Tools for genetic analysis in *Trypanosoma brucei* © George A.M. Cross

Introduction

Although this commentary and web page might not end up being either a comprehensive or an unbiased tutorial on genetic analysis in *T. brucei*, I hope you will find it useful. One major advantage of electronic publication is that it can be updated continuously. In fact, I might make this commentary available before I consider it complete. You can make suggestions to add to it, or contest opinions expressed in it, on our nascent TrypChat BLOG or by writing directly to me. Significant insights that are incorporated will be credited to their sources. Some opinions on transfection efficiency and genetic approaches may be biased by the fact that my laboratory mainly works with bloodstream-form (BF) trypanosomes, except where procyclic forms (PF) are used for specific reasons, including their intrinsic scientific interest, the need to grow large quantities of tryps in vitro, or where morphology is better preserved for microscopy, etc.

Forward genetics

Although it is possible to make classical genetic crosses between *T. brucei* strains, by co-transmission through *Glossina*, this is not a facile technique, it is not widely available, and the mechanism of genetic exchange (which often results in non-diploid progeny) is not understood. This is the subject of a brief 2008 review [18]. Although 'classical' genetics has potential utility for identifying genes responsible for specific and potentially multigenic traits (human infectivity or drug resistance would be obvious examples), and it is useful to have alternative approaches for identifying non-obvious genes responsible for specific functions or traits, there are alternative techniques for 'forward' genetics (RNAi libraries, transposon mutagenesis and — in principal but not so simple — chemical mutagenesis followed by genetic complementation), some of which will be discussed in more detail below.

Reverse genetics

Unlike some of its kinetoplastid cousins, *T. brucei* has several properties (listed below) that make it very amenable to genetic manipulation. Some manipulations require the use of transgenic cell lines that have been pre-engineered to express exogenous RNA polymerases and/or repressors.

1. *T. brucei* is easy to culture and clone on agar plates or in liquid media (see culture commentary and culture protocols on our lab web site)
2. It can be efficiently transfected by electroporation-based methods.
3. Homologous recombination is efficient and non-specific recombination does not appear to be a significant obstacle to gene-specific targeting that can be used to delete or epitope-tag endogenous genes, or perform other manipulations.
4. The Tet operator and TN10 Tet repressor combination can be used to efficiently and tightly regulate transcription mediated by endogenous Pol I, or by introduced bacteriophage RNA polymerases, allowing for regulated gene expression (including conditional gene knockouts) at several levels. Expression is induced by adding tetracycline and can be regulated to some extent by varying the tetracycline concentration (tetracycline probably accumulates in the cell, even when added at a low concentration). Expression levels can also be regulated by varying the promoter sequence, especially when using the T7 promoter, where well-characterized mutations can be used to vary transcription efficiency [39], or by using EP procyclin promoter mutants [1], but can also be strongly modulated by the choice of spliced-leader signal sequence [32] or by using 3´ UTR sequences that regulate mRNA stability [13, 16].
5. *T. brucei* has an intrinsic RNAi system that can be exploited for mRNA knockdown.
6. Up to six drug-resistance markers are readily available, together with markers for negative selection. Additional markers could probably be found, but the inclusion of multiple drugs in the culture cocktail is probably not a good idea.
7. The 'cre-lox' system works very efficiently and can be used for various purposes, principally to allow a drug-resistance marker to be removed and re-used, or to tag genes without materially affecting regulatory sequences in upstream or downstream sequences and mRNA UTRs.

How should you evaluate whether your favorite or candidate gene is essential?

Introduction

The first thing most people want to do when they identify a gene of interest is to know whether it is essential. It is likely that 50% of *T. brucei* genes are not essential under standard laboratory culture conditions. Deletion of many genes will not even affect growth rates in vitro, some genes may be essential in one life-cycle stage but not others (there are at least five clearly recognizable stages in the life-cycle of *T. brucei*), or may be essential for growth in the animal host or insect vector, but not in culture (as in the case of the pyrimidine synthesis pathway; see below), or there may be strain-specific differences that are not reflected in the strain that is predominantly used for *T. brucei* genetics. Deletion of genes involved in recombination or DNA repair may not have significant consequences in vitro, but may be critical in vivo.

Is your gene essential?

The way most people initially test whether a gene is essential, or if an obvious phenotype can assign a probable function, is to knock down expression by inducible RNA interference. RNAi has advantages (it is an easy test: in its simplest form it requires one PCR reaction and cloning step, followed by one transfection and selection step) and disadvantages (a negative result does not mean a gene is not essential, because knockdown may not be effective, or because clones that are inefficient in RNAi knockdown may be self-selected after transfection, or that the RNAi phenotype is unstable). RNAi approaches are discussed in more detail below.

The alternative approach to test whether a gene is essential is to attempt to delete both alleles (assuming there is a single gene copy or a single tandem gene array). Specific strategies for gene deletion are discussed below. The two advantages of this approach are that if the gene is not essential the result is unequivocal and, if its deletion causes a non-lethal phenotype, the cell line will be stable, in contrast to many RNAi or conditional knockout cell lines. In some cases, in the experience of several labs, deleting the second allele can be significantly more difficult than deleting the first, which can lead to the impression that the gene is essential. This might be because the cells need to adjust the expression of other genes to compensate for the loss of the knockout target. The obvious controls are to show that each allele can be deleted independently, with either of the drug-resistance knockout constructs. The best control is to perform a conditional knockout, where the first allele is deleted and an ectopic inducible copy is inserted in a tightly regulated pLEW100-type vector prior to deletion of the second endogenous allele. A conditional knockout involves more work than using RNAi as the initial approach, but it has benefits for subsequent regulation of an essential gene, due to the more precise control over expression. However, some of these benefits can be obtained by using a more tightly regulated RNAi vector than those that are commonly used (see below).

Strategies for gene deletion

With the dramatic decline in the cost of oligonucleotide synthesis, the first choice for making gene knockout constructs is generally to use PCR primers that contain up to 80 nt of the sequences flanking a coding sequence to amplify a drug-resistance cassette, then transfect the product. This approach is also widely used for epitope-tagging endogenous genes (see below). It often works, but results can vary dramatically among target loci, and efficiency is generally much lower than when longer targeting sequences are used. The relationship between target length and targeting efficiency has been systematically explored in one study [5] but only for one target. Anecdotal information from many laboratories, including ours, indicates that variation in efficiency is a significant problem for PCR-based targeting and that deleting the second allele can be especially difficult. A more specific example from [27] is worth quoting: *"[PCR-mediated tagging] works with vastly different efficiencies for different genes. In one particular case, we were completely unable to tag one of two tandemly arranged gene copies by PCR- tagging. Four transformations using 15–18 μg of PCR product per transformation did not result in a single correct transformant. Using the same tagging vector with larger, PCR-amplified 1.3 kb 5´ and 1 kb 3´UTR targeting fragments instead of the 80-mer oligonucleotides, resulted in high efficiency tagging of this previously refractory gene copy. One transformation using 13 μg of digested plasmid DNA resulted in >200 transformants. While this gene was apparently refractory to PCR-mediated tagging, its neighbouring homologue could be readily and reproducibly tagged with this technique".*

The more reliable but more costly and time-consuming (time is money) approach is to PCR-amplify longer (500 to 1,000 bp) regions upstream and downstream of the target coding sequence, clone these into a plasmid containing the drug-resistance cassette, then cut out and transfect the chimeric construct. This approach requires two cloning and DNA purification steps, in contrast to no cloning in the pure PCR-oligonucleotide approach. However, if you are likely to be interested in tagging your gene (see below), the tag could be introduced into the ectopic regulated copy at this stage, thereby saving time later.

When using either approach, it makes sense to use 'floxed' resistance markers (see below, and the separate web page), so drug resistance markers can be re-used. Apart from re-using markers, another advantage when having to assemble a knockout cassette by cloning is that only one version is necessary.

Strategies for regulated gene expression and conditional gene knockouts

Cell lines for regulated expression

Regulated gene expression has prerequisites: it has to be done in a transgenic cell line that expresses the Tet repressor and, depending on the vector, the T7 RNA polymerase. For these reasons, most regulated expression and RNAi strategies are performed in the Lister 427 'single-marker' (SM) BF cell line (maintained in G418) or PF 29-13 cells (maintained in G418 and hygromycin) expressing Tet repressor (TetR) and T7 RNA polymerase (T7RNAP) that were created by Liz Wirtz, in the late 1990s. These same cell lines can also be used for conditional knockouts of specific genes, subject to the availability of an adequate number of selectable markers (see 'the marker problem' below), if advantage has not been taken of using the cre-lox system for marker re-use.

Transcription equivalent to about twice Pol II read-through appeared to be necessary for adequate TetR expression, which led Liz Wirtz to use a T7 promoter, reduced to 10% of wild-type activity by a -10 A to T point mutation, to express TetR, in a T7RNAP background. This approach could possibly have been avoided by using a more efficient trans-splicing signal sequence, or by other means.

The single-marker cell line is available as BF only. It proved impossible for Liz Wirtz, for unknown reasons, to make an equivalent PF line. Although BF cells can be differentiated to PF and easily established as a stable cell lines, the re-transfection efficiency of recently differentiated PF is much lower than our long-established PF line, for unknown reasons. The SM and 'double marker' (13-90) BF (similar to PF 29-13) also lost regulation when differentiated to PF, for unknown reasons.

IT IS IMPORTANT to maintain cell lines in the appropriate drugs, to prevent loss of the various components of the expression system. This means 2.5 µg/ml G418 for the SM line and 15 µg/ml G418 plus 50 µg/ml hygromycin for 29-13 PF. Long-term subculture should be avoided, as is true for most cell lines.

We are currently (2008) making 'zero-marker' (ZM) TetR- and T7RNAP-expressing cell lines that we expect to be stable in the absence of drug selection, which will be a major improvement. These lines will not require continuing drug selection and they do not contain extraneous plasmid sequences.

These cell lines have several disadvantages that we and other labs have tried or are trying to eliminate. Lister 427 does not always always differentiate readily (let alone synchronously) from BF to PF in vitro. Clones expressing different VSGs (probably from different expression sites) seem to differ in their ability to differentiate, but this has not been explored systematically, and individual investigator's results also vary greatly. The other serious but misconceived disadvantage is that Lister 427 has long been thought to be incapable of completing the life cycle in *Glossina*, despite considerable evidence to the contrary, albeit unpublished until recently [30], except as detailed on our web site. In collaboration with Serap Aksoy at Yale, we also showed that Lister 427 'clone 3' retained transmissibility, as confirmed by Peacock et al. After transmission by Wendy Gibson in 1992, Lister 'clone 3' was re-established in culture by us in 1994 (a detailed pedigree is available at http://tryps.rockefeller.edu/DocumentsGlobal/lineage_Lister427.pdf). It is not clear, however, if transmissibility through *Glossina* will be a stable phenotype, as lines are repeatedly transfected, recloned and subcultured.

One approach has been to engineer transgenic lines of supposedly (more about this elsewhere) more biologically 'natural' clones and strains of *T. brucei*. Unfortunately, such strains require some degree of adaptation to grow efficiently in culture, which might eliminate the supposed differences between these 'wild-type' strains and Lister 427. For good reasons, one consortium [1] decided to engineer the TetR and T7RNAP into TREU 927, the so-called 'genome strain' [35]. Due to a strain mix-up, however, the cells

engineered by Alibu et al. 2005 and the TetR line in van Deursen et al. 2001 was STIB 247, not TREU 927 (personal communication from the principal investigators). The 2005 paper also adds confusion by referring to Lister 427 as STIB 427.

Expression vectors and insertion sites

Most of the expression cassettes that we have made and use were designed to integrate into an rRNA spacer region, and use either T7 RNAP or a *T. brucei* Pol-I promoter to drive expression of the gene of interest.

pLEW20 is the prototypical regulated single-promoter (GPEETp) expression vector. In the prototypical pLEW100 dual-promoter vector, a T7 promoter drives the selectable marker and TetR-regulated GPEET promoter drives the gene of interest (incidentally, there do not appear to be any significant differences between the RP and GPEET flavors of procyclin promoters). The dual-promoter vectors typically give 1,000 to 10,000-fold induction and about two-fold lower induced reporter levels and five-fold lower background levels than pLEW20, which means an uninduced level of less than one molecule of Luciferase per cell. 100 ng/ml tetracycline gives maximum induction.

pLEW82 has a regulated T7 promoter and gives the highest expression levels, about 2- to 5-fold higher than pLEW20 in PF and 20- to 50-fold higher than the pLEW100 PARP promoter in BF (see figure 5 in [39]). A more recent version of this vector is pUB39 [8], in which VSG UTRs lead to 2- to 3-fold higher expression in BF, the maximum that has been obtained.

Once you have integrated your expression cassette, it is essential to maintain phleomycin selection to prevent its loss. This represents a conundrum with the original pLEW82-type vectors, where a low level of tetracycline must also be included to maintain phleomycin resistance, but this is unnecessary in the improved version — pLEW82v4. pLEW100 is a better vector with a far lower background (uninduced) expression level and high regulation. It is generally harder, however, to obtain transformants with pLEW100 than with pLEW82, for reasons we can guess, supported by some evidence. It is possible that the pLEW100 cassette is not tolerated in all rRNA loci. Once integrated, the pLEW100 cassette is easily deleted in the absence of continuing drug selection. The pLEW100 cassette also appears to persist somewhat longer in cells after transfection, since cells in which the cassette has not integrated take a few days longer to die during phleomycin selection.

FOR THESE AND OTHER REASONS, we have made several improvements to pLEW100 and pLEW82. pLEW100 versions with different drug-resistance genes, a TetR-regulated rRNA promoter, three Tet operators and two T7 terminators, and a version in which the T7 promoter is replaced by a GPEET promoter to drive HYG, give efficient integration and high-level highly regulated expression (I wrote this before becoming aware of a problem, which we have yet to solve, with the luciferase levels achievable with the pLEW100 version containing adjacent rRNA and GPEET promoters). pLEW82v4 allows selection for integration in the absence of tetracycline, but remains very leaky, compared to pLEW100 vectors.

Other investigators [37] have proposed that vector integration into minichromosomal 177-bp repeats reduces leakiness, but the best regulation in the cited study was far worse than we see with the rRNA spacer, and we have not found any advantage in minichromosome targeting, even for the dual opposing T7 promoter plasmids used for RNAi. Indeed, the 177-bp targeting efficiency appeared much lower than for the rRNA locus, especially when the selectable marker was BLE rather than HYG (our unpublished data), suggesting that the minichromosomes are refractory to transcription (silenced), which might be unsurprising and might also explain why the frequency of targeting appeared to be independent of the copy number of the target sequence [38]. It has been shown that, as known in other organisms, that recombination is more efficient at a transcribed *T. brucei* locus [2].

Another approach to reducing background expression from vectors using a tet-regulated T7 promoter was to make cell lines in which expression of the T7RNAP was also tet-regulated [1].

There is one long-recognized issue in rDNA targeting, which is that not all targeting to the 'non-transcribed spacer' alleles gives equal results. Different alleles are probably not equally silent, as appears to be indicated by some recent experiments by Nicolai Siegel, in our lab (unpublished data). There are nine (haploid number) copies of the rRNA genes dispersed among chromosomes I, II, III and VII in the TREU 927 genome. These loci, except the single copy on Chr I, which is perhaps dangerously close to a

chromosome end, are probably conserved in Lister 427, although the total number of copies may not be. Growth rates of some organisms have been linked to the number of rRNA genes, and it is possible that the more virulent Lister 427 strain of *T. brucei* could have amplified its rDNA. Different rRNA spacer loci are flanked by different genes and, although we have not demonstrated a problem at these target loci, we do know that read-through transcription from a *T. brucei* Pol-I promoter or a full-strength T7 promoter inserted at other loci can give unexpected results [25] [10]) These examples also suggest a possible approach to using randomized over-expression for forward genetics.

On average, five clones from a single transfection give us three that are very tightly regulated (up to 10,000-fold, but averaging about 4,000-fold) and two that perform very poorly. Alsford & Horn took an approach to resolving this problem that I originally suggested to another lab in 1999 (!), which was to tag one rRNA allele that gave low background and use this cell line to target subsequent constructs to the same allele [4]. The resulting BF cell line (2T1) has a re-usable target incorporated into one of the two rRNA spacer regions on chromosome 2a. As Alsford & Horn pointed out in a subsequent paper [3], "An additional, unexpected but welcome outcome was increased homologous recombination efficiency stimulated by transcription of the target". This finely targeted approach is a major advantage for 'high throughput' gene screening by RNAi, for example, where the scale precludes the evaluation of multiple clones for each gene [34].

One often ignored but proven factor that can affect the background level expression level from TetR-regulated vectors is the presence of small amounts of tetracycline in bovine serum, especially in the USA, where tetracycline is routinely included in animal food, which is not permitted in some countries.

Gene expression using endogenous Pol-II read-through transcription

For constitutive expression of a tagged gene or an alien gene (or for investigating recombination parameters [5]) we and other labs have often targeted the Tubulin locus on chromosome I. Although targeting is very efficient, we have seen great variation in expression levels that cannot readily be explained but is probably due to multiple recombination and/or rearrangement events occurring during integration of constructs into the tandemly repeated tubulin genes. The clonal variation was most apparent in an extensive series of GFP integrations into this locus (Ina Ly, George Cross and Nina Papavasiliou, unpublished data).

Strategies for Conditional Gene Knockouts (assuming a single locus)

Strategy 1 for BF or PF Introduce an ectopic copy of your gene into SM or 29-13 PF cells on pLEW82, pUB39, or pLEW100. Knock out the wild-type alleles using 'floxed' positive-negative selection cassettes, with appropriate positive selectable markers (five are available) excising the markers subsequently with Cre recombinase.

Strategy 2 for BF or PF This strategy, introduced and used in several collaborations by Liz Wirtz, is probably obsolete today. You can introduce T7RNAP and TetR cassettes into your favorite trypanosome strain, using targeting vectors pLEW13 and pLEW90, after replacing the tubulin regions with the targeting sequences appropriate to the loci into which you want to place them, as part of a knock-out strategy. It may not be easy. There are several uncertainties, some of which are listed below. In the future, you will be able to create ZM (zero marker) stable T7RNAP and TetR cell lines, as described above for 427.

(i) The pLEW13 cassette depends, for T7RNAP expression, on read-through transcription in the targeted locus. This has not caused problems for us, so far.

(ii) The pLEW29 cassette and pLEW90 cassettes are driven by a 10% strength T7 promoter. Insertion of this promoter could interfere with transcription of downstream trypanosome genes. Although we have not experienced this problem, we still know little about how Pol II transcription works in trypanosomes.

(iii) If you use this approach, you will have to test and pick cell lines that regulate efficiently, as we did, using a standard reporter like pLEW100.

The Marker Problem

For a long time, our ability to conveniently manipulate genes was widely regarded as being limited by the number of selectable (drug-resistance) genes that work in *T. brucei*. Although additional drug-gene combinations could probably be devised, maintaining cells in the presence of multiple drugs is

undesirable. If you've found other drugs that work, let me know and I can add this information. It has been reported in Leishmania [17] that one allele can be knocked out then increasing drug concentration can force a gene conversion (loss of heterozygosity) to knockout the second allele. This approach did not work on the two occasions where we tried quite hard to do it in *T. brucei*. Christine Clayton may have additional (unpublished) information on *T. brucei*. Whether or not it works efficiently enough to be useful is doubtful, but could be locus-dependent. A better option is to use the cre-lox system (see below), which allows drug-resistance genes to be re-used.

Proven drug-resistance markers

MARKER	PROMOTER REQUIREMENT		SELECTION	CONCENTRATION (µg/ml)	
	BF	PF		BF	PF
NEO	Pol II	Pol II	G418	1 to 3 *	15
HYG	Pol II	Pol II	Hygromycin	4 to 5	25 to 50
BLE	?	?	Phleomycin	1 to 2.5	2.5
SAT	T7	Pol II	Nourseothricin	25	100
PUR	Pol II	Pol II	Puromycin	0.1	1.0
BSD	Pol II	Pol II	Blasticidin	5	10
HSV-TK	Pol II	not tested	Ganciclovir	5 to 30 or more **	not tested
TbPYR6-5 ('ura3')	Pol II	not tested	FOA	6	not tested

* 3 µg/ml was the minimum G418 concentration that killed 10^5 cells in 1ml in 2 days: 1 µg/ml was the minimum concentration that killed 100 cells in 1 ml in 2 days: transfectants with NEO inserted at silent ESs could be obtained using 2 µg/ml.

** see the cre-lox page on our web site.

? uncertain about required expression level. We use it mainly with a strong promoter, in the context of integrating our pLEW82 and pLEW100-style regulated expression cassettes, but have also integrated it downstream of a 'silent' ES promoter, using 1 µg/ml of phleomycin (Navarro, Wirtz & Cross, 1999)

The cre-lox system

In its most common use, the bacteriophage P1 CRE recombinase catalyzes the excision of a 'floxed' (flanked by *lox*) DNA sequence, one flanked by 34-bp *loxP* elements in direct orientation, leaving one residual *loxP* sequence at the site of excision. This system has found many uses, especially in situations where a cell-specific or stage-specific promoter can be used to induce CRE and delete a floxed gene in a specific tissue and/or at a desired stage in an organism's developmental cycle. It can also be used to induce chromosomal rearrangements, to target DNA to a specific site in the genome, for DNA inversion, or to create large deletions that might be hard to achieve by conventional knockout strategies. After the original proof that the cre-lox system works in *T. brucei* [6], we decided to focus on developing systems that would allow drug-resistance markers to be re-used for multiple gene deletions or other genetic manipulations. To eliminate the toxicity of CRE, we developed pLEW100-based CRE-expressing plasmids with negligible background activity, but we ultimately found that transient transfection of a CRE-expression vectors was highly efficient. More details, including updated constructs, can be found on the cre-lox page of this website.

General Comments

The baseline reference for this system is "CRE recombinase-based positive-negative selection systems for genetic manipulation in Trypanosoma brucei" [31] DOI PMC. I suggest that you read at least parts of this

paper, especially the discussion! Please note that we have only used this system in BF. The floxed constructs listed on the web site are the enhanced versions referred to in the discussion of the above paper. The most important features are:

(1) the floxed constructs are available with 5 positive-selectable markers.

(2) the version of the HSVTK that is used contains the mutations that rendered mammalian cells 40-fold more sensitive to GCV [7]. Although we found that this mutation had no effect on GCV sensitivity of *T. brucei* when expressed from the beta-TUBULIN locus in HMI-9 lacking thymidine, it could make a difference at other loci.

(3) these versions contain the most efficient splice site that we know of, downstream of the upstream *lox* site. This was done with the intention of ensuring high expression of the fusion protein at any targeted locus regardless of the SAS at that locus. Whether this in fact makes a difference has not been tested by us.

(4) as suspected might be the case, we have confirmed that deleting both alleles of the pyrimidine gene that is targeted by the prototypical constructs prevents the cells from establishing an infection in mice but does not affect growth in culture, where the media generally ensure an abundant supply of pyrimidines. A similar deficiency in pyrimidine synthesis was previously shown to prevent infection of mice by *Toxoplasma gondii* [15]. These results confirm the rationale for the presence of pyrimidine synthesis pathways in protozoa. There is an insufficient supply of pyrimidines in the host, despite the presence of efficient pyrimidine transporters in *T. brucei* [9].

Using the floxed vectors

The floxed targeting vectors can be tested at the pyrimidine locus the '*Ura3*' equivalent, after linearization with PvuII. The pyrimidine-targeting sites can be replaced with your targets of interest, either by PCR amplification of the floxed cassette with long primers terminating in the construct sequences or by cloning in longer targeting fragments to replace the PYR upstream and downstream targets.

As already noted, the efficiency of 'PCR targeting' can vary among target loci. With the floxed cassettes, we (Cross lab: Hee-Sook Kim) have found that PCR-targeting the second allele can be especially inefficient, probably because of the TK homology within the floxed cassettes containing different positive selectable markers. This 'second allele problem' can be obviated in two ways: either treat with cre to eliminate the first floxed cassette prior to targeting the second allele, which can now be done with the same cassette, or (if the PCR oligonucleotides are still not working), the first floxed allele can be PCR-amplified using more distant flanking primers and this product, with its much longer flanking regions, can be used to target the second allele after cre treatment to eliminate the first floxed allele (alternatively the PCR product could be cloned and the positive selectable markers switched via two of the unique flanking restriction sites, allowing both alleles to be targeted before cre-mediated excision).

In case of difficulty targeting even the first allele, the pyrimidine targets in the provided vectors can be replaced with other PCR-amplified target loci. First, clone the downstream target sequence, with BamHI and SbfI ends into a pyrFEKO vector digested with the same enzymes, then clone the upstream target sequence, with PvuII and HindIII ends into a pyrFEKO vector digested with the same enzymes. Linearize the vector with PvuII plus SbfI. For either cloned or PCR oligo targeting, proceed as follows.

Step 1 Transfect and select for drug resistance (hygromycin, puromycin, etc). You can probably avoid cloning prior to Steps 2 & 3.

Step 2 Repeat Step 1 to target the second allele.

Step 3 Transiently transfect with a pLEW100cre plasmid (see below) and selection with Gancyclovir (GCV). The amount of GCV you need for negative selection depends on the expression level. If the TK construct is in the VSG Expression Site, 5 micrograms per ml is sufficient but you may have to use at least 30 micrograms per ml of GCV at other loci. BF should be grown in HMI-9 without thymidine.

The original pLEW100-cre plasmid was provided by John Donelson [6] with his permission to distribute it along with the modified derivatives made by us.

All the pLEW100cre plasmids can be stably transfected, although in this case we suggest using the one with least (essentially no) uninduced activity, which is pLEW100cre-EP1-6G. Any of the pLEW100 cre versions can be used for our recommended transient transfection expression of Cre. The version from which the Tet operator has been deleted is intended for transient use in Tet repressor-expressing cells, to avoid having to us tetracycline induction, in case this would induce other tet-regulated insertions in the same cell line.

RNA interference (RNAi)

Strategies for RNAi-mediated gene knockdown

The simplest (but not the best) way to knock down gene expression is to clone the coding sequence into one of the purpose-made vectors between dual opposing inducible T7 promoters, as first demonstrated by the Donelson and Englund laboratories [20, 36]. The most widely vectors being the 2T7TAblue and pZJM, from the laboratories of David Horn and Paul Englund, respectively [1, 36]. This approach suffers from two serious disadvantages. First, expression from these vectors, with their strong T7 promoters, is very leaky, which sometimes can make it impossible to obtain clones when the target genes are essential. Second, using the entire coding sequence can result in off-target effects. The first disadvantage can be avoided by using a hairpin (stem-loop) construct in the tightly regulated pLEW100 vectors, at which time the second problem can be eliminated by a more discriminating choice of sequence for RNAi — or the use of multiple alternative sequences in the RNAi vector, and other controls, as discussed later. The pLEW100 vectors use strong intrinsic pol-I promoters, which are 2–3-fold less active than T7 polymerase, but the slightly lower transcription level is probably outweighed by better control of conditional expression.

Conditional RNAi is also subject to the intrinsic deficiencies of the current widely used TetR and T7RNAP cell lines, as mentioned above, although putting the T7 polymerase under control of the Tet repressor, creating a 'doubly-inducible' cell line, can decrease leaky RNAi expression from dual T7 RNAi vectors [1].

It has been widely assumed that making stem-loop constructs requires more (and more problematical) cloning steps than using the dual opposing T7 promoter systems. This is not true. A target sequence can be amplified using primers containing different restriction sites on each end. After amplification, the product is digested with one restriction enzyme, then the mixture is self-ligated, cut with the second enzyme, and cloned into a pLEW100-series vector digested with the same enzyme. Good-practice approaches to design and quality control for RNAi is exemplified in a 2007 paper describing the generation of a genome-wide RNAi library for *Drosophila* [11].

Evaluating phenotypes generated by RNAi knockdown

There are several major considerations in evaluating RNAi results. The first is that a gene might be essential but this might not be apparent from an RNAi experiment, either because the target mRNA is not knocked down enough, or because only clones that somehow resist (or adapt to resisting) the RNAi effect are selected for. The second major disadvantage of RNAi is ensuring that the observed phenotype is directly attributable to the target gene.

Using inducible RNAi libraries for forward genetics

As pioneered in the Englund lab [12, 23, 24], RNAi libraries of random-sheared size-fractionated genomic DNA have considerable potential for forward genetics, although screening and cloning is very labor-intensive, and the approach suffers from the inherent disadvantages of the dual T7 promoter vector and transfection efficiencies currently limit feasibility to procyclic forms. A similar vector was used for a pilot large-scale gene-by-gene RNAi knockout study [34]. The best approach to using an RNAi library would be to follow the lead from *Drosophila* [11], making a gene-by-gene genome-wide custom library, but this approach is beyond the reach of individual laboratories, unless funds can be acquired to generate a community-wide resource, as has been recommended by several investigators. Due to the administrative difficulties in preserving and distributing a library of transgenic trypanosomes, a plasmid library would be made available to any investigator requesting it.

Episomes

Episomal vectors for *T. brucei* are not well developed. This is unfortunate, as they have some proven and potential advantages as a genetic tool. Ideally, an episomal vector would replicate and segregate as a stable single-copy entity, in both bloodstream and procyclic forms, and have high transformation

efficiency. The vector that comes closest to these objectives is pT13-11, which derived these desirable properties by an empirical cloning strategy [29]. This plasmid stably transforms PF at an efficiency close to 1 in 1,000 cells. It is stably maintained so long as G418 selection is applied, but lost within a few cell divisions in the absence of selection. This makes it potentially useful for testing if a gene is essential: if an essential gene is expressed from the episome, the two endogenous alleles would be dispensable and the plasmid should then self-select for persistence. This approach has been reported in BF [26] but, because the slight variant of pT13-11 that was used (pT11-bs) is not efficiently replicated and/or propagated, it is possible that the episome became integrated into the genome in this study. Although there has been some further work to identify the sequences necessary for the maintenance of pT13-11 in PF [14], our understanding of this plasmid remains incomplete, and it and its derivatives give equivocal results in BF. With our current knowledge of the trypanosome genome, it should be possible to take a more hypothesis-driven approach to designing a better episomal vector.

Two other laboratories have tackled this topic, but none have derived a plasmid as useful as pT13-11. Also using a random genomic DNA cloning strategy, Agabian's lab showed that a kDNA minicircle-derived sequence could provide plasmid replication and maintenance functions. Although potentially useful for complementation, as shown in a later proof-of-principle experiment [33], this plasmid has the major disadvantage of replicating as a concatamer. Mary Gwo-Shu Lee's group have made an extensive study of another series of plasmids that could have some utility, although they owe their stability to rapid integration into minichromosomes (Lee, personal communication). Both Lee [22] and Patnaik [28] have reported the construction of Trypanosome Artifical Chromosomes (TACs): the Patnaik TACs are essentially linearized variants of pT13-11 to which telomeres were added. TACs or tagged endogenous minichromosomes could be useful for studying certain aspects of chromosome replication and segregation in *T. brucei*, but have no obvious toolbox utility.

Patnaik's pT13-11 is the most useful episome so far, but far from ideal. Nevertheless, it forms the basis of the first successful approach to mariner transposon mutagenesis in *T. brucei* [21] (see below). Improvements could make it valuable for genome-wide RNAi screening, especially in BF.

Mutation and complementation

Chemical mutation and complementation is feasible in procyclic forms (the transfection efficiency of bloodstream forms is too low with current technology), but genome-wide RNAi has many advantages. Proof-of-principal that complementation can work in PF was provided in 1996 [33], but has not been used since.

Transposon mutagenesis

A mutagenesis system based on the mariner transposon offers an alternative approach to forward genetics [21], although subsequent experiments suggest that the high rate of allelic gene conversion reported at the originally investigated locus, which resulted in the generation of diploid mutants, is atypical. Therein lies the main drawback of using transposon mutagenesis in a diploid organism. However, some single-allele random transposon insertions would be expected to show phenotypes; by creating dominant-negative effects, for example. The main advantage of the system we developed is that its application is independent of transfection efficiency, and allows very large transposon-insertion populations to be generated and screened.

Gene (protein) tagging

Two principal approaches have been used to tag trypanosome proteins for visualization, purification and co-immunoprecipitation, using a variety of short epitopes, for which well-characterized antibodies are available, or with larger tags such as various flavors of fluorescent proteins (GFP, YFP, etc), or to allow tandem affinity purification (TAP tagging). The more 'traditional' and still useful approach has been to express a tagged ectopic gene from a regulated expression vector such as pLEW100 or pLEW82. As the cost of oligonucleotide synthesis has declined and the transfection efficiency of BF increased, PCR tagging of endogenous genes has become a simpler option in most situations. A crucial point is that it must always be shown that the tagged version can substitute for the native gene. This normally means that endogenous untagged copies must be deleted after the tagged copy is created. A principal rationale for tagging an endogenous allele is to maintain expression at a similar level as for the natural gene. If deletion of a gene shows no phenotype, however, it is impossible to know whether a tagged version is functional.

In many cases, tagging constructs (whether tagging endogenous or ectopic genes) were created ad hoc, using PCR primers that contained the tag sequence in frame with the target gene. In the case of ectopic copies, the initial constructs were often made so that they could be re-used to permit subsequent genes-of-interest to be tagged by insertion at convenient restriction sites. In this way, a series of constructs that allow fusion of any gene of interest and its ectopic expression can be constructed. An extensive series of useful constructs from the Carrington lab (http://web.mac.com/mc115/iweb/mclab) [19] exemplifies this approach. The main concern with this approach is that these expression vectors are generally driven by strong T7 or Pol-I promoters and contain highly efficient trans-splicing signal sequences and 5´ and 3´ UTRs from highly expressed genes, and might, therefore, express tagged genes at unnaturally high levels or at inappropriate stages in the cell cycle or life cycle. This is the main reason for adopting in-situ tagging.

The Carrington laboratory has also produced an extensive range of N- and C-terminal in-situ TAP-tagging and epitope-tagging (up to 12 copies of the HA tag) vectors (http://web.mac.com/mc115/iweb/mclab) [19] and the Seebeck laboratory has produced a smaller but very useful unified set of constructs for in-situ epitope-tagging at a protein C-terminus http://www.izb.unibe.ch/res/seebeck/taggingvectors.php [27]. Tagging vectors from both sources are available with alternative drug-resistance genes. The only potential disadvantage of these vectors is that they replace the endogenous 3´ or 5´ UTR. This could affect the expression level of the tagged gene (sequences that regulate RNA stability have been identified in the 3´ UTRs of several genes). For this reason, we have made floxed versions of some of the Seebeck tagging vetcors, which will only add a 34-bp loxP footprint to the endogenous 3´ UTR.

In most cases, tagging the C-terminus of the target protein is fine. Sometimes, however, only the N-terminal end can be tagged without compromising function.

Frequently asked questions and answers

Q I am only interested in expressing an alien gene in trypanosomes, or attempting to over-express a trypanosome gene. I don't want to knock out any endogenous gene. What is the simplest approach?

A You can use our ready-made cell lines and simply insert a copy of your gene into one of the targeting plasmids pLEW82v4 or one of the more extensive and preferable pLEW100 series, depending upon the degree of regulation and expression desired.

References

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

2. Alsford S, Horn D. RNA polymerase I transcription stimulates homologous recombination in *Trypanosoma brucei*. Mol Biochem Parasitol 2007;153:77-9.

3. Alsford S, Horn D. Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*. Mol Biochem Parasitol 2008;161:76-9.

4. Alsford S, Kawahara T, Glover L, Horn D. Tagging a *T. brucei RRNA* locus improves stable transfection efficiency and circumvents inducible expression position effects. Mol Biochem Parasitol 2005;144:142-8.

5. Barnes R, Mcculloch R. *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. Nucleic Acids Res 2007;35:3478-93.

6. Barrett B, Lacount DJ, Donelson JE. *Trypanosoma brucei*: a first-generation CRE-loxP site-specific recombination system. Exp Parasitol 2004;106:37-44.

7. Black M, Newcomb T, Wilson H, Loeb L. Creation of drug-specific *Herpes simplex* virus type 1 thymidine kinase mutants for gene therapy. Proc Natl Acad Sci U S A 1996;93:3525-9.

8. Boehme U, Cross GAM. Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. J Cell Sci 2002;115:805-16.

9. De Koning HP, Bridges DJ, Burchmore R. Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy. FEMS Microbiol Rev 2005;29:987-1020.

10. Dhir V, Allen CL, Field MC. Perturbation of local endogenous expression by insertion of Pol I expression constructs into the genome of *Trypanosoma brucei*. Exp Parasitol 2005;109:198-200.

11. Dietzl G, Chen D, Schnorrer F, Su K, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson B. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 2007;448:151-6.

12. Drew ME, Morris JC, Wang Z, Wells L, Sanchez M, Landfear SM, Englund PT. The adenosine analog tubercidin inhibits glycolysis in *Trypanosoma brucei* as revealed by an RNA interference library. J Biol Chem 2003;278:46596-600.

13. Engstler M, Boshart M. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in *Trypanosoma brucei*. Genes Dev 2004;18:2798-811.

14. Espinal A, Quijano J, Hunt C, Lorenzo R, Mulligan C, Sampson M, Sauchelli M, Patnaik P. A 10 base-pair sequence within Domain III of the GPEET procyclin promoter is essential for the autonomous replication of a plasmid in procyclic *Trypanosoma brucei*. Mol Biochem Parasitol 2007;151:224-8.

15. Fox BA, Bzik D. De novo pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*. Nature 2002;415:926-9.

16. Furger A, Schurch N, Kurath U, Roditi I. Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation. Mol Cell Biol 1997;17:4372-80.

17. Gueiros-Filho FJ, Beverley SM. Selection against the dihydrofolate reductase-thymidylate synthase (DHFR-TS) locus as a probe of genetic alterations in *Leishmania major*. Mol Cell Biol 1996;16:5655-63.

18. Hide G. Visualizing trypanosome sex. Trends Parasitol 2008;24:425-8.

19. Kelly S, Reed J, Kramer S, Ellis L, Webb H, Sunter J, Salje J, Marinsek N, Gull K, Wickstead B, Carrington M. Functional genomics in *Trypanosoma brucei:* a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. Mol Biochem Parasitol 2007;154:103-9.

20. Lacount DJ, Bruse S, Hill KL, Donelson JE. Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. Mol Biochem Parasitol 2000;111:67-76.

21. Leal S, Acosta-Serrano A, Morris JC, Cross GAM. Transposon mutagenesis of *Trypanosoma brucei* identifies glycosylation mutants resistant to Concanavalin A. J Biol Chem 2004;279:28979-88.

22. Lee MGS, E Y, Axelrod N. Construction of trypanosome artificial minichromosomes. Nucleic Acids Res 1995;23:4893-9.

23. Morris JC, Wang Z, Drew ME, Englund PT. Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. EMBO J 2002;21:4429-38.

24. Motyka SA, Englund PT. RNA interference for analysis of gene function in trypanosomatids. Curr Op Microbiol 2004;7:362-8.

25. Motyka SA, Zhao ZX, Gull K, Englund PT. Integration of pZJM library plasmids into unexpected locations in the *Trypanosoma brucei* genome. Mol Biochem Parasitol 2004;134:163-7.

26. Nagamune K, Nozaki T, Maeda Y, Ohishi K, Fukuma T, Hara T, Schwarz RT, Sutterlin C, Brun R, Riezman H, Kinoshita T. Critical roles of glycosylphosphatidylinositol for *Trypanosoma brucei*. Proc Natl Acad Sci USA 2000;97:10336-41.

27. Oberholzer M, Morand S, Kunz S, Seebeck T. A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. Mol Biochem Parasitol 2006;145:117-20.

28. Patnaik PK, Axelrod N, Van Der Ploeg LHT, Cross GAM. Artificial linear mini-chromosomes for *Trypanosoma brucei*. Nucleic Acids Res 1996;24:668-75.

29. Patnaik PK, Kulkarni S, Cross GAM. Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. EMBO J 1993;12:2529-38.

30. Peacock L, Ferris V, Bailey M, Gibson W. Fly transmission and mating of *Trypanosoma brucei brucei* strain 427. Mol Biochem Parasitol 2008;160:100-6.

31. Scahill MD, Pastar I, Cross GAM. CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. Mol Biochem Parasitol 2008;157:73-82.

32. Siegel TN, Tan KSW, Cross GAM. A systematic study of sequence motifs for RNA trans-splicing in *Trypanosoma brucei*. Mol Cell Biol 2005;25:9586-94.

33. Sommer JM, Hua SB, Li FS, Gottesdiener KM, Wang CC. Cloning by functional complementation in *Trypanosoma brucei*. Mol Biochem Parasitol 1996;76:83-9.

34. Subramaniam C, Veazey P, Redmond S, Hayes-Sinclair J, Chambers E, Carrington M, Gull K, Matthews K, Horn D, Field M. Chromosome-wide analysis of gene function by RNA interference in the african trypanosome. Eukaryot Cell 2006;5:1539-49.

35. Van Deursen FJ, Shahi SK, Turner CMR, Hartmann C, Guerra-Giraldez C, Matthews KR, Clayton CE. Characterisation of the growth and differentiation in vivo and in vitro of bloodstream-form *Trypanosoma brucei* strain TREU 927. Mol Biochem Parasitol 2001;112:163-71.

36. Wang Z, Morris JC, Drew ME, Englund PT. Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. J Biol Chem 2000;275:40174-9.

37. Wickstead B, Ersfeld K, Gull K. Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. Mol Biochem Parasitol 2002;125:211-6.

38. Wickstead B, Ersfeld K, Gull K. The frequency of gene targeting in *Trypanosoma brucei* is independent of target site copy number. Nucl Acids Res 2003;31:3993-4000.

39. Wirtz E, Hoek M, Cross GAM. Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. Nucl Acids Res 1998;26:4626-34.