAAGGAGACA
TTTATCGACACGGTACGCTCCGGAGCTCTGCGGCGGGTGCGAGGGAATGTGACCGTTATCTCAGACTCCATGTTGCCTCAGGTGGAGGCAGCGGCCAATGATCTGTTAC
TAAACAATGTAAGCGATGGCAGGTTGAGCAATGTGCTATCGAAGATTGGTCGTGAACCTCTGCTGTCTGGGGAGGTATCGCAGGTTGATGTGGCGCTTATGCTTGCTAA
GGACGCACTGAAAGACTTTCTCAAGGAGGTAGATAGTTTGGTGTTGAACACAACCCTTGCTTTTCGCAAGTTGCTAATCACCAATGTCTACTTCGAGTCTAAGCGGTTA
GTGGAGCAAAAGTGGAAGGAACTCATGCAGGAAGAAGCAGCCGCGCAGTCAGAGGCTATTCCACCACTTTCACCCGCTGCCCCCACAAAGGGCGAA

>T b brucei 427 Tb427tmp.160.2970

CAACTCCAAAGGTTGGGTGCTCCACTACTTAAAAGGCTTGTGGGGGGATGCATACGCCAATCAACGGCGCCGATTATGCCATGTGTTGTTGTTAGTGGCTCGGGTGTTT
TTTTGACTCCCGTGCGTACTTATATGCCGCTCCCCAATGATCAAAGTGACTTCTCACCGTACATCGAGATTGATTTGCCAAGTGAAAGCCGCATCCAATCACTGCATAA
GAGCGGACTTGCGGCACAGGAGTGGGTTGCATGTGAAAAAGTGCATGGGACAAACTTTGGTATCTACTTAATTAACCAGGGAGACCACGAGGTTGTGAGGTTTGCAAAG
CGTAGTGGCATCATGGACCCAAATGAGAATTTCTTTGGTTACCACATCCTTATCGACGAGTTCACAGCACAAATTCGTATTCTAAATGACTTATTGAAACAAAAATATG
GACTGAGCCGTGTTGGGCGTTTGGTGCTCAATGGAGAACTGTTTGGTGCCAAGTATAAGCACCCACTCGTTCCAAAGAGTGAGAAGTGGTGCACGCTACCAAACGGGAA
GAAGTTCCCTATCGCCGGTGTTCAAATACAGAGGGAACCCTTTCCACAATATAGTCCAGAGCTCCATTTCTTTGCATTTGACATTAAGTACAGTGTGAGTGGCGCGGAA
GAGGACTTCGTGCTGCTTGGTTACGACGAGTTCGTCGAGTTTTCTTCCAAAGTGCCTAACCTGTTATACGCGCGGGCTCTTGTTCGTGGGACGTTGGATGAGTGTTTAG
CGTTTGATGTGGAGAATTTCATGACACCGCTGCCCGCGTTGCTTGGCTTAGGTAATTACCCCCTTGAGGGGAATTTGGCAGAGGGTGTGGTCATTCGTCACGTACGTCG

CGGGGATCCGGCGGTAGAGAAACACAACGTCTCGACTATAATCAAACTCCGCTGCTCCAGTTTCATGGAGTTGAAGCACCCCGGTAAGCAGAAGGAACTGAAGGAGACA
TTTATCGACACGGTACGCTCCGGAGCTCTGCGGCGGGTGCGAGGGAATGTGACCGTTATCTCAGACTCCATGTTGCCTCAGGTGGAGGCAGCGGCCAATGATCTGTTAC
TAAACAATGTAAGCGATGGCAGGTTGAGCAATGTGCTATCGAAGATTGGTCGTGAACCTCTGCTGTCTGGGGAGGTATCGCAGGTTGATGTGGTGCTTATGCTTGCTAA
GGACGCACTGAAAGACTTTCTCAAGGAGGTAGATAGTTTGGTGTTGAACACAACCCTTGCTTTTCGCAAGTTGCTAATCACCAATGTCTACTTCGAGTCTAAGCGGTTA
GTGGAGCAAAAGTGGAAGGAACTCATGCAGGAAGAAGCAGCCGCGCAGTCAGAGGCTATTCCACCACTTTCACCGGCTGCCCCCACAAAGGGCAAA

>T b gambiense Tbg972.9.2300

CAACTCCAAAGGTTGGGTGCTCCACTACTTAAAAGGCTTGTGGGGGGATGCATACGCCAATCAACGGCGCCGATTATGCCATGTGTTGTTGTTAGTGGCTCGGGTGTTT
TTTTGACTCCCGTGCGTACTTATATGCCGCTCCCCAATGATCAAAGTGACTTCTCACCGTACATCGAGATTGATTTGCCAAGTGAAAGCCGCATCCAATCACTGCATAA
GAGCGGACTTGCGGCACAGGAGTGGGTTGCATGTGAAAAAGTGCATGGGACAAACTTTGGTATCTACTTAATTAACCAGGGAGACCACGAGGTTGTGAGGTTTGCAAAG
CGTAGTGGCATCATGGACCCAAATGAGAATTTCTTTGGTTACCACATCCTTATCGACGAGTTCACAGCACAAATTCGTATTCTAAATGACTTATTGAAACAAAAATATG
GACTGAGCCGTGTTGGGCGTTTGGTGCTCAATGGAGAACTGTTTGGTGCCAAGTATAAGCACCCACTCGTTCCAAAGAGTGAGAAGTGGTGCACGCTACCAAACGGGAA
GACGTTCCCTATCGCCGGTGTTCAAATACAGAGGGAACCCTTTCCACAATATAGTCCAGAGCTCCATTTCTTTGCATTTGACATTAAGTACAGTGTGAGTGGCGCGGAA
GAGGACTTCGTGCTGCTTGGTTACGACGAGTTCGTCGAGTTTTCTTCCAAAGTGCCTAACCTGTTATACGCGCGGGCTCTTGTTCGTGGGACGTTGGATGAGTGTTTAG
CGTTTGATGTGGAGAATTTCATGACACCGCTGCCCGCGTTGCTTGGCTTAGGTAATTACCCCCTTGAGGGGAATTTGGCAGAGGGTGTGGTCATTCGTCACGTACGTCG
CGGGGATCCGGCGGTAGAGAAACACAACGTCTCGACTATAATCAAACTCCGCTGCTCCAGTTTCATGGAGTTGAAGCACCCCGGTAAGCAGAAGGAACTGAAGGAGACA
TTTATCGACACGGTACGCTCCGGAGCTCTGCGGCGGGTGCGAGGGAATGTGACCGTTATCTCAGACTCCATGTTGCCTCAGGTGGAGGCAGCGGCCAATGATCTGTTAC
TAAACAATGTAAGCGATGGCAGGTTGAGCAATGTGCTATCGAAGATTGGTCGTGAACCTCTGCTGTCTGGGGAGGTATCGCAGGTTGATGTGGTGCTTATGCTTGCTAA
GGACGCACTGAAAGACTTTCTCAAGGAGGTAGATAGTTTGGTGTTGAACACAACCCTTGCTTTTCGCAAGTTGCTAATCACCAATGTCTACTTCGAGTCTAAGCGGTTA
GTGGAGCAAAAGTGGAAGGAACTCATGCAGGAAGAAGCAGCCGCGCAGTCAGAGGCTATTCCACCACTTTCACCGGCTGCCCCCACAAAGGGCGAA

>T evansi STIB805 (unpublished)

CAACTCCAAAGGTTGGGTGCTCCACTACTTAAAAGGCTTGTGGGGGGATGCATACGCCAATCAACGGCGCCGATTATGCCATGTGTTGTTGTTAGTGGCTCGGGTGTTT
TTTTGACTCCCGTGCGTACTTATATGCCGCTCCCCAATGATCAAAGTGACTTCTCACCGTACATCGAGATTGATTTGCCAAGTGAAAGCCGCATCCAATCACTGCATAA
GAGCGGACTTGCGGCACAGGAGTGGGTTGCATGTGAAAAAGTGCATGGGACAAACTTTGGTATCTACTTAATTAACCAGGGAGACCACGAGGTTGTGAGGTTTGCAAAG
CGTAGTGGCATCATGGACCCAAATGAGAATTTCTTTGGTTACCACATCCTTATCGACGAGTTCACAGCACAAATTCGTATTCTAAATGACTTATTGAAACAAAAATATG
GACTGAGCCGTGTTGGGCGTTTGGTGCTCAATGGAGAACTGTTTGGTGCCAAGTATAAGCACCCACTCGTTCCAAAGAGTGAGAAGTGGTGCACGCTACCAAACGGGAA
GAAGTTCCCTATCGCCGGTGTTCAAATACAGAGGGAACCCTTTCCACAATATAGTCCAGAGCTCCATTTCTTTGCATTTGACATTAAGTACAGTGTGAGTGGCGCGGAA
GAGGACTTCGTGCTGCTTGGTTACGACGAGTTCGTCGAGTTTTCTTCCAAAGTGCCTAACCTGTTATACGCGCGGGCTCTTGTTCGTGGGACGTTGGATGAGTGTTTAG
CGTTTGATGTGGAGAATTTCATGACACCGCTACCCGCGTTGCTTGGCTTAGGTAATTACCCCCTTGAGGGGAATTTGGCAGAGGGTGTGGTCATTCGTCACGTACGTCG
CGGGGATCCGGCGGTAGAGAAACACAACGTCTCGACTATAATCAAACTCCGCTGCTCCAGTTTCATGGAGTTGAAGCACCCCGGTAAGCAGAAGGAACTGAAGGAGACA
TTTATCGACACGGTACGCTCCGGAGCTCTGCGGCGGGTGCGAGGGAATGTGACCGTTATCTCAGACTCCATGTTGCCTCAGGTGGAGGCAGCGGCCAATGATCTGTTAC
TAAACAATGTAAGCGATGGCAGGTTGAGCAATGTGCTATCGAAGATTGGTCGTGAACCTCTGCTGTCTGGGGAGGTATCGCAGGTTGATGTGGTGCTTATGCTTGCTAA

GGACGCACTGAAAGACTTTCTCAAGGAGGTAGATAGTTTGGTGTTGAACACAACCCTTGCTTTTCGCAAGTTGCTAATCACCAATGTCTACTTCGAGTCTAAGCGGTTA
GTGGAGCAAAAGTGGAAGGAACTCATGCAGGAAGAAGCAGCCGCGCAGTCAGAGGCTATTCCACCACTTTCACCGGCTGCCCCCACAAAGGGCGAA

>T congolense TcIL3000.9.1420

CGCCCCCAGCGGTTATGTGTTTCTGTGGCAACCAGGCTTAGCCACGTTTGCGCTCACTGGCGCACTGCGTCTGTGGCCCTGTGCCCTGCAGTGTGTGGTTCGAGTAGTA
TTTTTACTTCCACGCGCGCGTACATGCCGCTTCCCAATGACCAAAGTGATTTTTCGCCGTATATTGAGATTGACTTGCCGAGCGAAAGTCGAATCCAGCAGCTTCACAA
GAGTGGACTTGGAGGACAGGAGTGGGTTGCGTGCGAGAAGGTCCACGGGACGAACTTTGGCATATATCTTATCAATAATGGAGACCATGAAACAGTGCGGTTTGCTAAG
CGCAGTGGTATCATGGACCCCAATGAGAACTTCTTCGGTTACCATATACTCATTGATGAGTTCACCGCGCAGATCCGAATTTTGAATGACTTGCTCAAGCAGAAATATG
GACTGAGCCGTATTGGCCGCTTAGTGCTAAATGGTGAACTGTTTGGTGCCAAGTACAAGCATCCCCTTGTCCCCAAGAGCGAGAAATGGTGCACGTTGCCGAATGGGAA
GCGGTTTCCGATTGCTGGTGTTCAGATCCAGCGCGAACCGTTTCCACAGTACAGCCCCGAACTGCACTTCTTTGCCTTCGACATTAAGTACAGCGTGAGCGGTGCCGAG
GAGGATTTTGTACTGCTGGGTTATGACGAGTTTGTAGAGTTCTGCTCCAAGGTGCCTAACTTACTGTATGCACGGGCCATTGTCCGTGGAACGTTGGATGAGTGTTTGG
CATTTGATGTGGAAAACTTCACTACACCCCTGCCCGCACTGCTTGGTCTGGGGAACTATCCGCTCGAAGGCAACCTTGCGGAAGGTGTCGTCATACGCCATGTCCGCCG
CGGGGATCCAGCGGTGGAAAAACACAATGTTTCCACCATCATCAAGTTACGTTGTTCGAGCTTCATGGAATTGAAGCACCCAGGTAAACAGAAGGAGTTGAAGGAAACA
TTTATCGATACCGTGCGTTCAGGGGCACTGCGGCGGGTGCGAGGAAATGTGACGGTAATCACTGACTCCATGCTCCCACAGGTGGAGGCAGCAGCAAACAACTTATTGT
TGAACAATGTGAGTGATGGAAGGCTAAGCAATGTTCTGTCAAAAATCGGCCGCGAGCCGCTGCTTTCGGGAGAGGTGTCGCAGCATGATGTGATTCTGATGCTTGCCAA
GGATGCATTGAAGGATTTCCTGAAGGATGTGGATAGTTTAGTCCTGAACACAGGCCTCGCATTCCGTAAGCTGTTGATTACTAACGTGTTCTTTGAGTCTAAGCGTTTG
GTGGAGCAGAAGTGGAAGGAGCTTATGAAGGAAGAGGCAGCCGCGCATGACGAAGCCACCGAGGAACTCCAGAGCAAGGGCCCCACGGAGAGTACA

>T vivax TvY486_0901490

ATTGTCAAGCGAGTATCAGCCCTTTTAGGTAGCGGTGTTCACCAACATTGGTGGTACGAAGTTCGGCGCCTGGCGCCTGGTAGCGCAACGGTCTCCGTCCTTCACAGTG
CCAGGCGGTACTACATGCCACTTCCGAGTGACCAGACGGACTTTTCTCCGTATATTGAAATTGACCTGCCTAGTGAGAGCCGAATTCAGGCCATTCACAAGAGTGGCCT
GGGAGCACAGGAGTGGGTTGCATGCGAAAAGGTTCACGGAACAAATTTTGGCATTTATTTGATCAACCAAGGCGACCAGGAGCTTTTGCGATTTGCCAAACGCAGTGGT
ATTATGGATCCAAATGAGAACTTCTTTGGGTACCACATACTTATTGATGAGTTTTCCGCACAGATTCGCATCCTGAACGACTTGCTAAAACAAAAATATGGACTGAGCC
GTATTGGCCGTCTTGTGCTAAACGGCGAGTTGTTCGGTGCAAAGTACAAGCACCCGCTTGTTCCCAAAAGCGAAAAGTGGTGTACTCTACCCAACGGAAAGCGTTTTCC
AATTGCAGGTGTTCAGATACAGCGGGAGCCTTTCCCACAGTATAGTCCAGAGCTGCATTTCTTTGCATTTGATGTGAAGTATTCAGTGAGTGGAGCTGAGGATGACTTT
GTGCTACTCGGTTATGACGAGTTTGTGGAACTGTGCTCAAAGGTGCCGAACTTGCTGTACGCACGGGCCCTCGTGCGCGGAACACTTGACGAGTGTCTTGCATTTGACG
TGGAGAACTTTACAACACCCTTGCCCGCACTGCTTGGACTTGGCAACTATCCCCTTGAAGGGAATCTGGCCGAGGGTGTCGTTATTCGTCACGTTCGGCGTGGGGACCC
AGCGGTTGAGAAGCACAACGTGTCGACAATAATTAAGTTGCGCTGCTCTAGTTTTATGGAACTGAAGCATCCAGGTAAACAAAAGGAATTAAAGGAAACTTTTATTGAT
ACCGTACGCTCTGGCGCCTTGCGGCGGGTGCGCGGCAATGTTACCGTCATTGCCGATTCCATGCTCCCACAGGTGGAAGCAGCAGCAAACAATCTGCTCCTGAACAACG
TCAGCGATGGCAGGCTGAGCAACGTTCTTTCAAAAATTGGTAGGGAACCGCTGCTCTCCGGAGAGGTGTCGCAGAAAAGTGTCGCAGTGATGCTTGCCAAGGACGCTCT
GAAGGACTTTCTGAAGGATGTAGATCCGCTCGTGCTCAATACGGGCTTGTCCTTCCGCAAGCTGCTAATATCTAACGTTTACTTCGAGTCCAGGCGCCTCGTTGAATCA
AAGTGGAAGCAACTCCTTAGTGAGGAGGCAGTTGCCCAAAATGAAGCGGACGGAGCGGATGCTGCTGCTGTTGGATCTGGTGCGCGG

>T cruzi (CLBEL) Tc00.1047053511585.20

AGGGGAGTGGCGTTGGTGAGGGGGTGGATTCGACGGCCTTCCTCCATCTTTTCTCTTTTTGGCAGGAGTGTGATGCACGGCACGACGACCTTTGCTGTGGCCACACGCA
CGTATATGCCGCTACCGAATGACCAGAGTGATTTTTCTCCCTACATTGAAATTGATTTGCCGAGCGAGAGCCGCATTCAGGCGATAAATAAGAGCGGCCTCGGCGCCCA
AGATTGGGTGGCCTGCGAGAAGGTGCATGGCACCAATTTCGGTATCTATCTCATCAACCTTGGCGACAATGAAGTGGTGCGATTCGCGAAGCGTAGTGGGATTATGGAC
CCCAACGAGAACTTCTTTGGGTACCACATCCTCATTGACGAGTTCACTGCCCAAATCCGCATTTTAAGCGACTTACTGAAGCAAAAGTACGGCTTAGGTCGGATTGGGC
GCGTCGTTCTTAATGGCGAGCTTTTTGGTGCGAAGTATAAACATCCCCTTGTTCCCCCGAGCGAGAAGTGGTGCACAATGCCGAATGGGAAGCGATTCCCCATCGCCGG
TGTCCAAATTCAACGGGAGGCGTTCCCGCAGTACAGCCCCGAGTTACATTTCTTTGCCTTTGATATTAAGTACTCTGTGAGTGGGGCGGAGGAGGACTTTGTGCTGCTT
GGATACGACGAGTTTGTGGAGTTTTGCTCCAAGATTCCAAATCTGCTTTACTCCAAAGCACTTGTGCGCGGGACCCTTGACGCATGCCTGGCGTTTGATGTGGAGAACT
TCCACACCCCTCTGCCGGCACTTTTGGGCCTTGGCAACTACCCGCTGGAGGGGAATCTCGCGGAGGGCGTGGTCATCCGACACGTGCATCGTGGGGATCCGGCGACTGA
AAAGCACAATGTCGCGACGATCTTGAAGCTGCGCTGCTCCAGCTTTATGGAGCTGAAGCACCCGGGCAAGCAGAAGGAGTTGAAGGAAACGTTTATTGATACAGTGCGC
ACAGGGGCTCTGCGACGTGTACGCGGGGACGTTACAGTGATTGCAGATTCCATGCTTCCGCAGGTGGAGGCCGCCGCGAATAATCTACTGTTAAACAATGTGAGCGATG
GTAGGCTTAGCAATGTTCTCTCCAAGATTGGGCGGGAACCGTTACTCTCTGGGCAGGTTTCACAGAATGATGTTTCCCTAATGCTTGCCAAGGACGCTCTGAAGGATTT
TCTCAAGGAGGTGGACGCGTTAGTGCTTAACACGAGTCTCGCTTTTCGAAAGATGCTTATTTCTAATGTGTACTTTGAGTCCAAAAGACTCGTTGCGCAGAAATGGAAG
GAACTCATGAAAGAGGAGGAAACGGCGCAGCAGGAGGCGGAGGCGGCCGCATTGAAGGCCCCTCCAGTAGCCGCC

>L tarentolae AY148476.1 (NCBI)

CGTCGACTGGCACTGCGTTGTGCGCCACGCTGCTGTCGTGCCACTCTGTGCAACAGTCGAGGCAGCGGCAGCGTCCTGAGGCGCCAGGCACCGGCGCTCGTACCGAGGG
CCGCCGTCTCTCCGCCATGCCTCCCCACCGCCAGTACCGCCTGCTTGCAGAGCCTGGCCACGTCGCGCCGCTGCTACCTGCCGCTTCCGAAAGACCAGGACGACTTCTC
GGCCTACGCGGAGATTGACCTACCCACAGAGACGCGCATCGAGGCAATCAGGCGCACGGTGATCGCCAACCAGGAGTGGGTCGCGTGCGAGAAGGTGCACGGCACAAAC
TTTGCCATCTACCTCATCAACGAAAATGAGGTGCGGTTCGCAAAGCGCAGCGGCATCATGGACCCGAACGAGCACTTCTTCGGCTACCACCTCCTCATCGACGACTTCA
CGGCGCAGGTCCGGGCGTTGTGCGCGCTGCTGAAGCGCAAGTACGGCGTCACAGGTCGCATGGGCCGCGTGGTGCTTCACGGGGAGCTGTTCGGGGCCAAGTACAAGCA
TCCGCTGGTGCCCACGAGCAAGAAGTGGTGCACGCTGCCGAATAAGAAGCGGATTCCGATCTCGGGGGTGGAGATACAGAGCGAGCCGTTCCCGCAGTACAGCCCGGAG
CTTCACTTCTTCGCCTTCGACGTCAAGTACTCCGTGTCGGGTGCTGAGGAGGACGTGGTGTTGCTGCCGTTCGACGACTTCACGGAGGTGTGCGCGCAGGTGCCCAATC
TCCTGTACGCGAGGTCGCTGGTGCGTGGCACACTGGACGAGTGTCTCGCCTTCGATGTGGAGAACTTTGTCACACCGCTGCCCGCGCTGCTCGGTCTGGGCAACTACCC
TCTCGAAGGCAACCTCGCCGAGGGCGTCGTCATACGCCACGTGCGCCGCGGCGACCCGGCGGTGGAGAGCAGCGGGGTCTCGACGATCATCAAGCTGCGCTGTTCTTCC
TTCATGGAGCTCAAGCATCCCGGCAAGCAGCAGGAACTGAAGGCAACGTTCCTTGACACCGTGCGTGCCGGCGCGCTGCAACGCGTGCGCGGAGGCAAGAAAGTGACGG
TGCTGTTGGACGCGTTGCTGCCGAACCTAGAGGCAGCCGCAAACGCGCTGCTGCTGAACAACGTCAGCGAGGGTCGTGTGAGCAACGTGTTGTCGAAGATCGGCCGCGA
GCCGCTGCTGACCGGCAAAGTAACGCAGCACGACGTGGTGCAGATGCTCGCGCAGGATGCCCTGAAGGACTTCCTCAAGGAGACGGACCCGGTCATATTGAACACGGCC
TTATCCTTTCGCAAGACGCTCATCCGAAGCGTGTACTTGGCGGCGGAGGAGCTGCTGCAGGGAGAATGGAAGCGCATCATAGACCGACTACAGGCGTCCCAGGCAGAGA
TGGACGCCGCGATGGCCGCGCAGGAGAAGGCAGAGGCGCAG

REL2

>T b brucei 927 Tb927.1.3030

TTGCGTCGCCTCGGTGTACGTCACTTCCGGCGAACGCCGCTCCTTTTTGTCGGTGGGGACGGCAGCATTTTTGAGCGCTACACAGAAATCGACAACTCCAACGAGCGGC
GGATTAATGCGCTGAAGGGATGCGGTATGTTTGAGGATGAGTGGATTGCCACTGAGAAGGTTCATGGAGCGAACTTCGGTATCTACTCCATTGAAGGTGAGAAAATGAT
CAGATACGCAAAGCGCAGTGGCATTATGCCTCCGAACGAGCACTTCTTTGGCTATCATATATTGATCCCTGAGTTGCAGCGATACATAACTTCTATTCGTGAGATGTTG
TGCGAGAAGCAGAAGAAGAAGCTACACGTTGTGCTCATTAATGGAGAGTTGTTTGGAGGAAAATATGACCACCCTAGCGTTCCAAAGACACGGAAAACGGTCATGGTGG
CTGGTAAGCCGCGAACAATAAGCGCCGTGCAGACCGACTCTTTTCCCCAATATAGCCCAGATCTCCACTTTTACGCATTCGACATAAAGTACAAGGAGACAGAGGATGG
AGACTACACTACTCTCGTCTACGATGAAGCCATAGAGCTGTTTCAGCGCGTTCCCGGGCTCCTGTACGCGAGGGCGGTCATTCGAGGGCCAATGTCAAAGGTTGCTGCA
TTTGATGTGGAACGCTTTGTCACAACCATCCCCCCTCTTGTTGGTATGGGCAACTACCCACTGACGGGCAACTGGGCTGAGGGGCTTGTCGTGAAGCACTCGAGGCTGG
GAATGGCGGGATTCGACCCCAAAGGTCCAACGGTGTTGAAGTTTAAATGCACAGCGTTCCAGGAGATATCCACAGACCGAGCGCAAGGGCCTCGCGTAGATGAAATGAG
GAACGTACGAAGGGACTCCATCAATCGCGCCGGTGTGCAATTGCCAGACTTAGAGAGTATTGTGCAAGATCCAATTCAACTCGAAGCTTCGAAACTACTCCTTAATCAT
GTGTGCGAGAATCGTCTCAAAAATGTGCTATCAAAGATCGGCACGGAGCCCTTCGAGAAGGAGGAAATGACACCAGACCAGTTGGCCACACTTCTTGCCAAAGACGTCC
TGAAAGACTTTCTAAAAGATACCGAACCCTCCATTGTAAATATACCGGTACTCATCCGCAAGGATCTGACTAGGTACGTTATATTTGAGTCCAGGCGGCTAGTTTGCTC
ACAGTGGAAGGACATTCTGAAGCGTCAGTCTCCTGACTTTAGCGAA

>T b brucei 427 Tb427.01.3030

TTGCGTTGCCTCGGTGTACGTCACTTCCGGCGAACGCCGCTCCTTTTTGTCGGTGGGGACGGCAGCATTTTTGAGCGCTACACAGAAATCGACAACTCCAACGAGCGGC
GGATTAATGCGCTGAAGGGATGCGGTATGTTTGAGGATGAGTGGATTGCCACTGAGAAGGTTCATGGAGCGAACTTCGGTATCTACTCCATTGAAGGTGAGAAAATGAT
CAGATACGCAAAGCGCAGTGGCATTATGCCTCCGAACGAGCACTTCTTTGGCTATCATATATTGATCCCCGAGTTGCAGCGATACGTAACTTCTATTCGTGAGATGTTG
TGCGAGAAGCAGAAGAAGAAGCTACACGTTGTGCTCATTAATGGAGAGTTGTTTGGAGGAAAATATGACCACCCTAGCGTTCCAAAGACACGGAAAACGGTCATGGTGG
CTGGTAAGCCGCGAACAATAAGCGCCGTGCAGACCGACTCTTTTCCCCAATATAGCCCAGATCTCCACTTTTACGCATTCGACATAAAGTACAAGGAGACAGAGGGTGG
AGACTACACTACTCTCGTCTATGATGAAGCCATAGAGCTGTTTCAGCGCGTTCCCGGGCTCCTGTACGCGAGGGCGGTCATTCGAGGGCCAATGTCAAAAGTTGCTGCA
TTTGATGTGGAACGCTTTGTCACAACCATCCCCCCTCTTGTTGGTATGGGCAACTACCCACTGACGGGCAACTGGGCTGAGGGGCTTGTCGTGAAGCACTCGAGGCTGG
GAATGGCGGGATTCGACCCCAAAGGTCCAACGGTGTTGAAGTTTAAATGCACAGCGTTTCAGGAGATATCCACAGACCGAGCACAAGGGCCTCGCGTAGATGAAATGAG
GAACGTACGAAGAGACTCCATCAATCGCGCCGGTGTGCAATTGCCAGACTTAGAGAGTATTGTGCAAGATCCAATTCAACTCGAAGCTTCGAAACTACTCCTTAATCAT
GTGTGCGAGAATCGTCTCAAAAATGTGCTATCAAAGATCGGCACGGAGCCCTTCGAGAAGGAGGAAATGACACCAGACCAGTTGGCCACACTTCTTGCCAAAGACGCCC
TGAAAGACTTTCTAAAAGATACCGAACCCTCCATTGTAAATATACCGGTACTCATCCGCAAGGATCTGACTAGGTACGTTATATTTGAGTCCAGGCGGCTAGTCTGCTC
ACAGTGGAAGGACATTCTGAAGCGTCAGTCTCCTGACTTTAGCGAA

>T b gambiense Tbg972.1.1840

TTGCGTTGCCTCGGTGTACGTCACTTCCGGCGAACGCCGCTCCTTTTTGTCGGTGGGGACGGCAGCATTTTTGAGCGCTACACAGAAATCGACAACTCCAACGAGCGGC
GGATTAATGCGCTGAAGGGATGCGGTATGTTTGAGGATGAGTGGATTGCCACTGAGAAGGTTCATGGAGCGAACTTCGGTATCTACTCCATTGAAGGTGAGAAAATGAT
CAGATACGCAAAGCGCAGTGGCATTATGCCTCCGAACGAGCACTTCTTTGGCTATCATATATTGATCCCCGAGTTGCAGCGATACGTAACTTCTATTCGTGAGATGTTG
TGCGAGAAGCAGAAGAAGAAGCTACACGTTGTGCTCATTAATGGAGAGTTGTTTGGAGGAAAATATGACCACCCTAGCGTTCCAAAGACACGGAAAACGGTCATGGTGG
CTGGTAAGCCGCGAACAATAAGCGCCGTGCAGACCGACTCTTTTCCCCAATATAGCCCAGATCTCCACTTTTACGCATTCGACATAAAGTACAAGGAGACAGAGGGTGG
AGACTACACTACTCTCGTCTATGATGAAGCCATAGAGCTGTTTCAGCGCGTTCCCGGGCTCCTGTACGCGAGGGCGGTCATTCGAGGGCCAATGTCAAAAGTTGCTGCA
TTTGATGTGGAACGCTTTGTCACAACCATCCCCCCTCTTGTTGGTATGGGCAACTACCCACTGACGGGCAACTGGGCTGAGGGGCTTGTCGTGAAGCACTCGAGGCTGG
GAATGGCGGGATTCGACCCCAAAGGTCCAACGGTGTTGAAGTTTAAATGCACAGCGTTTCAGGAGATATCCACAGACCGAGCACAAGGGCCTCGCGTAGATGAAATGAG
GAACGTACGAAGAGACTCCATCAATCGCGCCGGTGTGCAATTGCCAGACTTAGAGAGTATTGTGCAAGATCCAATTCAACTCGAAGCTTCGAAACTACTCCTTAATCAT
GTGTGCGAGAATCGTCTCAAAAATGTGCTATCAAAGATCGGCACGGAGCCCTTCGAGAAGGAGGAAATGACACCAGACCAGTTGGCCACACTTCTTGCCAAAGACGCCA
TGAAAGACTTTCTAAAAGATACCGAACCCTCCATTGTAAATATACCGGTACTCATCCGCAAGGATCTGACTAGGTACGTTATATTTGAGTCCAGGCGGCTAGTCTGCTC
ACAGTGGAAGGACATTCTGAAGCGTCAGTCTCCTGACTTTAGCGAA

>T evansi STIB805 (unpublished)

TTGCGTTGCCTCGGTGTACGTCACTTCCGGCGAACGCCGCTCCTTTTTGTCGGTGGGGACGGCAGCATTTTTGAGCGCTACACAGAAATCGACAACTCCAACGAGCGGC
GGATTAATGCGCTGAAGGGATGCGGTATGTTTGAGGATGAGTGGATTGCCACTGAGAAGGTTCATGGAGCGAACTTCGGTATCTACTCCATTGAAGGTGAGAAAATGAT
CAGATACGCAAAGCGCAGTGGCATTATGCCTCCGAACGAGCACTTCTTTGGCTATCATATATTGATCCCCGAGTTGCAGCGATACGTAACTTCTATTCGTGAGATGTTG
TGCGAGAAGCAGAAGAAGAAGCTACACGTTGTGCTCATTAATGGAGAGTTGTTTGGAGGAAAATATGACCACCCTAGCGTTCCAAAGACACGGAAAACGGTCATGGTGG
CTGGTAAGCCGCGAACAATAAGCGCCGTGCAGACCGACTCTTTTCCCCAATATAGCCCAGATCTCCACTTTTACGCATTCGACATAAAGTACAAGGAGACAGAGGGTGG
AGACTACACTACTCTCGTCTATGATGAAGCCATAGAGCTGTTTCAGCGCGTTCCCGGGCTCCTGTACGCGAGGGCGGTCATTCGAGGGCCAATGTCAAAAGTTGCTGCA
TTTGATGTGGAACGCTTTGTCACAACCATCCCCCCTCTTGTTGGTATGGGCAACTACCCACTGACGGGCAACTGGGCTGAGGGGCTTGTCGTGAAGCACTCGAGGCTGG
GAATGGCGGGATTCGACCCCAAAGGTCCAACGGTGTTGAAGTTTAAATGCACAGCGTTTCAGGAGATATCCACAGACCGAGCACAAGGGCCTCGCGTAGATGAAATGAG
GAACGTACGAAGAGACTCCATCAATCGCGCCGGTGTGCAATTGCCAGACTTAGAGAGTATTGTGCAAGATCCAATTCAACTCGAAGCTTCGAAACTACCCCTTAATCAT
GTGTGCGAGAATCGTCTCAAAAATGTGCTATCAAAGATCGGCACGGAGCCCTTCGAGAAGGAGGAAATGACACCAGACCAGTTGGCCACACTTCTTGCCAAAGACGCCC
TGAAAGACTTTCTAAAAGATACCGAACCCTCCATTGTAAATATACCGGTACTCATCCGCAAGGATCTGACTAGGTACGTTATATTTGAGTCCAGGCGGCTAGTCTGCTC
ACAGTGGAAGGACATTCTGAAGCGTCAGTCTCCTGACTTTAGCGAA

>T congolense TcIL3000.1.1450

CTGCGCCATCCCACCGCCCTCTACTTCCGGCGTACAGCTGTTCTGCTTGCCGAAGAGGCTGGTGCTGTCTTCGAACGTTACAGTGAGATTGACAACTCCAATGAGAGGC
GAATAAACGCCCTGAAGGCCTGTGGAATGTTTGAGGACGAGTGGATAGCAACGGAGAAAGTTCACGGCGCTAACTTTGGTATTTACTCCATTGAGAATGAAAAAATGAT
ACGCTACGCCAAGAGGAGCGGCATCATGCCTCCGAGTGAGCATTTCTTTGGTTACCACATATTGATCCCGCAACTACAAAAGTGTATAACCTCCATCCGCGAAATGCTG
TGCGAAAAGCTTAAGAGGAAGTTGCACATTGTACTCGTTAACGGTGAGTTGTTTGGAGGAAAGTACGACCACCCAAGCGTCCCAAAGACGCGGAAGACAGTCATGGTGG

CGGGAAAGCCACGAGTCATAAGCGCCGTACAGACAGACTCCTTTCCCCAGTACAGCCCAGATCTTCATTTCTATGCATTTGACATCAAGTACAAGGAAACAGCAGAGGG
AAATTACACAACTCTCGTTTACGACGATGCCATTCAGTTGTTTCAGAGTGTCCCCGGCCTGCTGTACGCAAGGGCCATCATTCGGGGCCCCATGTCAAAAGTTGCTGCA
TTTGACGTGGAGCATTTCACCACGACCATTCCTCCCCTCGTTGGGATGGGAAACTACCCACTGACAGGGAACTGGGCTGAAGGGCTTGTGGTGAAGCACGCCCGGCTTG
GAACGCCGGGGTTTGAGCCAAAAGGGCCGACGGTGTTGAAGTTCAAATGCACAGCCTTCCAGGAGATATCCACTGATCGAAGGCAGGGACCCCGTGTGGACGAAATGGA
AAATGTTCGTCGGGACTCCATAAGTCGATCGGGTGTGCAGCTGCCTGACTTGCAGAGCGTTATACAGGATCCAGTGCAGCTTGACGCGGCAAAACTGCTCCTCGACCAC
GTCTGTGACAATCGTCTCAAAAACGTTCTCTCGAAGATTGGCACCGAGCCATTTGAAAAACAAGACATGACCCCAGATCAGTTAGCAACACTCCTGGCGAAGGACGCAC
TGAAGGACTTCTTGAAGGACACAGAGCCCTCCATTGTGAGCACGCCCATACTAACCCGCAGGGATATGGCGAGGTATGTTTTGTTTGAGTCCAGGAAGTTGGTATGTTC
ACAGTGGAAGGCTATATTGAAACGCCAGTCCCCAGAAGCTGATGTG

>T vivax TvY486_0101350

CGACGCATATTTGTCGCACTCTCCCTCAGGAGATCGCCGTTCTTGCTCGTTGAGGATCGTGGTAACATCTTTGAGCGCTACACTGAGATCGAAAACTCTAATGAACGGC
GAATCAACGCCCTGAAATCGTGTGGAATGTTCGAGGATGAATGGATTGCCACAGAAAAAGTTCACGGCGCTAACTTTGGGATCTACTCAATAGAGGGTGAGAAGACTAT
ACGCTACGCCAAGAGAAGTGGAATTATGCCTCCCTCTGAGCATTTCTTCGGGTACCATGTGCTAATTCCACAGTTACAACAGTACATAACGTCTGTTCGCGAGATGCTC
TGCGACAAGATGCAAAAAAAGCTCCATACTGTCCTCATCAATGGTGAACTATTCGGTGGCAAGTATGACCATCCGAGTCTCCCAAAGACAAGAAAGACGGTTATGGTGG
CAGGCAAGCAGCGGACGATTAGTGCGGTACAGACGGATTCGTTTCCTCAATACAGTCCAGATCTTCATTTTTACGCATTTGATATCAAGTACAAGGAGAGCGCCGAGTC
TGATTACATTTCTCTTGTGTTTGACGCTGCAACGGAACTGTTCCAGAAGGTTCCAGGCCTTCTGTACGCTAAAGCAATTATTCGAGGACCCATGTCGAAGGTAGCTGCC
TTTGACGTAGAGCACTTTGTCACAACAATCCCTCCTCTTGTCGGTATGGGTAATTACCCACTCGCGGGGAACTGGGCTGAGGGCCTTGTTGTAAAGCACGCGCGACGTG
GAGAGGTTGGGTTTGACCCGAAAGGGGTGACAATTTTAAAGTTCAAATGCACCGCTTTTCAGGAAATATCTACCGATAGGAGGCAAGGGCCCCGTGTTGATGAGATGGA
GAGTGTACGACGCGACTCAATTGTGCGGTCGGGTGTTCAGCTTCCTGATCTGGCCAGTGTGATCCAAGACCCGGTGCAGCTAGAGGCCACACGCCATCTTCTTGATCAT
ATTTGCGAAAACCGATTGAAGAACGTTCTTTCCAAGATAGGGACGGAACCGTTTGAGAAGGAAGTAATTGTTCCGGACCAGTTGGCAACACTCTTGGCCAAAGATGCTC
TGAAGGATTTTCTAAAAGATGTGGATCCATCTGTTGTGAGTGTGCCTGTTTTAACTCGCAAGGATATGGTTCGCTACTTGCTTCTTGAGTCACGACGGCTGGTGTGCAC
GCAGTGGAAAAGCATTTTGAAGCGTCACTCTGCAGAGTCTGAGGCT

>T cruzi (CLBEL) Tc00.1047053506363.110

CTGCGGCGCCACTTTCAGCTGTTTCTTCGTAGGACGCCCCTGTGGCTTGCTGACGATGGTTCTTCTCTTTTTGAGCGTTACACAGAAATTGAGAACTCTAACGAACGAC
GGATGAGTGCATTGAAGGCATGCGGAATGTTTGACGATGAGTGGATTGCCACAGAGAAGGTTCATGGCGCCAACTTTGGGATTTACTCCATCGAAGGTGAGAAAACCAT
ACGTTACGCGAAAAGAAGCGGTATTATGCATCCCACTGAACATTTCTTTGGCTATCATATGTTGGTACCGCAGTTGCAGCAGTACATCACCGTCACCCGTGAGATGCTC
TGTGACAAGTTGCAGAAGAAGTTGCACACTGTTCTTATAAACGGCGAGTTGTTTGGGGGGAAGTACGACCACCCAAGTCTTCAAAAGACGCGAAAAACGGTTATGGTGG
CAGGTAAACCCCGGACAATTAGTGCCGTGCAGACGGATTCGTTTCCTCAGTACAGTCCTGATCTTCATTTTTATGCGTTTGACATCAAATATAAGGAGACTAACGAGGC
TGAATACGTTACTCTCACATTTGATGATGCAACTGAATTGTTTAAGAGGGTTCCTGGATTACTGTATGCGAGGGCCATCATTCGGGGCCCAATGTCGAAGGTGGCGGCC
TTTGATGTGGAGCACTTTGTCACCACCATTCCGCCGCTTGTCGGGATGGGCAACTACCCATTAAAGGGTAATTGGGCAGAGGGACTTGTAGTGAAGCATGCAAAGCGAG
GCACGCCTGGCTTTGACCCCAAGGGATTGACCATTTTGAAGTTTAAGTGCACCGCCTTTCAGGAGATCTCGACAGACCGGCGTCAAGGGCCTCGCGTTGACGAAATGGA
GAGTGTGCGACGGGATTCAATTAGCCGCTCGGGTATTCAGCTCCCTGCTCTGGAAAGCATCATACATGATCCAGTGCAGCTGGAAGCATCGAAGTTTCTTCTGGATCAC

ATCTGTGAGAATCGCCTGAACGCCGTCCTGTCAAAGATTGGAACGGACCCCTTTGAGAAGCAGGAGATGACGCCGGATGATCTGGCGACACTGTTGGCCAAGGATGCCT
TGAAGGATTTTCTCAAGGAGGCGGAACCTGCCATAGTGAACACGCCCATTCTCACTCGCAGGGACATGGCAAGATACGTTCTATTTGAATCGAGGCAGCTGGTCTGTTC
ACGTTGGAAGGCGATTTTACAACGACAGACGGCCGATGTTGTTGAG

>L tarentolae AY148475.1 (NCBI)

CTTCGGCGATGCCTGCTTGCACGGCTCGTACGGCGCTCGCCAGTTCTCTTCTCAGACGAAAAGGAAGCGCTGTTTGAGCGCTATAGTGAGATTGAAAACGCCCACTCGC
GACGCATTGAAGCACTGAAGGACGCGGGTCTCTTCAATGATGAGTGGATCGCTACGGAGAAGGTACATGGTGCCAACTTCGGGATATACTCAACAGAGCATGGGAAGAT
TATTCAGTACGCGAAGCGGAGCGGCATCATGCCGCCACACGAACACTTTTTTGGCTACCACGTTCTCATACCTGATCTCGCAAGGTATATCAAACAAGGGCGAGAACTC
CTTACGGAACAGCTCGGCGTGTCACCCACCACCGTAATCATGAACGGGGAGCTCTTTGGTGGCAAGTATGACCATCCGAGCGTGTCAAAGCGGCGGCAGAGCGTGATGG
TGGCCGGCAAGTCACGCACCATCACAGCCGTGCAGAACAATGCCTTCCCACAGTACTCCCCAGACCTGCATTTCTACGCCTTTGACATCAAGTACCGCATGACCGAAGA
GGAAGAGTACCGCACCATGACGTATGACCAAGCCTTGGCCATCTTCGAGAAAATTCCGGGCCTGCTCTACGCGAGGGCTATCATTCGAGGGCCGCTCAGCAAGGTGGTT
GCGTTCGATGTGGAGAACTTTGCCACGACAATTCCATCGCTGGTGGGCATGGGCGACTACCCGCTCAAGGGTAACTGGGCAGAGGGTCTCGTCGTAAAGCACTGTCGTC
GTGGAGAGTCCGGCTTTGACGCAAAAGGCGTCACTATCATGAAGTTCAAGTGTACCGCCTTCCAGGAGATCTCGACTGATCGCCGCCAGGGGCCTCGCGTAGACGCGCT
GGAGGAGGTACGGCGGCAAGCCATACAAGCCTCTGGCGTCCAGCTGCCGGACATGGAGAGCGTCATCCGCGACTCCGAGGTGCGCAAAGCGGCACATCACCTGCTCAAC
CACGTATGTGAAAATCGCCTACAGAGTGTCCTCTCCAAGATCGGTACAGACCCATTCGAGACGCAGTCTATGACCCCTGCAGAGCTTTCTACGCTGCTCGCAAAGGATG
CGCTGAAGGACTTCTTGAAGGAGGCGGATGCAAAGATCGTCAACACACCGCTACTCGTGCGCCGGGAGATGACCCGCTACGTTCTCTTCGAGGCACGCAAGTACGTTGC
CAGTAGATGGAAAGGCATCGTCGGGCAGCAAAAGGAGGCGGACAGAAAATCGTGC

KREPA3

>T brucei 927 Tb927.8.620

AAGCGTGTTACTTCACATATTTCGCGGTTTTTACCACTTGTGTTGTCGCAGCGTGGCTTGTCGACTTACGCATCACCTCATGTCTCGTCCATGCGTTACTATGGAGCTA
CAAAATGTCTTTTGGCTTCCACACCAGACACTCCCTCCTTTCAGTGCGGCGAATGCGGTAAGGCTTTTCGTCTTATAAATGCCTTAAATCATCACATTATGACCAAGCA
CGCAGGGAAGGCAAAGGCTATGATGAACAAGGGGGGTAAGTTGGAGGAAGTAAACCCTGAAGAAATAAAAAACAAACCGCAAGGGGCAATGTCGCAACCAACATCACCG
CCACCGTCATCTTCGACGTCCGGTACGGAGGCCGCATCAACCTCACCAACGCATTCAAGCTTTCCCGGAATACCGTTTTCACCCGTGGGTGGCATAGGGCTCGTTGGAA
CTCCAGTAGGTGCTGCCTCTCATGTCGTTAATACTACTACCACTGCTGCTAATTCTGCTTCTGGAAATGCCCTTTCCGACGAGAACGCGGACAAAAAAACATTTGTCTG
TACGATTTGTCAGAAAACATTTCGGTTGGAAGCCGCCCTTCAACACCATTACCAGGCGAAGCACAACATGGAGATGCCGACATCATCTTCAAGTTCTGGCGGCGCCAGT
GCTCAACCAGTGCTGCAGGGTGGCGCCACTACAGCAGGTGTTGGAAGTGTGGGTTTTTCCCATACTGAGGAGGAAACTGGCAGATCAGCCATGGGAACACAATATGTTC
ACAGTCAAGAAACAATACTCCCGCAAGCCCCGCAATATCATCTTGATGTCGCGCCGAACGCACCTGAGGAGGGTGAAGTGGCTGCACATTGGCGTTGTGTCAATCATTG
TGTTATGTTGGGCGTGGTGCAAAACATTCAGGAAGGGTTTGTATTTGAAGACAAAGTTCTCCAGTTTACACTCATCACAGACTTCGAGGGACCCTCTCCTGGGGACCCG
GATAAGGACTTTCATACAGTTCGTGTGTTTGATAGCGATTATAGTTCGAGGGTGAAAGAGCAACTCCGTGACGGTGAGTGGTTCCTCGTCACTGGCAGACTGCGCATGG
TGCCTCAATATGATGGTTCCATGCGAAAGTATTACCACTATCCTGTCATACAAGTACACCCGGGCTGTGGGTCAGTGTTGAAGGTG

>T brucei 427 Tb427.08.620

AAGCGTGTTACTTCACATATTTCGCGGTTTTTACCACTTGTGTTGTCGCAGCGTGGCTTGTCGACTTACGCATCACCTCATGTCTCGTCCATGCGTTACTATGGAGCTA
CAAAATGTCTTTTGGCTTCCACACCAGACACTCCCTCCTTTCAGTGCGGCGAGTGCGGTAAGGCTTTTCGTCTTATAAATGCCTTAAATCATCACATTATGACCAAGCA
CGCAGGGAAGGCAAAGGCTATGATGAACAAGGGGGGTAAGTTGGAGGAAGTAAACCCTGAAGAAATAAAAAACAAACCGCAAGGGGCAATGTCGCAACCAACATCACCG
CCACCGTCATCTTCGACGTCCGGTACGGAGGCCGCATCAACCTCACCAACGCATTCAAGCTTTCCCGGAATACCGTTTTCACCCGTGGGTGGCATAGGGCTCGTTGGAA
CTCCANTGGGTGCTGCCTCTCATGTCGTTAATACTACTACCACTGCTGCTAATTCTGCTTCTGGAAATGCCCTTTCCGACGAGAACGCGGACAAAAAAACATTTGTCTG
TACGATTTGTCAGAAAACATTTCGGTTGGAAGCCGCCCTTCAACACCATTACCAGGCGAAGCACAACATGGAGATGCCGACATCATCTTCAAGTTCTGGCGGCGCCAGT
GCTCAACCAGTGCTGCAGGGTGGCGCCACTACAGCAGGTGTTGGAAGTGTGGGTTTTTCCCATACTGAGGAGGAAACTGGCAGATCAGCCATGGGAACACAATATGTTC
ACAGTCAAGAAACAATACTCCCGCAAGCCCCGCAATATCATCTTGATGTCGCGCCGAACGCACCTGAGGAGGGTGAAGTGGCTGCACATTGGCGTTGTGTCAATCATTG
TGTTATGTTGGGCGTGGTGCAAAACATTCAGGAAGGGTTTGTATTTGAAGACAAAGTTCTCCAGTTTACACTCATCACAGACTTCGAGGGACCCTCTCCTGGGGACCCG
GATAAGGACTTTCATACAGTTCGTGTGTTTGATAGCGATTATAGTTCGAGGGTGAAAGAGCAACTCCGTGACGGTGAGTGGTTCCTCGTCACTGGCAGACTGCGCATGG
TGCCTCAATATGATGGTTCCATGCGAAAGTATTACCACTATCCTGTCATACAAGTACACCCGGGCTGTGGGTCAGTGTTGAAGGTG

>T b gambiense Tbg972.8.220

AAGCGTGTTACTTCACATATTTCGCGGTTTTTACCACTTGTGTTGTCGCAGCGTGGCTTGTCGACTTACGCATCACCTCATGTCTCGTCCATACGTTACTATGGAGCTA
CAAAATGTCTTTTGGCTTCCACACCAGACACTCCCTCCTTTCAGTGCGGTGAGTGCGGTAAGGCTTTTCGTCTTATAAATGCCTTAAATCATCACATTATGACCAAGCA
CGCAGGGAAGGCAAAGGCTATGATGAACAAGGGGGGTAAGTTGGAGGAAGTAAACCCTGAAGAAATAAAAAACAAACCGCAAGGGGCAATGTCGCAACCAACATCACCG
CCACCGTCATCTTCGACGTCCGGTACGGAGGCCGCATCAACCTCACCAACGCATTCAAGCTTTCCCGGAATACCGTTTTCACCCGTGGGTGGCATAGGGCTCGTTGGAA
CTCCAGTAGGTGCTGCCTCTCATGTCGTTAATACTACTACCACTGCTGCTAATTCTGCTTCTGGAAATGCCCTTTCCGACGAGAACGCGGACAAAAAAACATTTGTCTG
TACGATTTGTCAGAAAACATTTCGGTTGGAAGCCGCCCTTCAACACCATTACCAGGCGAAGCACAACATGGAGATGCCGACATCATCTTCAAGTTCTGGCGGCGCCAGT
GCTCAACCAGTGCTGCAGGGTGGCGCCACTACAGCAGGTGTTGGAAGTGTGGGTTTTTCCCATACTGAGGAGGAAACTGGCAGATCAGCCATGGGAACACAATATGTTC
ACAGTCAAGAAACAATACTCCCGCAAGCCCCGCAATATCATCTTGATGTCGCGCCGAACGCACCTGAGGAGGGTGAAGTGGCTGCACATTGGCGTTGTGTCAATCATTG
TGTTATGTTGGGCGTGGTGCAAAACATTCAGGAAGGGTTTGTATTTGAAGACAAAGTTCTCCAGTTTACACTCATCACAGACTTCGAGGGACCCTCTCCTGGGGACCCG
GATAAGGACTTTCATACAGTTCGTGTGTTTGATAGCGATTATAGTTCGAGGGTGAAAGAGCAACTCCGTGACGGTGAGTGGTTCCTCGTCACTGGCAGACTGCGCATGG
TGCCTCAATATGATGGTTCCATGCGAAAGTATTACCACTATCCTGTCATACAAGTACACCCGGGCTGTGGGTCAGTGTTGAAGGTG

>T evansi STIB805 (unpublished)

AAGCGTGTTACTTCACATATTTCGCGGTTTTTACCACTTGTGTTGTCGCAGCGTGGCTTGTCGACTTACGCATCACCTCATGTCTCGTCCATACGTTACTATGGAGCTA
CAAAATGTCTTTTGGCTTCCACACCAGACACTCCCTCCTTTCAGTGCGGCGAGTGCGGTAAGGCTTTTCGTCTTATAAATGCCTTAAATCATCACATTATGACCAAGCA
CGCAGGGAAGGCAAAGGCWATGATGAACAAGGGGGGTAAGTTGGAGGAAGTAAACCCTGAAGAAATAAAAAACAAACCGCAAGGGGCAATGTCGCAACCAACATCACCG
CCACCGTCATCTTCGACGTCCGGTACGGAGGCCGCATCAACCTCACCAACGCATTCAAGCTTTCCCGGAATACCGTTTTCACCCGTGGGTGGCATAGGGCTCGTTGGAA
CTCCAGTRGGTGCTGCCTCTCATGTCGTTAATACTACTACCACTGCTGCTAATTCTGCTTCTGGAAATGCCCTTTCCGACGAGAACGCGGACAAAAAAACATTTGTCTG
TACGATTTGTCAGAAAACATTTCGGTTGGAAGCCGCCCTTCAACACCATTACCAGGCRAAGCACAACATGGAGATGCCGACATCATCTTCAAGTTCTRGCGGCGCCAGT
GCTCAACCAGTGCTGCAGGGTGGCGCCACTACAGCAGGTGTTGGAAGTGTGGGTTTTTCCCATACTGAGGAGGAAACTGGCAGATCAGCCATGGGAACACAATATGTTC

ACAGTCAAGAAACAATACTCCCGCAAGCCCCGCAATATCATCTTGATGTCGCGCCGAACGCACCTGAGGAGGGTGAAGTGGCTGCACATTGGCGTTGTGTCAATCATTG
TGTTATGTTGGGCGTGGTGCAAAACATTCAGGAAGGGTTTGTATTTGAAGACAAAGTTCTCCAGTTTACACTCATCACAGACTTCGAGGGACCCTCTCCTGGGGACCCG
GATAAGGACTTTCATACAGTTCGTGTGTTTGATAGCGATTATAGTTCGAGGGTGAAAGAGCAACTCCGTGACGGTGAGTGGTTCCTCGTCACTGGCAGACTGCGCATGG
TGCCTCAATATGATGGTTCCATGCGAAAGTATTACCACTATCCTGTCATACAAGTACACCCGGGCTGTGGGTCAGTGTTGAAGGTG

>T congolense TcIL3000.8.100

TTGTGCGTGAGCTCCCTCGTTCTCCGCCGCGTACGTGCACCGCTTTTATCGCGCTGCCGGGCCGCTCACACATCGTCCCTTTTTTGTTTAAAGCGGTACCACGCAGTGA
CTCCGCGCTTCCTCGCACCAAAGACAGATGCAGCATCTTTTCAGTGCAGTGAGTGCGGAAAGGCCTTTCGTCTGATCAATGCGCTAAACCATCACATTATGACAAAGCA
CGGGGGCAAGGCGAAGGCTATGGTAAATCGAGAGGGCAAACTGGAAGAGGTTGCACCCGATGACGTAACTAATAAACAATCGACAGACATGTTGCAGTCACAGCGACGG
CCGCAGTCCACCGCAAATAAGGGCACGGAGTATACGGCACCATCTCCAACACAGGTGAACATTCCCGGCATGCCGTTCTTCCCCATGTCATCTGTGAACCCCATTGGGA
GTTCGTCTCCTGCTCCGCAAAGTTCTCCTGAGACAGCCGCTTCTACTGCCGCTGGGGACAACACAGTGACGGATGGGGATGCGGACAAAAAGTCGTTTGTTTGCACAAT
CTGCCAAAAAACGTTTCGTTTAGAAGCCGCCCTCCAGCACCACTATCAGGCGAAACACAATATGGAAATGCCCACTGCATCCTCTGGTGGTGGTGGTGGTGCTACCGCA
TCACGGCCTGCTTCAGTGTCCACTTCTTCAGCCTCTGGGGTTATGGGAAGTGCTTTTGGTCACACCGACGACAACGCAAGCAAAACCACGTCTGGTGCGCAATACATCC
ACAGCCAGGAGACTGTGCTTCCACAAGCGCCGCAATACCATCTTGACGTTGCCCCGAACGCTCCGGAGGAGGGGGAGGTGGCTGCCCACTGGCGTTGCGTGAATCACTG
CGTTTTAATGGGCACTGTGCAGAACATTCAGGAGGGCTACGTCTTCGAAGAGAAGGTCCTGCAGTTCACGCTTATCACTGACTTCGAGGGGCCCTCACCAGGAGACCCC
GACAAGGACTTCCACACCGTTCGTGTGTTTGATGGGGAGTACGGCTCGAGAGTAAAGGAGCAGCTGAAAGAGGGGGATCACTTCCTGGTGACTGGGCGGCTCCGCATGG
TACCACAGTACGACCCTTCCATGCGGAAGTATTACCACTACCCAGTTGTGCAGGTGCATCCAGGCTCCGGCTCCGTGCTAAAGGTT

>T vivax TvY486_0800080

AGGAGTATCGCATCCCGCATCTCTCGGCATGCCTCGCAGCTGCTCCCTCAGGACGTGTGCAGCTCACACTTCGCATCTAGAGGAGTACCATTGCGCTATTACCGCGTGA
TGGGGGTTTGGTATGCTCCCAAGCAGGAGGCAGCGATGTATCAGTGTAGTGAGTGCGGTAAAACCTTTCGACTCTTTAATGCTCTAAACCACCACATCATGACAAAGCA
TGCGGGGCAGGCAAAAGCCATGATAAGTAAGGAGGGCACACTGGAGGAAGTCAGGCCAGAAGATGCGCAGCCCAAGCAACAAGCAACGGTACACGCACAATCGGTGGGG
GCAGCGGACACAGGAGGCGCAACCCCTGTCCACTCCGCCTTCCCAGGAATGGCGTGCTCACCCCTCGGCGGACCTGCAGCGTTCGGGGCGCCTTTCAACGTTAAGCAAG
CCGCATCCAATTCAGAGAAAGGAAGCCCTCCCTCTGAATCTGCCGTAAGGGATGATGACGCTGATAAAAAGTTGTTTGTTTGTACAGTGTGCCAAAAAACTTTTAGACT
TGAAGCAGCGTTGCAGCACCACTACCAGGCTAAACACAACATGGAGATGCCTTCGACCTCTACTCATGCTGCATCCTCTTCAAGTGGGTCTGCACGCACTTCAGTGTCG
GGTGGAGGCGGGGGTAACGCCTCAGCGGCACCCAGGCCGTCGGAGGAGTCGGGAAAATCCACAAGTGGCACACAGTATGTTCACAGCCAGGAGACAGTGCTCCCTCAGG
CCCCCCAGTATCACCTGGACGTAGCGCCTAATGCTCCTGAGGAGGGTGAAGTTGCTGCGCACTGGCGCTGTGTAAACCACTGCGTTATGATGGGAAAGGTGCAGAACAT
TCAGGAGGGATATGTTTTTGAGGACAAGGTGGTGCAGTTCACGTTAATTACAGACTTTGAAGGCCCCTCGCCCGGAGACCCCGACAAGGACTTCCATACAGTCCGAGTT
TTTGAAGAAGCTTATAGTGAACAGGTACGGAGTCGGCTGAAGGAAGGAGAATCGTACTTGGTTACAGGGCGCCTGCGTATGGTACCACAATACGATGATTCTATGAAGA
AGTACTACCACTACCCCGTCATTCAGGTCCATCCGGGCTCTGGATCGGTGTTGAAGGTG

>T cruzi (CLBEL) Tc00.1047053510857.40

CGGCGGATCGCATCACATTTGTCACAGCGGACATATCGAGCGTTGTTTCCTCGAGGCGGAGGATTTTACGCCCCGCCGTTCACTATTTCGACTCGCTTCTACCGAGAAA
CGTCGCTGTTGCAGGCCCCAAAGACGGAGACTCCCGCGTACCAATGCGGTGAGTGTGGAAAGACGTTTCGACTGCTTAACGCATTGAATCATCACATTATGACAAAGCA
CGCGGGACACGCGAAGGCGATGGTGCTAAAAGCAGGCAAACTGGAGGAGGTCAAGCCAGAGGAGGCAAAAAACAAGTCATCACATGCTACAGACGCCACGCATGTACCG
TCGACAACAGCAACGGGAGCCACCATGTCAGCCACACAACTTCCAACGGGATTTCCCGGCATGACGTATTCTCCTCTTGGAGGCACCGCGGCATTTGGATCACCTCTTG
GTGTTGCGTCGATGCCCCTGACATCGACGGCGGCAGCAGGTGCTCCCACCAATGCAGGAAAAAATGCTTCCGGGACAGAAAGTTCCGTGGCCGATGAGGATGCCGACAA
GAAACTTTTTGTTTGCACCATTTGCCAGAAGACTTTCCGTTTAGAGGCTGCCCTGCAGCACCACTACCAAGCCAAACATAACATGGAAATGCCGACGAGTTCTACAAGC
GCTAGTAGTACTCGAGCGGCATCTCAAAGCAGCACTAATCTGGGGGCGGCAAACGCCATTTTTGGAGGTGCCACGGATGACGCAGCAAAGTCTACTACCGGCGTTCACT
ATGTTCATAGCCAAGAATCTGTTCTTCCACAGGCACCACAGTATCACCTTGATGTTGCACCTAACGCACCGGAAGAGGGAGATGTGGCAGCTCACTGGCGCTGCGTAAA
CTACTGCGTTGTGCTGGGCCCTGTGCAGAACATTCAGGAGGGCTACGTATTTGAAGAGAAGGTTGTGCAATTTACTTTAATCACAGACTTTGAAGGTCCCTCTCCCGGA
GATCCTGATAAGGACTTTCACACCGTAAGAATTTTTGATAATTCCTTCTGTGAGCAGATGAAAAAAGAACTAAAGGAGGGCGATCGTTTTCTCGTAACTGGTCGCTTGC
GCATGATACCACAGTATGACAGTGCGATGAAGAAGTATTATCACTACCCCGTCATTCAGGTACACCCGGGATGTGGGTCTGTTGTAAAGGTG

>L tarentolae LtaP07.1150

CGACGCATTTCTTTGGCAACATGTCAGCGACGGTGGCTTGCTTCGGCACTGCTCCCCCGCGCCGGGAAGGGAGGAACGCCATGGTCGGGGCATAGATCCATTCTCAATA
CCGCGGCGCCTCTCACGACCGCGATTCGAGACCTGCGCACGACTTCGGTTTTCCGTGCCCCGAAGGATAATGCACAGTATCAGTGCGGCGAATGCGGCAAGACTTTCCG
CCTCGTGAACGCTCTCAACCATCACATTATGACACGCCACGGCAACAATGCCAAGGCACTGATGAAGAAGGACGGCAAGCTGGTCCCTGTAGAGACAGAGCAGCTCAAA
AGTGCGGCGCATGGTGGGTCCTCCTCATCACCCACAGGAACGGCGAGCCCCGCAACTTCAGCTGGGGCTTCCACAGCAGCCTCTTTTTCTCCACTATCTGGCAGCGTCG
CTTCCCCCTTCTCGGTGCCCTTCGCCGCCCCATTTGGTGGAGCCGCCGCCTCCATAGGAGGGGCAGCATCTTCCTCGGTTCCCGGGGCGCTGCAGGCCCCTGCTCCACC
AGCCACCTCGACGACCGGCAGAGTTACCGCAAGCAATAGCGAGAATGGAAACGACACCGGCGGGGCCGTTGAGGAGGCAGAGAAGCGCATGTTCGTCTGCACAGTGTGC
CAGAAAACGTTTCGGCTCGAGGCAGCGCTACAGCACCACTATCAGGCGAAGCACAACATGGACATGCCGACGTCCGCCTCGACTCCCTCTTCTCTCGGTGGCGCCAGCA
CGGCACCGGGGACTGGGGGCGCCTCCGTGCCCGCTCCAGGAGAGCCTGGTGCCAGTGCTGGTGGTAGTGGTACCACGTTTGGTGGCGTGCCTGGCGTAGCTGCAGACGG
TACCGCAGCCCCTGTGAACACTTCTGGCTTTAGTGCTGCGCAGTACGTTCGGCAGCAGGAAGGGGCACTGCCTGACGCACCGCAGTACCACCTAGACGTGGCCCCGAAC
GCCCCAGAGGAGGGCGACATCGCCGCACACTGGCGCTGCGTGAATATATGTGTGCTGATGGGCGACGTGCAGGAAGTAGAGGAGGGGTACGTGTTTGAGGATCATGTGC
TGCAGTTCACTGTTGCGACCGAGTTTGCTACCCCGGCCGCTGGTGACCCTGACATGGACTTCCACACGGTGCGTGTGTACGGGCACGAGTTTTGGGCCCCGTTGAAGGC
GGACGTGCAGAGGGGTGGCCGCTTCCTCGTCACGGGGCGACTACGCATGGTGCCGCAGTTCGATACGCAGCTCAAGAAGTACTACCATTATCCTGTGATACATGTCTTT
GCCGGCACTGGTAACGTGATACGCGTT

Appendix 8

Below is tabulated the read outs from Datamonkey HyPhy analysis (Chapter 4).

Where indicated sites of negative or purifying selection are listed.

REL1

Data summary

6 sequences with **1** partition

These sequences have not been screened for recombination. Selection analyses of alignments with recombinants in them using a single tree may generate misleading results.

Partition **1**: 489 codons 0.842436 subs/site

Inferred rate distribution

Rate class has dN>dS Rate class has dN<dS

Rate Class	1	2	3	4	5	6	7	8	9	Summary
dS	1.41	1.41	1.10	1.10	1.41	0.53	0.53	1.10	0.53	
dN	0.00	0.10	0.00	0.10	0.73	0.00	0.10	0.73	0.73	
dN-dS	-1.41	-1.31	-1.10	-1.00	-0.69	-0.53	-0.43	-0.38	0.19	Mean dN-dS: -0.884
Prob.	0.000	0.000	0.437	0.294	0.000	0.097	0.065	0.088	0.019	Std.Dev : 0.307

Found no positively selected sites (Error: Reference source not found significance level Error: Reference source not found)

Found **298** negatively selected sites (50 significance level)

Codon	E[dS]	E[dN]	Normalized E[dN-dS]	Posterior Probability	Bayes Factor
60	0.977685	0.017089	-0.960596	0.99999	2062.91
62	1.04666	0.0201391	-1.02652	0.99999	1933.62
63	0.978111	0.0838068	-0.894305	0.999739	75.7119
64	0.932416	0.013633	-0.918783	0.999994	3533.01
65	0.98705	0.0190393	-0.968011	0.999967	592.955
66	0.98553	0.0136341	-0.971896	0.999996	5361.86
68	1.00043	0.017936	-0.982498	0.999987	1548.38
69	1.02091	0.0198546	-1.00106	0.999984	1260.03
71	1.02573	0.0181812	-1.00754	0.999988	1646.73
72	1.01257	0.0150179	-0.997557	0.999996	5198.54
73	1.05099	0.0153257	-1.03566	0.999998	7935.23
74	0.983433	0.0826545	-0.900779	0.999731	73.5264
75	1.02271	0.0201401	-1.00257	0.999985	1337.62
77	1.03671	0.019479	-1.01723	0.999983	1132.64
78	0.879629	0.017453	-0.862176	0.999974	770.404
79	1.02491	0.0181794	-1.00673	0.999988	1639.69
80	0.954408	0.0222965	-0.932112	0.99986	141.657
81	1.04705	0.0136345	-1.03341	0.999998	11016.4
82	1.04745	0.0864293	-0.961024	0.999802	99.9187
83	1.02092	0.0194671	-1.00145	0.99998	986.473
84	1.03417	0.0857859	-0.948384	0.999793	95.801
85	0.985583	0.018614	-0.966969	0.999981	1016.72
86	0.990126	0.0171389	-0.972987	0.999987	1570.42
92	1.0442	0.0860313	-0.958173	0.999795	96.468
97	1.0162	0.019602	-0.996599	0.999986	1391.52
99	0.999966	0.0167469	-0.983219	0.999997	6286.31
100	0.962815	0.0209147	-0.9419	0.999916	236.866
101	1.0172	0.0155729	-1.00162	0.999994	3115.02
102	1.00023	0.0142284	-0.986004	0.999999	15413.5
103	1.04376	0.0194697	-1.02429	0.999984	1255.15
104	1.01803	0.0178972	-1.00013	0.999991	2295.51
105	1.02963	0.0209113	-1.00872	0.999956	446.83
106	1.02491	0.0159136	-1.00899	0.999994	3493.92
107	1.01954	0.019134	-1.0004	0.999986	1382.31
108	0.974063	0.0151375	-0.958926	0.999994	3423.68
109	1.01375	0.0176574	-0.996092	0.999992	2439.42
110	0.981681	0.015194	-0.966487	0.999995	3748.09
112	1.03377	0.0168881	-1.01688	0.99999	2054.19
113	1.01278	0.0201416	-0.99264	0.999983	1164.87
114	1.08141	0.0212204	-1.06019	0.999904	206.7
115	0.977096	0.0170284	-0.960068	0.999987	1549.43
116	1.00372	0.0174328	-0.98629	0.999993	2658.45
121	1.04012	0.0252546	-1.01487	0.99989	180.621
124	0.997842	0.0198299	-0.978012	0.999959	478.64
125	0.995284	0.0150189	-0.980265	0.999996	4483.7
126	1.05403	0.015571	-1.03846	0.999996	5454.12
127	1.01548	0.0178982	-0.997578	0.999991	2215.48
128	0.933506	0.0167793	-0.916727	0.99999	1890.97
129	0.957592	0.0151334	-0.942459	0.999997	6209.64
130	1.00713	0.0189437	-0.988189	0.999985	1344.92
131	0.990126	0.0171389	-0.972987	0.999987	1570.42
132	0.999956	0.0138617	-0.986094	0.999998	9300.6

133	0.99515	0.017925	-0.977225	0.999987	1481.92
134	1.02729	0.0136323	-1.01366	0.999997	7806.7
135	0.995294	0.0179033	-0.977391	0.999989	1757.99
136	0.955634	0.0190184	-0.936616	0.999975	790.163
138	0.943996	0.0144313	-0.929564	0.999996	5030.31
139	1.01473	0.0150167	-0.999709	0.999996	5319.55
140	1.01722	0.0189498	-0.998273	0.999986	1438.19
141	0.943618	0.0201382	-0.92348	0.999972	694.986
142	1.00369	0.0157046	-0.987983	0.999994	3308.33
144	0.923862	0.0185463	-0.905315	0.999958	466.149
145	0.987338	0.0172543	-0.970084	0.999986	1433.91
146	1.02275	0.018177	-1.00458	0.999988	1598.74
148	0.995068	0.0146162	-0.980451	0.999997	6165.97
149	1.03715	0.0875544	-0.949595	0.999617	51.5937
150	1.0172	0.0155729	-1.00162	0.999994	3115.02
151	1.0476	0.0198605	-1.02774	0.999988	1686.68
153	1.02132	0.0177375	-1.00358	0.999991	2291.75
155	1.03163	0.0229643	-1.00867	0.999905	208.815
158	1.05482	0.022971	-1.03185	0.999934	301.677
159	1.0411	0.0220932	-1.01901	0.999919	243.958
160	1.04376	0.0181365	-1.02563	0.999992	2553.51
162	0.996471	0.0183798	-0.978091	0.999986	1378.3
163	0.995459	0.0201414	-0.975318	0.99998	984.016
164	0.930889	0.0196163	-0.911273	0.999963	529.759
168	0.936881	0.0169984	-0.919882	0.999988	1603.68
170	1.02615	0.0189686	-1.00718	0.999988	1603.14
171	0.929599	0.0167802	-0.912819	0.999989	1848.65
173	0.961864	0.0209228	-0.940942	0.999916	236.774
174	1.01889	0.0199487	-0.998939	0.999962	525.436
176	1.04173	0.0190703	-1.02266	0.99999	1914.84
177	1.02092	0.0194671	-1.00145	0.99998	986.473
178	0.993492	0.0218749	-0.971617	0.999882	167.643
179	1.00376	0.0150253	-0.988734	0.999996	4759.18
180	0.898841	0.0189114	-0.87993	0.999968	626.83
181	1.02184	0.0155703	-1.00627	0.999994	3337.35
182	0.95563	0.0177262	-0.937903	0.999988	1620.33
183	1.02486	0.0201402	-1.00472	0.999986	1372.61
184	1.0162	0.0179068	-0.998293	0.999991	2204.84
185	1.02007	0.0160888	-1.00398	0.999993	3018
186	1.0411	0.0136322	-1.02747	0.999998	9978.1
187	0.947796	0.0190075	-0.928789	0.999955	435.498
188	0.973804	0.0209083	-0.952896	0.999923	255.435
189	0.969176	0.0136448	-0.955532	0.999996	5150.8
191	1.00037	0.0145447	-0.985823	0.999999	13271
193	1.02107	0.0178972	-1.00317	0.999992	2378.05
194	0.999966	0.0167469	-0.983219	0.999997	6286.31
195	0.994953	0.0142248	-0.980729	0.999999	14699.9
196	1.01589	0.0151312	-1.00075	0.999996	4934.06
198	1.01228	0.0136318	-0.998652	0.999997	6660.49
199	1.02006	0.0178624	-1.0022	0.999991	2312.92
201	0.95563	0.0177262	-0.937903	0.999988	1620.33
204	1.05268	0.0136323	-1.03904	0.999998	12252.4
205	1.00171	0.0171418	-0.984564	0.999989	1722.76
206	1.03882	0.0864757	-0.952344	0.999694	64.5592
207	0.898841	0.0189114	-0.87993	0.999968	626.83
208	0.953261	0.0209089	-0.932352	0.99991	220.422

210	1.0431	0.0159421	-1.02716	0.999993	2843.47
211	1.0162	0.019602	-0.996599	0.999986	1391.52
213	1.02092	0.0194671	-1.00145	0.99998	986.473
214	1.02714	0.0824191	-0.944717	0.99979	94.2225
215	1.00557	0.014823	-0.990749	0.999997	5672.69
216	0.892542	0.0136345	-0.878908	0.999993	2690.67
217	1.01803	0.019591	-0.998436	0.999986	1448.12
218	1.02226	0.0201401	-1.00212	0.999985	1331.29
219	1.02231	0.0147339	-1.00757	0.999999	13685.9
220	1.02729	0.0136323	-1.01366	0.999997	7806.7
221	1.02108	0.0192071	-1.00187	0.999983	1166.27
222	1.0598	0.0214858	-1.03832	0.999944	352.179
223	1.01474	0.0161237	-0.998616	0.999993	2736.89
224	0.943996	0.0144313	-0.929564	0.999996	5030.31
225	0.957563	0.0152246	-0.942338	0.999994	3103.21
226	1.01826	0.0155704	-1.00269	0.999994	3191.7
227	1.01473	0.0150167	-0.999709	0.999996	5319.55
228	1.01374	0.0181696	-0.995573	0.999987	1469.31
230	0.95563	0.0177262	-0.937903	0.999988	1620.33
231	0.994853	0.0201392	-0.974714	0.99998	1010.24
233	0.838563	0.0209394	-0.817624	0.999848	129.795
235	1.04963	0.0190766	-1.03055	0.999991	2199.18
236	1.03956	0.01557	-1.02399	0.999995	4278.61
237	1.01803	0.019207	-0.998821	0.999982	1125.49
239	1.00372	0.0179367	-0.985787	0.999988	1592.76
241	0.945013	0.0209312	-0.924082	0.999907	213.329
242	0.953056	0.0225234	-0.930532	0.999848	130.229
243	1.01608	0.0195954	-0.996489	0.999967	601.645
246	0.944128	0.0176913	-0.926437	0.999984	1215.12
248	1.01257	0.0150179	-0.997557	0.999996	5198.54
250	1.01548	0.0192082	-0.996271	0.999982	1086.59
252	0.993011	0.0826121	-0.910399	0.999892	182.764
253	1.00765	0.0850477	-0.922605	0.999736	74.8415
256	1.07312	0.0136369	-1.05948	0.999999	21585.5
257	1.00206	0.0176226	-0.984438	0.999991	2310.99
258	1.05101	0.0222877	-1.02872	0.999919	244.851
259	1.04741	0.0219644	-1.02544	0.999936	308.06
260	1.00343	0.0201389	-0.983287	0.999982	1100.35
263	1.03112	0.0865042	-0.944611	0.999657	57.6536
265	1.01061	0.0209189	-0.989695	0.999945	359.105
266	0.903617	0.016739	-0.886878	0.999988	1629.7
267	1.01215	0.0193835	-0.992765	0.999982	1108.1
268	1.00457	0.0151314	-0.989434	0.999995	4345.92
269	1.04888	0.0212306	-1.02765	0.999893	184.936
270	1.00561	0.0181839	-0.987429	0.999985	1354.64
272	0.981903	0.0144762	-0.967426	0.999998	10998.3
273	1.06707	0.0217227	-1.04535	0.999926	267.06
274	1.00671	0.0155716	-0.991141	0.999993	2766.08
275	0.957563	0.0152246	-0.942338	0.999994	3103.21
276	0.982323	0.0186437	-0.963679	0.999978	887.781
277	0.838563	0.0209394	-0.817624	0.999848	129.795
278	1.02108	0.0192071	-1.00187	0.999983	1166.27
279	1.00043	0.0174321	-0.983	0.999992	2584.54
280	1.01002	0.0147813	-0.995243	0.999997	6243.45
282	0.83778	0.0151547	-0.822625	0.999989	1748.56
283	1.03189	0.0136361	-1.01826	0.999998	9256.14

284	0.991921	0.0223922	-0.969529	0.99989	179.311
285	0.95911	0.0136438	-0.945467	0.999996	4773.7
286	0.920256	0.0155759	-0.90468	0.999986	1464.08
287	1.00169	0.0217624	-0.979926	0.999899	196.6
288	1.00596	0.0195471	-0.986412	0.999897	192.282
289	1.01763	0.0190529	-0.998581	0.999986	1463.17
290	1.06421	0.0210261	-1.04318	0.999956	454.828
291	1.00137	0.0189531	-0.982413	0.999985	1286.43
292	1.00043	0.0174321	-0.983	0.999992	2584.54
293	1.02271	0.0201401	-1.00257	0.999985	1337.62
294	1.03407	0.0136383	-1.02043	0.999998	9460.1
295	0.982879	0.019061	-0.963818	0.999965	557.962
296	1.02936	0.0197203	-1.00964	0.999979	935.98
297	1.01006	0.0189791	-0.991083	0.999985	1302.99
298	1.01472	0.0179013	-0.996824	0.99999	2081.02
299	1.052	0.0203551	-1.03164	0.999942	339.855
300	1.03846	0.01557	-1.02289	0.999995	4156.18
301	1.02108	0.0192071	-1.00187	0.999983	1166.27
302	0.89964	0.018936	-0.880704	0.99997	665.409
303	1.02609	0.0209053	-1.00519	0.999953	424.147
304	0.912827	0.0209248	-0.891902	0.999889	179.026
305	1.06831	0.0167263	-1.05158	0.999992	2554.38
306	0.998753	0.017385	-0.981368	0.999992	2421.4
307	1.00369	0.0157046	-0.987983	0.999994	3308.33
308	1.04685	0.0209109	-1.02594	0.999966	580.368
310	0.912059	0.0167664	-0.895293	0.999988	1683.76
311	0.919457	0.0189725	-0.900485	0.999971	676.47
312	1.00375	0.0184336	-0.985316	0.999982	1126.38
313	0.979381	0.0136322	-0.965749	0.999996	4711.27
314	0.83856	0.0155758	-0.822984	0.999981	1040.14
316	1.04833	0.0194819	-1.02885	0.999985	1320.29
320	1.00374	0.0209142	-0.982828	0.999941	335.503
322	1.03858	0.0151304	-1.02345	0.999997	6589.43
323	1.04348	0.0163216	-1.02716	0.999991	2327.94
325	1.01803	0.0178972	-1.00013	0.999991	2295.51
326	1.05584	0.0225028	-1.03333	0.999896	189.898
327	0.906639	0.0167955	-0.889844	0.999988	1624.96
328	1.0197	0.0143013	-1.0054	0.999999	17207.7
329	1.00799	0.0144532	-0.993538	0.999998	8007.09
331	1.01371	0.0148128	-0.9989	0.999997	6165.35
332	0.999956	0.0138617	-0.986094	0.999998	9300.6
333	1.04418	0.019477	-1.0247	0.999984	1254.12
334	0.978987	0.0218963	-0.957091	0.999865	146.103
335	0.95563	0.0177262	-0.937903	0.999988	1620.33
336	1.01005	0.0158688	-0.994185	0.999994	3151.18
337	1.02712	0.013633	-1.01348	0.999997	7718.2
338	0.89964	0.018936	-0.880704	0.99997	665.409
339	1.04418	0.0181432	-1.02604	0.999992	2551.26
340	1.01548	0.0195924	-0.995886	0.999986	1397.26
342	1.03125	0.0197134	-1.01154	0.999979	959.743
343	0.982158	0.0225645	-0.959593	0.999872	154.039
344	0.95563	0.0177262	-0.937903	0.999988	1620.33
346	1.00825	0.0151335	-0.993118	0.999996	4514.39
347	0.957563	0.0152246	-0.942338	0.999994	3103.21
349	1.01256	0.0184242	-0.994135	0.999984	1229.51
350	1.03392	0.0151228	-1.0188	0.999997	6166.33

351	0.979077	0.0209232	-0.958154	0.999925	265.527
352	0.924422	0.0167713	-0.90765	0.999989	1797.5
354	1.03849	0.0191398	-1.01935	0.999989	1746.59
355	1.0485	0.0155694	-1.03293	0.999996	4965.01
356	0.90538	0.0224387	-0.882942	0.999804	100.954
358	0.960223	0.0179544	-0.942269	0.999978	893.735
359	0.932631	0.0209293	-0.911701	0.999899	196.867
360	0.931379	0.0186989	-0.912681	0.999973	740.81
361	1.02383	0.0193496	-1.00448	0.999985	1320.55
365	1.01931	0.0209089	-0.998406	0.999949	389.82
366	1.03988	0.015121	-1.02476	0.999997	6565.46
367	1.04986	0.0209111	-1.02895	0.999968	613.249
370	1.01374	0.0181696	-0.995573	0.999987	1469.31
371	0.927043	0.0840966	-0.842946	0.999636	54.2963
372	0.998679	0.084203	-0.914476	0.99973	73.2104
373	1.05398	0.0202327	-1.03375	0.999938	321.711
374	1.02824	0.0136314	-1.01461	0.999997	7807.37
377	1.01548	0.0192082	-0.996271	0.999982	1086.59
378	0.837782	0.0155745	-0.822207	0.99998	1007.66
379	1.02077	0.0155709	-1.0052	0.999994	3262.33
380	0.965294	0.0155744	-0.949719	0.99999	1940.08
381	1.02492	0.0176667	-1.00725	0.999993	2723.44
383	1.03025	0.0228248	-1.00742	0.999912	223.8
384	1.04422	0.0223797	-1.02184	0.999901	199.567
385	1.04668	0.0229286	-1.02376	0.999919	244.699
386	0.944123	0.017196	-0.926927	0.99999	2033.43
387	1.00376	0.0179115	-0.985852	0.999989	1862.64
388	1.03846	0.0209042	-1.01756	0.99996	500.689
389	1.00366	0.0145452	-0.989115	0.999999	13653.2
391	1.00832	0.0190552	-0.989268	0.999985	1328.52
392	1.00428	0.0171459	-0.987131	0.999979	949.312
394	0.944019	0.0143531	-0.929666	0.999998	10409.1
395	1.00376	0.0179115	-0.985852	0.999989	1862.64
396	1.01141	0.0209018	-0.990512	0.999944	354.906
397	1.05682	0.0210739	-1.03575	0.999949	391.321
398	1.01519	0.0162654	-0.998922	0.999993	2763.35
399	1.04418	0.0181432	-1.02604	0.999992	2551.26
400	0.990076	0.0172536	-0.972822	0.999987	1467.26
401	1.00832	0.0189441	-0.989374	0.999985	1361.6
402	1.05437	0.0176988	-1.03667	0.999987	1511.92
403	1.01713	0.0197677	-0.997364	0.999975	785.075
404	0.964302	0.0136326	-0.950669	0.999996	4419.48
405	0.981169	0.022707	-0.958462	0.999858	139.615
406	1.01121	0.0196853	-0.991522	0.999964	551.419
407	1.01166	0.0879883	-0.923673	0.999623	52.4744
408	1.01215	0.0193835	-0.992765	0.999982	1108.1
410	1.03646	0.0209107	-1.01555	0.999959	486.132
412	0.955632	0.0194018	-0.93623	0.99998	1008.71
415	1.01995	0.0209103	-0.999043	0.99995	393.578
418	0.999956	0.0138617	-0.986094	0.999998	9300.6
419	0.819679	0.0184319	-0.801247	0.999934	298.558
420	0.98164	0.0155705	-0.96607	0.999991	2254.03
422	1.01977	0.0181369	-1.00163	0.999988	1631.69
423	1.01512	0.0155718	-0.999552	0.999994	3110.62
424	0.92159	0.0224098	-0.899181	0.99982	109.877
425	1.01803	0.0178972	-1.00013	0.999991	2295.51

426	1.02573	0.0181812	-1.00754	0.999988	1646.73
427	1.01473	0.0150167	-0.999709	0.999996	5319.55
428	0.992816	0.0198679	-0.972948	0.999953	422.674
429	0.95563	0.0177262	-0.937903	0.999988	1620.33
432	0.995281	0.0184251	-0.976856	0.999981	1062.45
436	1.05801	0.0214127	-1.0366	0.999919	243.344
437	0.995149	0.0174215	-0.977728	0.999992	2472.04
438	0.940798	0.0151469	-0.925651	0.999993	2752.54
440	1.07341	0.0209741	-1.05243	0.999913	228.441
442	1.02061	0.0149849	-1.00562	0.999997	5946.51
443	0.986868	0.017447	-0.969421	0.999991	2117.65
444	0.95563	0.0177262	-0.937903	0.999988	1620.33
446	1.05679	0.021315	-1.03547	0.999951	406.479
447	1.04748	0.0170085	-1.03048	0.999991	2193.49
449	1.025	0.0866128	-0.938387	0.999744	77.2172
450	1.02718	0.0209119	-1.00627	0.999954	432.629
452	1.01555	0.0843388	-0.931207	0.999814	106.247
454	0.978786	0.0840384	-0.894747	0.999741	76.4831
457	1.06507	0.021751	-1.04332	0.999905	207.835
462	0.999966	0.0167469	-0.983219	0.999997	6286.31
463	0.95563	0.0177262	-0.937903	0.999988	1620.33
475	1.04418	0.019477	-1.0247	0.999984	1254.12

REL2

Data summary

6 sequences with **1** partition

These sequences have not been screened for recombination. Selection analyses of alignments with recombinants in them using a single tree may generate misleading results.

Partition **1**: 418 codons 0.776479 subs/site

Inferred rate distribution

Rate class has dN>dS Rate class has dN<dS

Rate Class	1	2	3	4	5	6	7	8	9	Summary
dS	1.12	1.12	1.05	1.05	1.12	1.05	0.24	0.24	0.24	
dN	0.01	0.05	0.01	0.05	0.21	0.21	0.01	0.05	0.21	
dN-dS	-1.10	-1.07	-1.04	-1.00	-0.91	-0.84	-0.22	-0.19	-0.02	Mean dN-dS: -0.916
Prob.	0.000	0.000	0.605	0.000	0.000	0.330	0.042	0.000	0.023	Std.Dev : 0.223

No rates with dN>dS were inferred for this datasets, suggesting that all sites are under purifying selection.

KREPA3

Data summary

6 sequences with **1** partition

These sequences have not been screened for recombination. Selection analyses of alignments with recombinants in them using a single tree may generate misleading results.

Partition **1**: 446 codons 1.53592 subs/site

Inferred rate distribution

Rate class has dN>dS Rate class has dN<dS

Rate Class	1	2	3	4	5	6	7	8	9	Summary
dS	1.36	1.29	1.36	1.29	1.36	1.29	0.29	0.29	0.29	
dN	0.01	0.01	0.08	0.08	0.23	0.23	0.01	0.08	0.23	
dN-dS	-1.35	-1.28	-1.27	-1.21	-1.12	-1.06	-0.29	-0.21	-0.06	Mean dN-dS: -0.892
Prob.	0.266	0.000	0.154	0.000	0.246	0.000	0.134	0.077	0.124	Std.Dev : 0.511

No rates with dN>dS were inferred for this datasets, suggesting that all sites are under purifying selection.