High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome

Sam Alsford,[1,6] Daniel J. Turner,[2,3,6] Samson O. Obado,[1,4] Alejandro Sanchez-Flores,[2] Lucy Glover,[1] Matthew Berriman,[2] Christiane Hertz-Fowler,[2,5] and David Horn[1,7]

[1] London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom; [2] The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom

African trypanosomes are major pathogens of humans and livestock and represent a model for studies of unusual protozoal biology. We describe a high-throughput phenotyping approach termed RNA interference (RNAi) target sequencing, or RIT-seq that, using Illumina sequencing, maps fitness-costs associated with RNAi. We scored the abundance of >90,000 integrated RNAi targets recovered from trypanosome libraries before and after induction of RNAi. Data are presented for 7435 protein coding sequences, >99% of a non-redundant set in the *Trypanosoma brucei* genome. Analysis of bloodstream and insect life-cycle stages and differentiated libraries revealed genome-scale knockdown profiles of growth and development, linking thousands of previously uncharacterized and "hypothetical" genes to essential functions. Genes underlying prominent features of trypanosome biology are highlighted, including the constitutive emphasis on post-transcriptional gene expression control, the importance of flagellar motility and glycolysis in the bloodstream, and of carboxylic acid metabolism and phosphorylation during differentiation from the bloodstream to the insect stage. The current data set also provides much needed genetic validation to identify new drug targets. RIT-seq represents a versatile new tool for genome-scale functional analyses and for the exploitation of genome sequence data.

[Supplemental material is available for this article. The sequence data from this study have been submitted to the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under accession no. ERA015558. The mapped RIT-seq data have been submitted to the TriTrypDB (http://tritrypdb.org/)]

With 60 million people at risk, ~30,000 deaths per year, and additional livestock infections, the neglected tropical diseases caused by the African trypanosome, *Trypanosoma brucei*, have a major impact across sub-Saharan Africa (Simarro et al. 2008). These tsetse-fly transmitted protozoan parasites diverged early from the eukaryotic lineage (Sogin et al. 1989) and represent the most genetically tractable organism within the Excavate domain.

Trypanosomatids display "unusual" molecular and biochemical features relative to their hosts and other, more intensively studied, eukaryotes. These include compartmentalized glycolysis (Michels et al. 2006), a single motile flagellum with a paraflagellar rod (Ralston et al. 2009; Portman and Gull 2010), a complex mitochondrial DNA structure known as the kinetoplast (Shlomai 2004), and extensive editing of kinetoplast-encoded mRNA transcripts (Stuart et al. 2005). In the nucleus, transcription is polycistronic and all mRNA transcripts are *trans*-spliced (Martinez-Calvillo et al. 2010). In addition, the African trypanosome undergoes antigenic variation of abundant glycosylphosphatidylinositol (GPI) anchored (Smith and Butikofer 2010) coat proteins, which is important to maintain an infection (Horn and McCulloch 2010). Environmental adaptation associated with transmission via the tsetse fly requires major developmental change in surface architecture and energy metabolism in particular (Fenn and Matthews 2007). In spite of these potential targets, current antitrypanosomal therapies suffer from problems of toxicity and limited efficacy (Wilkinson and Kelly 2009), and there is little prospect of a vaccine due to antigenic variation (Deitsch et al. 2009).

The TriTryp (*T. brucei*, *T. cruzi*, and *Leishmania major*) genome sequences were reported in 2005 (Berriman et al. 2005; El-Sayed et al. 2005a; Ivens et al. 2005) and comparative analysis revealed a high degree of synteny (El-Sayed et al. 2005b). The *T. brucei* genome sequence (Berriman et al. 2005) is a haploid mosaic of the 11 diploid chromosomes and comprises ~3 × 10^7 bp. Protein coding sequences are 1592 bp on average with a GC-content of 50% and intergenic regions are 1279 bp on average. Introns and *cis*-splicing are extremely rare and cotranscribed genes are not generally functionally related. A non-redundant set of protein-coding sequences comprises ~7500 genes with 64% of these annotated "hypothetical." Full exploitation of the TriTryp genome sequences requires the continued development of versatile approaches for genome-scale functional analyses. RNA interference (RNAi) mediated knockdown has proven to be an excellent functional analysis tool for *T. brucei* but has been largely limited to a piecemeal gene-by-gene approach (Subramaniam et al. 2006). We present an RNAi target sequencing (RIT-seq) approach that provides a high-resolution genomic scale readout of fitness following RNAi knockdown. Access to RIT-seq data for almost all *T. brucei* genes and to genetic "profiles" that reflect trypanosome biology should guide future research and facilitate drug-target prioritization efforts.

Present addresses: [3] Oxford Nanopore Technologies, Oxford Science Park, Oxford OX4 4GA, UK; [4] The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA; [5] University of Liverpool, Liverpool L69 3BX, UK.
[6] These authors contributed equally to this work.
[7] Corresponding author.
E-mail david.horn@lshtm.ac.uk.

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.115089.110. Freely available online through the *Genome Research* Open Access option.

Results

High-throughput parallel RNAi target sequencing

Illumina sequencing can be used to survey the representation of genetically distinct cells in a complex population (Langridge et al. 2009; van Opijnen et al. 2009) and we reasoned that this could be applied to RNAi-based phenotyping studies. To improve coverage, and critically, to optimize the reproducibility of knockdown, we used a meganuclease-based system (Glover and Horn 2009) to target RNAi cassettes to a single genomic locus validated for robust expression (Alsford et al. 2005). Briefly, an RNAi plasmid library, containing randomly sheared genomic fragments (Morris et al. 2002), was used to create an inducible library in bloodstream form *T. brucei* (Fig. 1). After transfection, the library was grown under non-inducing and inducing conditions (Fig. 2A) and genomic DNA was isolated from populations that survived induction for the equivalent of approximately nine to 20 population doubling times (see Methods). Adaptor-ligated sequencing libraries were prepared from each genomic DNA sample and used to amplify DNA fragments containing RNAi cassette-insert junctions in semi-specific PCR reactions; one primer was specific for the RNAi vector and the other for the Illumina adapter (Fig. 2B). Size-selected DNA was sequenced with 76 cycle paired-end runs on an Illumina GAII. Illumina sequencing reads containing a nine-base RNAi vector junction sequence were then mapped to the *T. brucei* reference genome using the SSAHA sequence alignment algorithm (Ning et al. 2001). As expected, we found no preferential representation of genomic fragments corresponding to protein-coding or noncoding regions in the uninduced population. Under inducing conditions, cytocidal or cytostatic defects diminish RNAi-target sequence representation relative to the expanding bulk population, revealing "cold spots" that correspond to predicted mRNA sequences (Fig. 3). We could thus deduce knockdown-associated loss-of-fitness phenotypes from these cold spots.

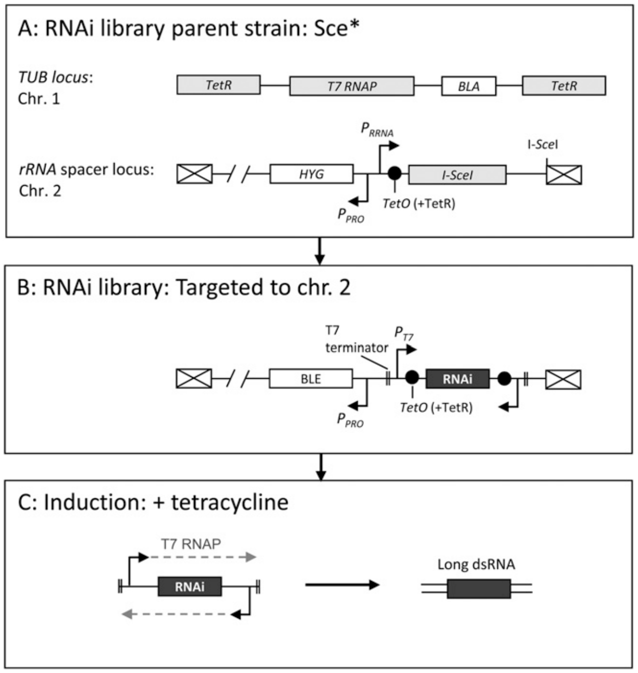

Figure 1. The RNAi library expression system. (*A*) The Sce* strain expresses the tetracycline repressor (TetR) and T7 phage RNA polymerase (RNAP) for the control of dsRNA expression, while inducible homing endonuclease (I-SceI) cleavage facilitates site-specific RNAi library integration at a locus that has been validated for reproducible and robust inducible expression. (*B*) RNAi plasmid library constructs replace the I-SceI gene and cleavage site. The RNAi vector consists of opposing T7 promoters regulated by Tet-operators. The RNAi target fragments serve as a template for the production of dsRNA and also provide unique sequence identifiers for each clonal population. (*C*) Tetracycline induction of dsRNA synthesis. These long (av. ~600 bp) dsRNAs generate a pool of heterogeneous siRNAs that mediate sequence-specific destruction of the cognate mRNA.

We examined four different induced samples (Figs. 2A, 3): bloodstream-form cells grown for three (BF^{D3}) or six days (BF^{D6}), insect/procyclic-form cells (PF), and cells induced throughout growth as bloodstream forms, differentiation and growth as procyclic forms (DIF). In the present study, we focused on protein-coding sequences (CDS) and scored the number of reads mapping to each of these regions in each experiment. This allowed identification of the "core essential genomic loci," where mRNA ablation is associated with loss-of-fitness in all four induced conditions (blue in Figs. 2A, 3), genes associated with more pronounced defects as bloodstream cells (red in Figs. 2A, 3) or as procyclic cells (green in Figs. 2A, 3), and genes associated with a differentiation defect (purple in Figs. 2A, 3).

RIT-seq digital data analysis and validation

We obtained counts of reads mapping to each CDS in each condition. Reads from a pair of replicate sequencing libraries yielded a correlation coefficient of $R^2 = 0.98$ (data not shown), indicating excellent reproducibility for library preparation, amplification, sequencing, and mapping. Examination of these RIT-seq digital data (Supplemental File 1A) readily revealed groups of genes associated with loss-of-fitness in either bloodstream or procyclic cells or loss-of-fitness in both life-cycle stages, and examples of each of these are shown in the scatter-plots in Figure 4A. Several additional groups of genes associated with loss-of-fitness in the BF^{D3} experiment are highlighted in Supplemental Figure S1. The number of reads per CDS from each of the four induced experiments was then compared with the uninduced condition using the DEGseq R package (Wang et al. 2010). We used the *Z*-score for statistical analysis and the Boolean (TRUE | FALSE) reports to find significant changes in reads mapped to each CDS (Supplemental File 1A). In the uninduced population, mapped sequence reads represented >90,000 independent RNAi target fragments, equivalent to >5 targets and >550 reads for each CDS. There are ~7500 non-redundant protein-coding sequences predicted in the assembled *T. brucei* genome and we obtained data for 7435 (>99%) of these. Significant loss-of-fitness was reported for 1908 and 2724 CDS in the 3- and 6-d bloodstream form time-course, respectively, 1972 CDS in the procyclics and 2677 CDS in induced and differentiated cells (see Fig. 4B). All *Z*-scores, in heat-map format, are presented on physical maps of each of the eleven *T. brucei* megabase chromosomes in Supplemental Figure S2.

To further validate the RIT-seq approach, we examined genes associated with the essential function of translation and cohorts of genes that were previously characterized, including potential drug targets (Fig. 4C; Supplemental File 1B). Factors associated with translation, with a previously reported lethal phenotype following RNAi in at least one life-cycle stage, and drug targets all displayed a pronounced mean loss-of-fitness in each of the four experiments. In contrast, genes known to be dispensable for growth in at least one life-cycle stage failed to display a pronounced mean loss-of-fitness in any experiment. Four of these dispensable genes did display

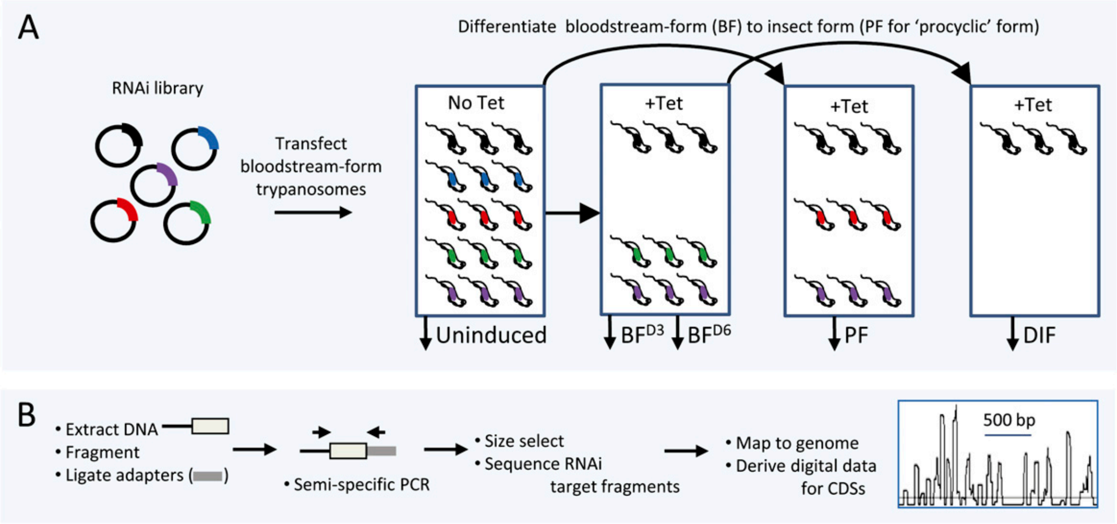
A
Differentiate bloodstream-form (BF) to insect form (PF for 'procyclic' form)
RNAi library
Transfect bloodstream-form trypanosomes
No Tet
+Tet
+Tet
+Tet
Uninduced
BFD3
BFD6
PF
DIF
B
• Extract DNA
• Fragment
• Ligate adapters
• Semi-specific PCR
• Size select
• Sequence RNAi target fragments
• Map to genome
• Derive digital data for CDSs
500 bp

Figure 2. RIT-seq. (*A*) Schematic of the RNAi library and the growth conditions analyzed. The RNAi target fragment provides a unique identifier for each cell and its progeny, while dsRNA production is induced by tetracycline (Tet) addition (Fig. 1). Five different possible outcomes are illustrated. (*B*) Schematic of amplification, sequencing, and mapping of the RNAi target fragments. Only sequences containing a terminal RNAi-vector junction sequence, GCCTCGCGA, were mapped. The box shows a sample region with sequence mapping frequency viewed in Artemis (Carver et al. 2008); each peak represents a unique RNAi target fragment.

a significant loss-of-fitness in the BF^{D3} experiment. However, the greatest reduction in read-count was eightfold compared to uninduced for *RAD51-5* and this is entirely consistent with the decreased growth rate reported for *RAD51-5* null strains (Proudfoot and McCulloch 2005). For each gene, we applied a four-set binary coding system that reflected the output from the DEGseq package in each of the induced experiments (Fig. 4D). This allowed us to identify 750 genes associated with significant loss-of-fitness in all four experiments (0-0-0-0 genes), 437 genes associated with loss-of-fitness in the three experiments involving growth as bloodstream form cells (0-0-1-0 genes), 219 genes associated with loss-of-fitness in the pair of experiments involving growth as procyclic cells (1-1-0-0 genes), and 545 genes associated with loss-of-fitness only in differentiated cells (1-1-1-0 genes).

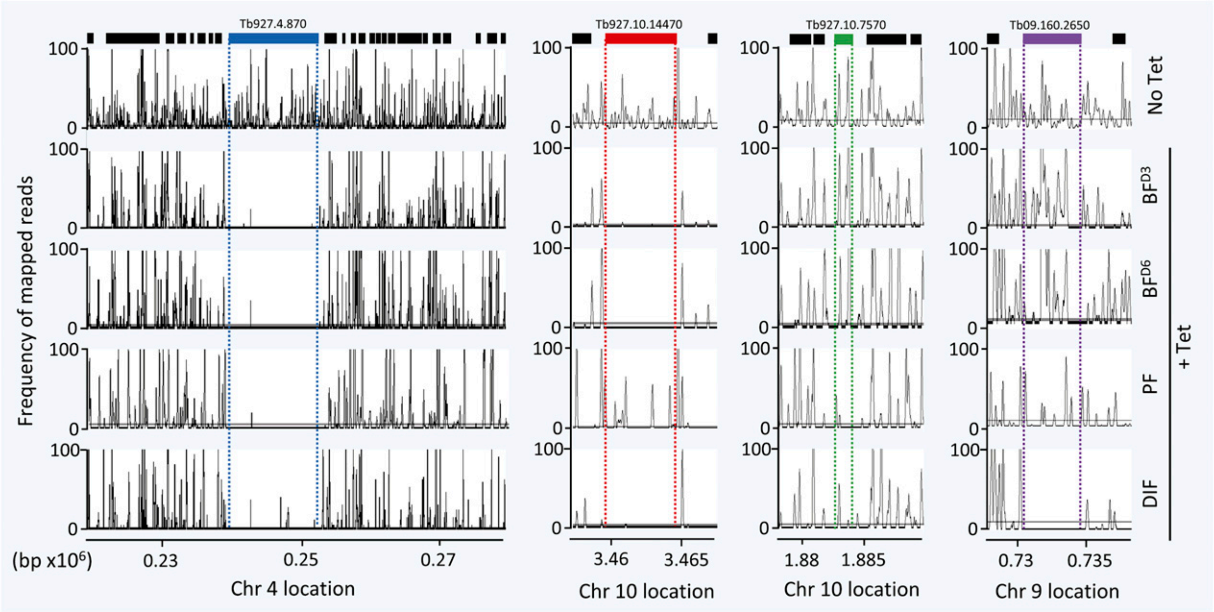

Figure 3. Example plots in Artemis of the four different loss-of-fitness patterns illustrated in Figure 2; mRNA ablation is associated with a defect in all four induced conditions (blue), a more pronounced defect in bloodstream form cells (red) or procyclic cells (green), or a differentiation defect (purple). Cells carrying RNAi target fragments that negatively impact fitness through dsRNA expression and RNAi-mediated ablation are relatively depleted as the population expands and these changes are reported by the depth of sequence coverage relative to the uninduced control (no Tet). Each peak represents a unique RNAi target fragment. The genes shown encode a dynein heavy chain (Tb927.4.870), an intraflagellar transport protein (Tb927.10.14470), dihydrolipoamide acetyltransferase (Tb927.10.7570), and an uncharacterized zinc ion binding protein (Tb09.160.265).

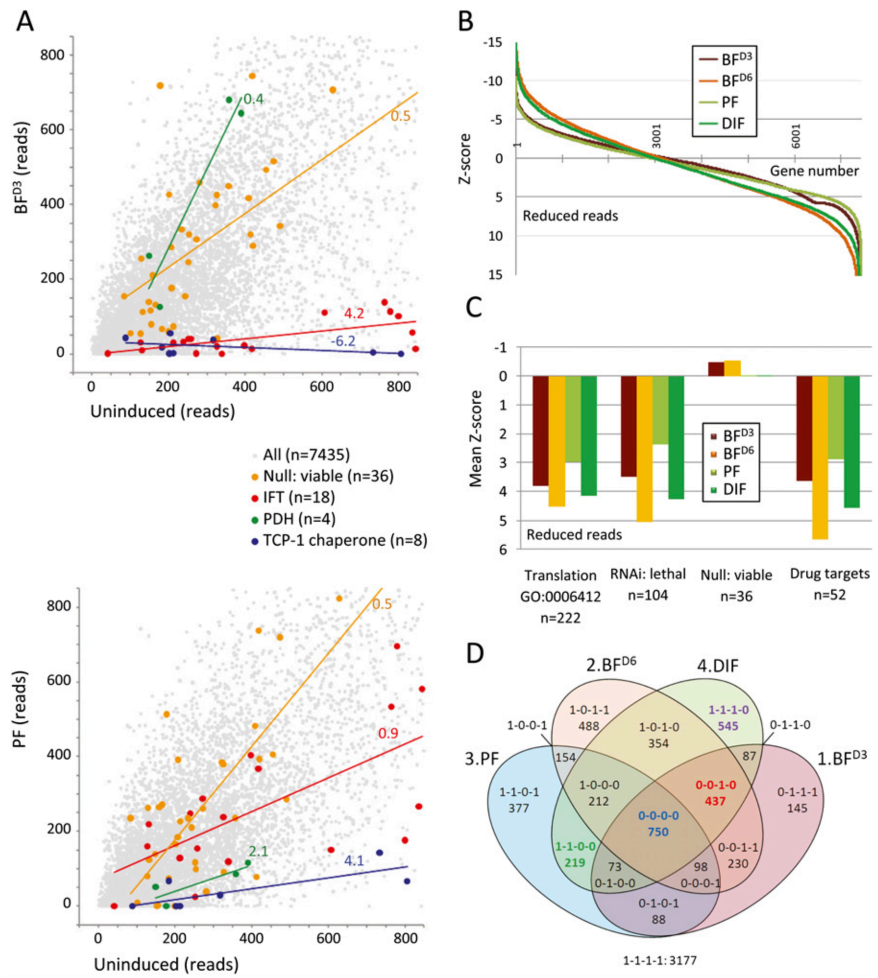

Figure 4. RIT-seq digital data analysis and validation. (*A*) The scatter-plots illustrate groups of genes that display different loss-of-fitness profiles in the BF^{D3} (*upper* plot) and PF experiments (*lower* plot). Trend lines and slope scores are shown. (IFT) Intraflagellar transport (all genes identified to date); (PDH) pyruvate dehydrogenase (four-subunit complex); (TCP-1) T-complex protein 1 (eight-subunit chaperone ring complex); (*n*) number of genes (see Supplemental File 1B for details). (*B*) The distribution of *Z*-scores for 7435 genes and for each of the four experiments based on DEG-seq analysis; a *Z*-score of >3.3 represents a significant loss-of-fitness. (*C*) The plot shows mean *Z*-scores, based on DEG-seq analysis, for genes with the GO-term annotation "translation" and groups of characterized genes for each of the four experiments. Genes were selected based on a combination of TriTrypDB and PubMed searches (See Supplemental File 1B for details). (*D*) The four-set Venn diagram shows the distribution of genes based on DEG-seq analysis. The binary codes score the output from each of the four experiments (1–4) in the order indicated, with "0" representing a significant loss-of-fitness. For example "0-0-1-0" genes are associated with a loss-of-fitness in all three experiments involving RNAi induction during growth as bloodstream-form cells. The DIF experiment involved RNAi induction throughout growth as bloodstream forms, differentiation and growth as procyclic forms, so any gene required for growth in either life-cycle stage or for differentiation should register a loss-of-fitness in this experiment. The four cohorts illustrated in Figure 2A and Figure 3 are highlighted (bold colored text).

The analysis above indicates that the RIT-seq outputs report RNAi induced fitness costs and identify the genes that underlie these defects at a genomic scale. The outputs also provide functional annotation for the high proportion of *T. brucei* genes encoding proteins with no prior functional assignment, many of which are trypanosomatid or species-specific. Among 7435 genes analyzed, 64% are annotated "conserved hypothetical" (4401) or "hypothetical" (388). Of 750 genes that display loss-of-fitness in all experiments, 370 (49%) were derived from these categories. Thus, despite significant under-representation (P < 0.0001), approximately half of essential functions in trypanosomes are encoded by "hypothetical" genes. The output also provides a catalog of new potential drug-target data; 1187 genes displayed a significant loss-of-fitness in all three experiments involving growth as bloodstream-form cells (BF^{D3}, BF^{D6}, and DIF) and can therefore now be considered at least partially validated, genetically, as potential targets. As expected, previously characterized potential drug targets, including several glycolytic enzymes and kinases (Supplemental File 1B), were over-represented (P < 0.0001) in this cohort.

RIT-seq for genome-scale profiles of growth and development

RIT-seq provides data for several thousand individual genes. To study fitness in the context of functions, processes, and compartments, we assessed the data in terms of Gene Ontology (GO) annotations, which provide structured controlled vocabularies to describe gene products (Supplemental File 2). We began by analyzing GO terms from a GO-slim set that displayed highly significant (P < 0.01) over-representation within the cohort of genes displaying loss-of-fitness in all experiments (Fig. 5A). Core essential functions such as "transcription," "translation," and "proteasome" are over-represented in this cohort, as expected. Another notable over-represented term was "RNA binding" (P < 0.0001). Since trypanosomatid genes are arranged in long polycistronic units transcribed from strand-switch regions lacking conventional promoters (Siegel et al. 2009), this output is consistent with the rarity of regulated transcription and the emphasis on post-transcriptional control of gene expression at the level of mRNA stability and translation. Our results identify 40 RNA-binding proteins that appear to be essential for growth in both major life-cycle stages.

We next analyzed the distribution of genes annotated with the term "RNA binding" and other groups of genes thought to contribute to post-transcriptional control of gene expression across the four cohorts highlighted in Figure 4D. This analysis revealed that "RNA binding" was specifically enriched in the 0-0-0-0 core essential function group (Fig. 5B). In contrast, a high proportion of ZC3H-zinc finger proteins, which often bind to the 3′-untranslated regions of mRNA, were specifically associated with stage or differentiation-specific loss-of-fitness. The larger cohort of "zinc ion binding" proteins was only moderately over-represented in the differentiation-specific loss-of-fitness group. These profiles suggested that many ZC3H proteins control mRNA subsets at different stages of the life cycle. There are a large number of kinases encoded in the

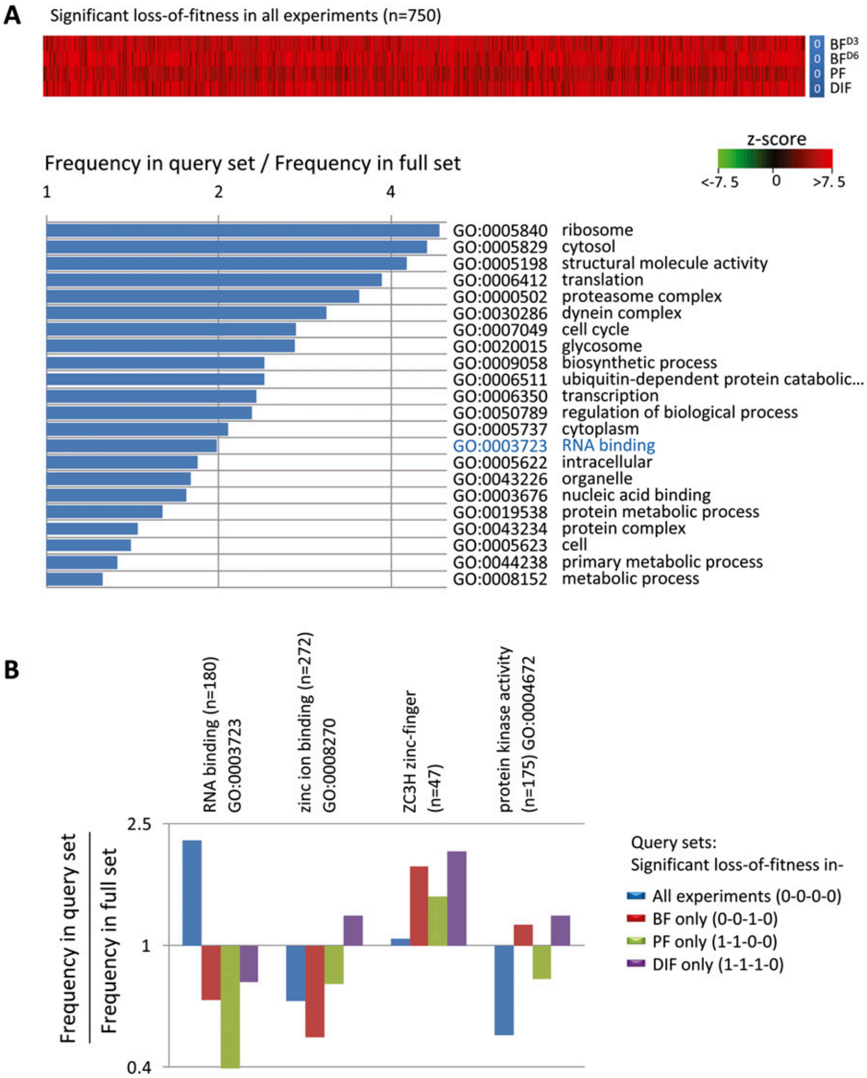

Figure 5. Genetic profile for genes associated with loss-of-fitness in all four RNAi-induced experiments. (*A*) Genes in the 0-0-0-0 group are presented as a *Z*-score heat-map above the associated GO-term profile. All highly significant ($P < 0.01$) associations from the GO-slim set are shown. See Supplemental File 2 for full GO-term analysis. The term shown in blue is analyzed in more detail in B and discussed in the text. (*B*) Distribution of cohorts of genes potentially associated with post-transcriptional control of gene expression. GO terms were used to extract GeneID lists from TritrypDB. See Supplemental File 1B for ZC3H GeneIDs. Binary coding is as described in the legend to Figure 4D.

trypanosome genome and under-representation of "protein kinase activity" ($P < 0.001$) among genes reporting loss-of-fitness in all experiments suggested a high level of redundancy or an emphasis on contingency functions (Fig. 5B). Among this latter group, however, the cdc2-related kinases were commonly associated with a significant loss-of-fitness ($P < 0.002$; see Supplemental Fig. S1J). Additional GO terms that display the range of possible loss-of-fitness profiles are shown in Supplemental Figure S3.

Differentiation from the bloodstream to the insect stage is accompanied by major developmental changes in metabolism, surface architecture, and cell-cycle control, but the genes underlying these differences remain largely unknown. GO terms associated with flagellar motility are strikingly over-represented in the bloodstream loss-of-fitness group and absent in the procyclic loss-of-fitness group (Fig. 6A). This is consistent with reports of the specific impact of flagellar motility defects on bloodstream-form growth (Ralston et al. 2009; Portman and Gull 2010), and we identify intraflagellar transport ($P < 0.0001$; see Fig. 4A) and axonemal radial spoke proteins ($P < 0.0001$) as important players in this differential effect. Both cohorts of genes encoding these factors display a pronounced loss-of-fitness in the bloodstream form (BF^{D3}) relative to the procyclic form (Fig. 6B). Motility in bloodstream-form cells may play a role in cytokinesis or in maintaining the high rate of traffic and nutrient uptake via the flagellar pocket (see Field and Carrington 2009). The proteasome complex is also over-represented in the bloodstream loss-of-fitness group and absent in the procyclic loss-of-fitness group (Fig. 6A; Supplemental Fig. S3). In this case, the α-ring subunits ($P < 0.0001$) are primarily responsible for the effect (Fig. 6B), suggesting a particularly important role for this subcomplex in bloodstream trypanosomes.

Glycolysis operates in peroxisome-like organelles known as glycosomes in trypanosomes and is thought to be the single source of ATP in the bloodstream (Michels et al. 2006), while the process of differentiation to the procyclic-form is accompanied by elaboration of mitochondrial cristae and a switch to oxidative phosphorylation. A constitutive dependence upon the glycosome is highlighted by over-representation of this term ($P < 0.01$) in the GO profile shown in Figure 5A but, consistent with the developmental transition, RNAi against glycolytic enzymes was associated with a more pronounced defect in the bloodstream form (Fig. 6C). In contrast, pyruvate dehydrogenase (PDH) components displayed a prominent association with defects specific to procyclic cells (Fig. 6C, $P < 0.0001$; also see Fig. 4A), suggesting a particularly important role for fatty acid metabolism in the insect mid-gut (see van Weelden et al. 2005). Increased reads associated with knockdown of the PDH complex in bloodstream-form cells may indicate a gain-of-fitness. Gene knockdowns associated with a significant gain-of-fitness in all four experiments number 147, including six genes annotated with the GO-term, ATPase activity ($P < 0.01$, data not shown). These latter genes play roles in contingency functions such as DNA repair (Tb11.02.3110) and toxin export (Tb927.8.2160, Tb11.01.8700) and their expression could retard growth under the conditions tested. Notably, both "mitochondrion" and "integral to membrane" are significantly under-represented in the core essential function and bloodstream loss-of-fitness groups and over-represented in the procyclic loss-of-fitness and DIF groups (Supplemental Fig S3; Supplemental File 2). These outputs are likely associated with the shift to oxidative phosphorylation and surface coat remodeling, respectively, during differentiation to the insect stage.

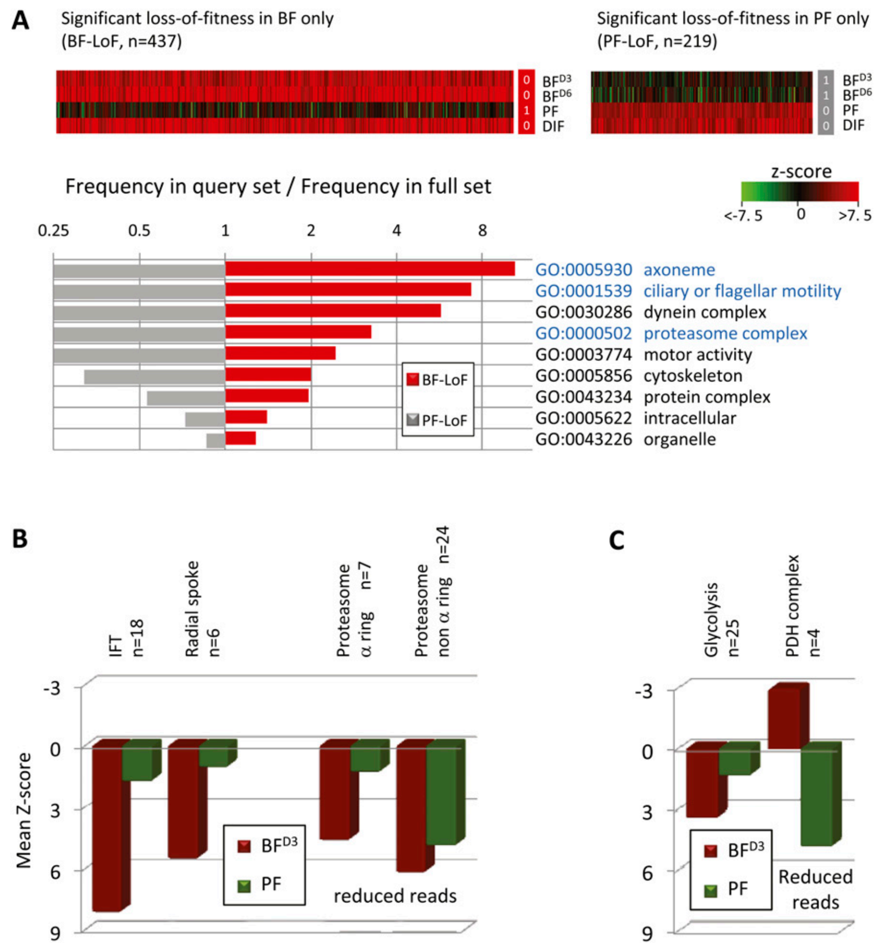

Figure 6. Developmentally regulated genetic profile for genes associated with loss-of-fitness in all three experiments involving the bloodstream form (BF) and compared to the equivalent procyclic form (PF) profile. (*A*) Genes in the BF (0-0-1-0, red bars) and PF (1-1-0-0, gray bars) groups are presented as a *Z*-score heat-map above the associated GO-term profile. All highly significant ($P < 0.01$) associations from the GO-slim set are shown. See Supplemental File 2 for full GO-term analysis. Every term shown is notably under-represented in the PF group. Terms shown in blue are analyzed in more detail in *B* and discussed in the text. (*B*) Mean *Z*-scores for cohorts of genes associated with flagellar motility and the proteasome complex. (IFT) Intraflagellar transport. (*C*) Mean *Z*-scores for cohorts of genes associated with energy metabolism. (PDH) Pyruvate dehydrogenase. (*B–C*) See Supplemental File 1B for details of cohorts of genes analyzed.

The pathways underlying the major developmental changes that accompany adaptation to the insect host environment could be targeted in the mammal to reduce virulence or to block transmission. Tricarboxylic acid cycle intermediates serve as a trigger for this developmental transition and a citrate transporter involved in the process has been identified (Dean et al. 2009), but other factors involved in this environmental sensing remain largely unknown. A GO-term profile of the cohort of genes associated with differentiation (Fig. 7A) revealed over-representation of lipid metabolic process ($P < 0.01$; also see Supplemental Fig. S3), carboxylic acid metabolism ($P < 0.01$), oxidoreductase activity ($P < 0.01$), and phosphorylation ($P < 0.02$). The genes encoding factors implicated in carboxylic acid transport and propagation of this differentiation signal are shown (Fig. 7B). We also present the group of genes for which RNAi confers a specific gain-of-fitness during differentiation (Fig. 7B). This latter group comprises only 15 genes and is validated by the presence of the known differentiation inhibitory kinase, ZFK (Vassella et al. 2001).

Discussion

RNAi libraries and second-generation sequencing have both revolutionized genomics studies. We have combined these tools to map knockdown-associated fitness costs to the predicted protein-coding sequences of the African trypanosome genome. This RIT-seq approach reports a wide dynamic range of RNAi-induced fitness defects at nucleotide-level resolution and at a genomic scale. RIT-seq has several advantages over related approaches. For example, unlike mutagenesis approaches, the knockdown approach is not limited by ploidy and facilitates the study of genes that are essential for growth. In addition, microarray-based approaches can suffer from problems of cross-hybridization artifacts and a narrow dynamic range, whereas high-throughput sequencing has low background noise and a wide dynamic range (Wilhelm et al. 2008; Montgomery et al. 2010). Known trypanosome biology is broadly supported by our RIT-seq data while some findings could be considered unexpected. For example, based on loss-of-fitness profiles, the full set of kinases may not represent a particularly rich source of attractive drug targets, the proteasome α ring may perform a function of particular importance in the bloodstream form, and lipid metabolism may be more important for the differentiation process than previously appreciated.

The RIT-seq data were validated by several means but it is worth considering the parameters that affect data quality, genome coverage, and the origin of false positives or false negatives. Parallel analysis has the advantage that all genes are tested and analyzed in the same controlled environment, i.e., within a single culture flask. DNA fragment amplification is also carried out in parallel and the process of mapping sequences back to the reference genome negates the need to manipulate or monitor large numbers of samples. Importantly, sequences were only mapped and counted if they contained a correctly positioned 9-bp "tag" derived from the RNAi expression cassette. We found that ~20% of these reads failed to map to the assembled reference chromosomes, the majority of which likely derive from repetitive subtelomeric regions (Berriman et al. 2005). More than 99% of the non-redundant gene-set was represented and the data set includes more than five RNAi target fragments per gene on average, indicating excellent genome coverage in the library and the RIT-seq readout. Cloning or amplification artifacts can be a particular problem when manipulating (G+C) biased genomes (Kozarewa et al. 2009) but, in the current case, *T. brucei* protein coding sequences present a (G+C) content of 50%, amplification was kept to a minimum, and quantitative PCR was used for normalization (see Methods). Although the average number of reads per protein coding sequence

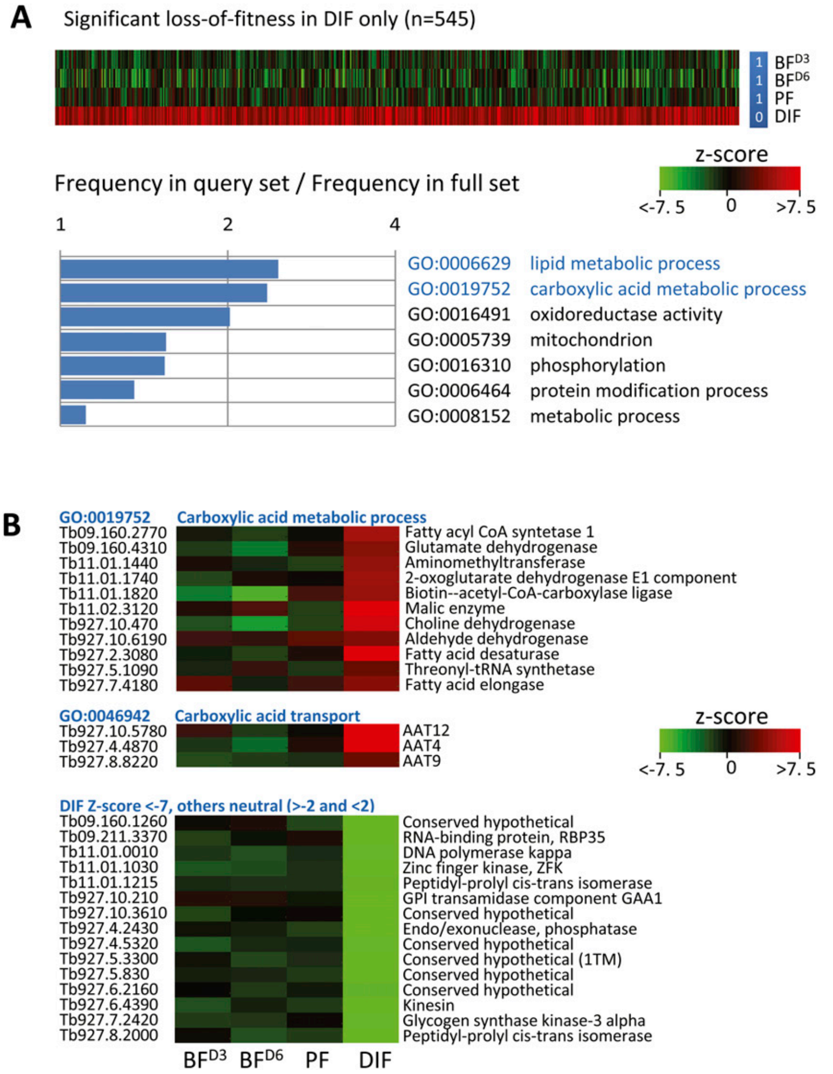

Figure 7. Genetic profile for genes associated with loss-of-fitness only in the differentiation experiment. (*A*) Genes in the 1-1-1-0 group are presented as a *Z*-score heat-map above the associated GO-term profile. All significant ($P < 0.03$) associations from the GO-slim set are shown. See Supplemental File 2 for full GO-term analysis. Terms shown in blue are analyzed in more detail in *B* or discussed in the text. (*B*) Examples of cohorts of genes displaying loss-of-fitness (*upper* panels) or gain-of-fitness (*lower* panel) associated with differentiation. The *upper* panels show the genes in the 1-1-1-0 group that are associated with the GO terms "carboxylic acid metabolic process" or "carboxylic acid transport." (AAT) Amino acid transporter. The *lower* panel shows the genes with a low *Z*-score in the DIF experiment (increased reads) but otherwise neutral *Z*-scores.

was >550, more sequencing would likely increase confidence for short genes with less coverage. Regulated repression of an RNAi cassette can be insufficient to prevent expression of toxic amounts of dsRNA in some cases (Alibu et al. 2005) and this could lead to under-representation of certain essential genes in a *T. brucei* library, but we see no evidence for this problem when using the current system (see α and β-tubulin and the clathrin heavy chain, for example Tb927.1.2340, Tb927.1.2330, and Tb927.10.6050, respectively, in Supplemental File 1A).

Using the RIT-seq approach, false negatives, where a growth defect should be observed but is not detected, could arise due to insufficient dsRNA expression from the RNAi construct. False positives, where a growth defect should not be observed but is detected, could arise primarily through off-target effects. Some such effects can be tolerated here due to fivefold coverage. The use of a library containing RNAi target fragments with a mean length of 600 bp also represents a balance between minimizing false negatives, due to short target fragments with reduced RNAi potency, and minimizing false positives, due to the targeting of two adjacent genes by any single fragment. The long dsRNAs generate a pool of heterogeneous small-interfering RNAs (siRNAs) that mediate sequence-specific destruction of the cognate mRNA. This is thought to increase the knockdown effect and, at the same time, dilute off-target effects of any individual siRNA. To further minimize false negatives due to position effects, we used a meganuclease-based system whereby the vast majority of RNAi cassettes are integrated at a locus validated for reproducible and robust inducible expression (Alsford et al. 2005; Glover and Horn 2009). This approach also minimizes the off-target effects that can occur following integration at unexpected locations in the genome (Motyka et al. 2004). Revertants with a rearranged or otherwise dysfunctional RNAi cassette would not be expected to produce false negatives, because they arise at low frequency and because many will lack the template for sequencing (Chen et al. 2003).

The RIT-seq output reflects the "division of labor" among various protein classes in *T. brucei*. For example, among RNA-binding proteins and protein kinases, 40 (22%) and 9 (5%), respectively, appear to be essential for growth in both major life-cycle stages, while 14 (30%) of the ZC3H zinc-finger proteins are implicated in life-cycle stage specific growth or the process of differentiation. By providing a controlled vocabulary, GO terms allow for a standardized representation of gene product attributes across species and databases. Currently only 38% of *T. brucei* genes are associated with any GO-term annotation but our findings indicate that GO terms are excellent tools for the analysis of RIT-seq data sets. Progress in GO-term annotation, facilitated by new findings, will undoubtedly reveal additional GO terms associated with essential processes in trypanosomes. Meanwhile, groups of genes significantly over-represented in the cohorts analyzed here present opportunities for further research; 38 such associations are shown in Figures 5–7, and an additional 23 ($P < 0.05$) are highlighted in Supplemental File 2.

Although about 80% of yeast genes can be deleted with no obvious phenotypic consequence in rich medium, almost all yeast genes contribute to optimal growth under a variety of stress conditions (Hillenmeyer et al. 2008). In the current study, 25%–37% of *T. brucei* genes appear to be required for optimal growth in each experiment. This is consistent with a chromosome-wide knockdown analysis of genes on *T. brucei* chromosome 1 that revealed growth defects for 23% of those genes (45/197) in the bloodstream form (Subramaniam et al. 2006). By comparing our RIT-seq experiments, we show that knockdown of 10% of genes is associated with a loss-of-fitness in all experiments and that knockdown of 43% of genes is not associated with a loss-of-fitness in any experiment. Consistent with our use of optimized growth conditions, the GO-term, response to stress, is under-represented or absent in all four loss-of-fitness cohorts analyzed above (see Supplemental File 2 and Supplemental Fig. S3). This output also indicates that a 37°C to 27°C temperature shift, used to trigger differentiation, does not constitute a stress for trypanosomes. Thus, like yeast, many trypanosome genes may only contribute to optimal growth under suboptimal conditions.

The current data set will greatly facilitate the identification of attractive drug targets in this neglected-disease pathogen. There are ~100 genes annotated as "lethal" following RNAi in *T. brucei* but more than half of these have only been analyzed in insect-stage cells (Supplemental File 1B). Our RIT-seq analysis reveals more than one thousand genes that can now be considered genetically validated potential targets in bloodstream-form parasites. This will greatly aid prioritization efforts using open-access tools such as those found at http://TDRtargets.org/ (Crowther et al. 2010).

In summary, the RIT-seq output provides genome-scale fitness data under a variety of growth conditions. The GO-term analysis and genetic profiles recapitulate known features, thereby validating the approach, and identify new features. In every case, the genes that underlie essential features of trypanosome biology are now revealed. The current data set and resources should accelerate discovery in understanding the biology of these important pathogens, and should also facilitate drug-target prioritization efforts. Finally, the RIT-seq approach we describe here is versatile and can be widely applied beyond trypanosomes. With a contextual output, responses to a range of environments and selective pressures can be assessed.

Methods

T. brucei growth and manipulation

Bloodstream-form *T. brucei*, MiTat 1.2, clone 221a cells were maintained (Alsford et al. 2005) and transfected by electroporation as described (Glover and Horn 2009) except that cytomix was used for all transfections other than for library generation. Sce* cells were checked for I-SceI-facilitated transfection efficiency and validated for robust tetracycline (Tet)-regulated (1 μg mL^{-1}) *tubulin* RNAi; using this system, we obtained >250,000 transformants/experiment and >90% morphologically FAT cells (Ngo et al. 1998) after 24 h induction. Differentiation to the procyclic-stage was triggered in vitro by transferring the cells to glucose-free DTM medium (Overath et al. 1986) supplemented with citrate and *cis*-aconitate at 27°C. Note that the differentiation process displays some differences dependent upon the strain analyzed (Fenn and Matthews 2007); the current strain is "monomorphic." BF^{D3}, BF^{D6}, PF, and DIF cultures were maintained throughout as populations of >5 × 10^{6} cells and grown under inducing conditions for the equivalent of ~10, 20, 9, and 13 (7 + 6) population doubling (PD) times, respectively (1PD ~7 h for BF cells and ~24 h for PF cells).

Plasmid construction and strain assembly

pT7^{BLA} and pRPa^{Sce*} were used to derive Sce* cells from 2T1 cells (Alsford et al. 2005). pT7^{BLA} replaced a *BLE* gene with a *BLA* gene and added a T7 RNA polymerase gene at the *tubulin*/TetR locus on chromosome 1 while pRPa^{Sce*} added the homing endonuclease gene and cleavage site to the tagged *rRNA* spacer locus on chromosome 2. pT7^{BLA} was constructed by replacing an MluI-BstZ171 fragment and a SmaI-RsrII fragment in pLew13 with a MluI-SmaI fragment from pAZA and a PCR-amplified *BLA* fragment, respectively. pRPa^{Sce*} was constructed by cloning an I-SceI coding sequence in the pRPa^{TAG} vector (Alsford and Horn 2008) and engineering an I-SceI cleavage site at the *Bsp*120I site. pT7^{BLA} was digested with EcoRV-NsiI prior to transfection. Full oligonucleotide details are available on request.

Library construction

The estimated 11× genome-coverage (complexity of 500,000 fragments) RNAi plasmid library (Morris et al. 2002) comprised randomly sheared genomic DNA (mean ~600 bp) cloned in a vector for the Tet-inducible expression of dsRNA. Our ~750,000-clone *T. brucei* library was constructed in the Sce* strain (Fig. 1) using 30 μg plasmid DNA with three transfections using an AMAXA Nucleofector 3 h after Tet-addition (Glover and Horn 2009). Tet was removed during the transfection procedure. To validate the library, we amplified RNAi inserts from a small number of clones and confirmed random distribution of targets across the genome.

DNA sequencing

For paired-end Illumina sequencing, we took ~8 μg of genomic DNA from each sample and prepared standard Illumina sequencing libraries, with a mean insert size of 250 bp, up to the point of adaptor ligation (Bentley et al. 2008). We then normalized these adaptor-ligated libraries by qPCR, by reference to a dilution series of a standard library. Reactions were set up as follows: Ad_T_qPCR1: 5′-CTTTCCCTACACGACGCTCTTC-3′ (desalted), Ad_B_qPCR2: 5′-ATTCCTGCTGAACCGCTCTTC-3′ (desalted), 2× SybrGreen Master Mix, 10 μM Syb_FP5, 10 μM Syb_RP7, 1/100× ligated library DNA in an Applied Biosystems Step One Plus qPCR machine, using the default cycling conditions. To amplify DNA fragments containing the RNAi cassette-insert junction, we used 500 ng of ligated library DNA quantified in this way in semi-specific PCR reactions. We performed 22 cycles of PCR, with a 67.5°C annealing temp and 40 sec extension step, using a forward primer that consisted of a RNAi cassette-specific 3′ end, tailed with the Illumina P5 sequence (bold): TbrF1: 5′-**AATGATACGGCGACCACCGAGATCTACAC**CCTGCAGGAATTCGATATCAAG-3′ and a custom reverse primer that annealed to the Illumina adapter at the 3′ end and which was tailed with the Illumina P7 sequence (bold): V3.3: 5′-**CAAGCAGAAGACGGCATACGAGAT**CGGTACACTCTTTCCCTACACGACGCTCTTCCGATCT- 3′.

Following amplification, we cleaned each product using a spin column (Qiagen) and ran the entire eluate from each column in one lane of a 2% agarose gel at 5V cm^{-1} for 30 min in 1× TBE. We excised a gel band corresponding to a fragment size of ~400 bp and extracted the DNA (Quail et al. 2008). We quantified products by SYBRGreen qPCR (Quail et al. 2009) and sequenced them with a 76 cycle paired-end run, on an Illumina GAII, using a custom sequencing primer for read 1: TbrR1SF1: 5′-GAATTCGATATCAAGCTTGGCCTGTGAG-3′ and the Illumina read 1 sequencing primer for read 2: Illumina read 1: 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′.

Informatics

Paired-end Illumina sequencing reads for each condition were mapped to the *T. brucei* 927 reference genome (Berriman et al.

2005) as follows: Sequences containing a terminal RNAi-vector junction sequence (GCCTCGCGA, one "mismatch" tolerated) were trimmed to remove nine bases from the beginning of read 1 and a similar number of bases from the end of read 2, giving a final length of 67 bases. Reads were then mapped with SSAHA2 (Ning et al. 2001) using a kmer size of 40. After mapping, we obtained coverage and uniqueness plots and, for each CDS, a count of reads mapping uniquely, applying a normalization factor of five million reads for all technical replicates and for all conditions. The number of reads mapping to each CDS in each induced condition was compared with the uninduced condition using the DEGseq R package typically used to normalize transcriptome data sets (Wang et al. 2010). In this case, we carried out pairwise comparisons with each induced sample and the uninduced control and used the *Z*-score and the Boolean (TRUE | FALSE) reports to find significant changes in reads mapped to each CDS. A mechanism used by trypanosomatids to compensate for the general lack of transcriptional control is to duplicate genes in tandem. Since these sequences were excluded from the unique mapping set, we repeated the normalization and DEGseq analysis with the repetitive (raw) read count for all CDS. Uninformative data sets for tandem genes were extracted and replaced with raw data for 157 genes (Supplemental File 1C), each considered to be representative of a tandem. Pseudogenes, genes annotated as "hypothetical unlikely" and repetitive *VSG*, expression-site associated and retrotransposon hotspot gene families were excluded from the analysis. Gene cohorts were assembled and analyzed using TriTrypDB (http://tritrypdb.org/), GeneDB (http://www.genedb.org/), GOToolbox (http://genome.crg.es/GOToolBox/) (Martin et al. 2004), and a Generic GO slim set (http://geneontology.org/) with eleven terms added to capture key features of trypanosome biology. *P*-values were calculated using a χ^2 test.

Acknowledgments

We are extremely grateful to James Morris (Clemson University) and Paul Englund (Johns Hopkins University) for the RNAi plasmid library, to Bernard Dujon (Pasteur Institute, Paris) for the pSCM52 plasmid, and to George Cross (Rockefeller University) for the pLew13 plasmid. We also thank LSHTM colleagues, Tim Perkins for helpful discussions, and John Kelly and Brendan Wren for comments on the draft manuscript. This work was funded by the Wellcome Trust; grant numbers 083648 and 089172 at the LSHTM and 079643 and 085775/Z/08/Z at The Wellcome Trust Sanger Institute.

References

Alibu VP, Storm L, Haile S, Clayton C, Horn D. 2005. A doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei*. *Mol Biochem Parasitol* **139:** 75–82.

Alsford S, Horn D. 2008. Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*. *Mol Biochem Parasitol* **161:** 76–79.

Alsford S, Kawahara T, Glover L, Horn D. 2005. Tagging a *T. brucei RRNA* locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol* **144:** 142–148.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53–59.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309:** 416–422.

Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA. 2008. Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24:** 2672–2676.

Chen Y, Hung CH, Burderer T, Lee GS. 2003. Development of RNA interference revertants in *Trypanosoma brucei* cell lines generated with a double stranded RNA expression construct driven by two opposing promoters. *Mol Biochem Parasitol* **126:** 275–279.

Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, Nwaka S, Ralph SA, Roos DS, Van Voorhis WC, et al. 2010. Identification of attractive drug targets in neglected-disease pathogens using an *in silico* approach. *PLoS Negl Trop Dis* **4:** e804. doi: 10.1371/journal.pntd.0000804.

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

Deitsch KW, Lukehart SA, Stringer JR. 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7:** 493–503.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, et al. 2005a. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309:** 409–415.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, et al. 2005b. Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309:** 404–409.

Fenn K, Matthews KR. 2007. The cell biology of *Trypanosoma brucei* differentiation. *Curr Opin Microbiol* **10:** 539–546.

Field MC, Carrington M. 2009. The trypanosome flagellar pocket. *Nat Rev Microbiol* **7:** 775–786.

Glover L, Horn D. 2009. Site-specific DNA double-strand breaks greatly increase stable transformation efficiency in *Trypanosoma brucei*. *Mol Biochem Parasitol* **166:** 194–197.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, et al. 2008. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* **320:** 362–365.

Horn D, McCulloch R. 2010. Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol* **13:** 700–705.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309:** 436–442.

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* **6:** 291–295.

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, et al. 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* **19:** 2308–2316.

Martin D, Brun C, Remy E, Mouren P, Thieffry D, Jacq B. 2004. GOToolBox: Functional analysis of gene datasets based on Gene Ontology. *Genome Biol* **5:** R101. doi: 10.1186/gb-2004-5-12-r101.

Martinez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martinez LE, Manning-Cela RG, Figueroa-Angulo EE. 2010. Gene expression in trypanosomatid parasites. *J Biomed Biotechnol* **2010:** 525241. doi: 10.1155/2010/525241.

Michels PA, Bringaud F, Herman M, Hannaert V. 2006. Metabolic functions of glycosomes in trypanosomatids. *Biochim Biophys Acta* **1763:** 1463–1477.

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464:** 773–777.

Morris JC, Wang Z, Drew ME, Englund PT. 2002. Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. *EMBO J* **21:** 4429–4438.

Motyka SA, Zhao Z, Gull K, Englund PT. 2004. Integration of pZJM library plasmids into unexpected locations in the *Trypanosoma brucei* genome. *Mol Biochem Parasitol* **134:** 163–167.

Ngo H, Tschudi C, Gull K, Ullu E. 1998. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci* **95:** 14687–14692.

Ning Z, Cox AJ, Mullikin JC. 2001. SSAHA: A fast search method for large DNA databases. *Genome Res* **11:** 1725–1729.

Overath P, Czichos J, Haas C. 1986. The effect of citrate/cis-aconitate on oxidative metabolism during transformation of *Trypanosoma brucei*. *Eur J Biochem* **160:** 175–182.

Portman N, Gull K. 2010. The paraflagellar rod of kinetoplastid parasites: From structure to components and function. *Int J Parasitol* **40:** 135–148.

Proudfoot C, McCulloch R. 2005. Distinct roles for two *RAD51*-related genes in *Trypanosoma brucei* antigenic variation. *Nucleic Acids Res* **33:** 6906–6919.

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5:** 1005–1010.

Quail MA, Swerdlow H, Turner DJ. 2009. Improved protocols for the Illumina genome analyzer sequencing system. *Curr Protoc Hum Genet* **Chapter 18:** Unit 18.12. doi: 10.1002/0471142905.hg1802s62.

Ralston KS, Kabututu ZP, Melehani JH, Oberholzer M, Hill KL. 2009. The *Trypanosoma brucei* flagellum: Moving parasites in new directions. *Annu Rev Microbiol* **63:** 335–362.

Shlomai J. 2004. The structure and replication of kinetoplast DNA. *Curr Mol Med* **4:** 623–647.

Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA. 2009. Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* **23:** 1063–1076.

Simarro PP, Jannin J, Cattand P. 2008. Eliminating human African trypanosomiasis: Where do we stand and what comes next? *PLoS Med* **5:** e55. doi: 10.1371/journal.pmed.0050055.

Smith TK, Butikofer P. 2010. Lipid metabolism in *Trypanosoma brucei*. *Mol Biochem Parasitol* **172:** 66–79.

Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA. 1989. Phylogenetic meaning of the kingdom concept: An unusual ribosomal RNA from *Giardia lamblia*. *Science* **243:** 75–77.

Stuart KD, Schnaufer A, Ernst NL, Panigrahi AK. 2005. Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* **30:** 97–105.

Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D, Field MC. 2006. Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. *Eukaryot Cell* **5:** 1539–1549.

van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6:** 767–772.

van Weelden SW, van Hellemond JJ, Opperdoes FR, Tielens AG. 2005. New functions for parts of the Krebs cycle in procyclic *Trypanosoma brucei*, a cycle not operating as a cycle. *J Biol Chem* **280:** 12451–12460.

Vassella E, Kramer R, Turner CM, Wankell M, Modes C, van den Bogaard M, Boshart M. 2001. Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of *Trypanosoma brucei*. *Mol Microbiol* **41:** 33–46.

Wang L, Feng Z, Wang X, Zhang X. 2010. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26:** 136–138.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453:** 1239–1243.

Wilkinson SR, Kelly JM. 2009. Trypanocidal drugs: Mechanisms, resistance and new targets. *Expert Rev Mol Med* **11:** e31. doi: 10.1017/S1462399409001252.

Received September 9, 2010; accepted in revised form February 8, 2011.

CSH PRESS

GENOME RESEARCH

High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome

Sam Alsford, Daniel J. Turner, Samson O. Obado, et al.

Genome Res. 2011 21: 915-924 originally published online March 1, 2011
Access the most recent version at doi:10.1101/gr.115089.110