DEVELOPING CRISPR/CAS9 FOR GENOME-WIDE GENE EDITING IN THE HUMAN PATHOGEN *TRYPANOSOMA CRUZI*

by

DUO PENG

(Under the Direction of Rick Tarleton)

ABSTRACT

The protozoan parasite *Trypanosoma cruzi*, the causative agent of Chagas disease, can infect nearly all mammals and infections are usually lifelong if not treated. The mammalian host's CD8+ T cell response is crucial to controlling this intracellular parasite. However, the development of the host CD8+ T cell response is significantly delayed compared to that against other pathogens. A significant proportion of host CD8+ T cells in *T. cruzi* infection is specific for peptides derived from the *trans*-sialidase (TcTS) gene family. The reference genome of *T. cruzi* encodes 3209 TcTS genes and it is hypothesized that this large TcTS gene family repertoire aids in immune evasion. One way to test this hypothesis is to create a TcTS-knockdown *T. cruzi* line. Here, I aimed to build that foundation by knocking-down of the expression of TcTS genes and as well as knocking-out the coding capacity of silent TcTS, so as to impede the restoration of TcTS expression by recombination. To accomplish this, I adapted the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas (CRISPR-associated) system for use in *T. cruzi*. CRISPR/Cas9 system is a genome editing tool adapted from bacteria. Using this system, I demonstrated rapid and efficient gene knockout in *T. cruzi*, sequence editing using DNA donor template and the ability to knockdown a gene family of 65 members. I next

developed a computation tool: the eukaryotic pathogen gRNA design tool (EuPaGDT, http://grna.ctegd.uga.edu) capable of designing CRISPR/Cas9 gRNA sequences targeting genes at a genome-wide scale. EuPaGDT's unique algorithm allows on-target search, which enumerates all the gene family members that a gRNA targets. Using this tool, I designed 21 gRNAs, that when combined, target more than 87% of TcTS capable of producing peptides and 73% of TcTS at least 30% of full gene length. I sequentially delivered the 21 gRNAs in complex with Cas9 protein and associated repair templates to *T. cruzi*. The resulting *T. cruzi* parasites population showed a 55% decrease in TcTS-binding fetuin staining. Sequencing of the gene editing sites in TcTS revealed successful editing of TcTS at multiple sites.

INDEX WORDS: *trans*-sialidase, CRISPR/Cas9, gRNA, *T. cruzi*

DEVELOPING CRISPR/CAS9 FOR GENOME-WIDE GENE EDITING IN THE HUMAN PATHOGEN *TRYPANOSOMA CRUZI*

by

DUO PENG

B.S., Wuhan University, China, 2010

M.S., University of Georgia, USA, 2016

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

DEVELOPING CRISPR/CAS9 FOR GENOME-WIDE GENE EDITING IN THE HUMAN PATHOGEN *TRYPANOSOMA CRUZI*

by

DUO PENG

Major Professor:	Rick Tarleton
Committee:	Jessica C. Kissinger
	Robert Sabatini
	Ronald D. Etheridge

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

DEDICATION

This thesis is dedicated to my parents, Jingpian Peng and Jiamei Liu, as well as Xiaochen Zhu, who became my wife during this time.

ACKNOWLEDGEMENTS

I want to first of all thank Dr. Rick Tarleton, without whose guidance and supervision this thesis work is simply not possible. I still remember the first time I walked into his office telling him I wanted to simultaneously do a Ph.D. in cellular biology and a M.S. in bioinformatics. He supported my passion and directed me in pursuit of an interdisciplinary training. It allowed me to embark on challenging projects and the intellectual reward is tremendous. I extend my thanks to my committee members: Dr. Jessica Kissinger, Dr. Robert Sabatini, and Dr Ronald Etheridge for their advice, assistance, and academic support, for their constructive questions during my thesis committee meetings.

To the wonderful people currently and once in the Tarleton Research Group: without your support and accompaniment, I couldn't complete my Ph.D. journey, let alone enjoy it. Dr Samarchith Kurup, thank you for sharing your research experience and all the great research ideas, I miss our late-night discussions. To my Ph.D. student peers (a.k.a. the TRG lunch crew): Phil Yao, Weibo Zhang, and Angela Pack, in addition to well-accompanied lunches, thank you for your friendship and support in and out of the lab. Phil, I miss our late-night food trips after lab. I want to thank Todd, Angel, Gretchen, Juan, Bharath, Wei, Arlene, Brooke, Britton, Fernando, Huifeng, Molly, Natasha, Dilan, Lia, Donna for being great colleagues. To Georgia Wilson, Alexandria Purcell, Lilith South, Evelina Kravchuk, thank you for being great undergraduate helpers.

I want to thank Julie Nelson of the Flow cytometry core facility, Dr. Muthugapatti Kandasamy of the Biomedical Microscopy Core, who have been very helpful to and supportive of my research.

There are too many of my friends to name individually, however special thanks are given to Yang Yang, Xuebin Wei, Haiwei Luo, Liangjiao Xue, Ciro, Lexiang Ji, Chenming Sun, Lina Sun and Zheng Ruan.

Lastly, I would like to thank my parents for their endless love, support and faith in me that kept me going on this journey. To my wife, Xiaochen Zhu, thank you for your unconditional sacrifices and supporting my pursuit of science.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 CHAGAS DISEASE AND *TRYPANOSOMA CRUZI*

Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, affects 6-7 million people worldwide (http://www.who.int/mediacentre/factsheets/fs340/en/, March 2017). Though Chagas disease affects mostly underdeveloped, rural regions of Latin America, it has spread to other continents due to global traveling. About 59,000-108,000 cases of Chagas disease are estimated in Europe ((WHO) 2010), and over 300,000 people in the United states and 4,500 in Japan are infected (Gascon, Bern et al. 2010, Bern, Kjos et al. 2011, Imai, Maeda et al. 2015). Chagas disease has both acute and chronic phases. If not treated promptly in the acute infection phase, most infections enter a chronic phase in which progressive and debilitating conditions develop over decades. *T. cruzi* persistence in patients is thought to impair smooth muscle, peripheral nerve and cardiac muscle and associated nerves, causing digestive diseases, cardiac arrhythmia or progressive heart failure, which is responsible for sudden death in later years of infection.

The transmission of Chagas disease in nature is primarily through contact with feces or urine of hematophagous insect vector triatomine bugs deposited after a blood meal. *T. cruzi* can also be transmitted congenitally or by blood transfusions. Transmission of *T. cruzi* to humans is promoted by poor housing conditions that provide both a habitat for the triatomine bugs and access to sleeping, unsuspecting humans. The presence of an animal reservoir promotes

transmission of *T. cruzi* to humans, as many infected animals such as dogs, cats, and other wild animals live in close association with humans and human dwellings (Cohen and Gurtler 2001).

Diagnosis of Chagas disease is complicated by extremely low parasite levels in the blood during the chronic infection phase and a lack of sensitive, dependable and non-invasive procedures for parasite detection. Widely used diagnostic strategies are based on serological identification of anti-*T. cruzi* antibodies produced by the host immune system. This diagnosis method suffers from reliability issues rooting from poor standardization of serological test reagents and cross-reactivity with antibodies produced by other pathogens (Camargo, Segura et al. 1986, da Silveira, Umezawa et al. 2001).

There are 2 drugs used for treatment of Chagas disease, benznidazole and nifurtimox. Both drugs can cure Chagas disease at close to 100% rate if administered at the onset of the acute phase infection, but both are significantly less effective when treatment starts in the chronic infection phase. Severe side effects may result in discontinuation of treatment in up to 20% of cases (Urbina 2015). Currently, there is no vaccine for Chagas disease.

1.2 *T. CRUZI* IN THE MAMMALIAN HOST

T. cruzi has four life cycle stages, 2 in the mammalian host and 2 in the insect vector. In the insect vector, *T. cruzi* multiplies as epimastigotes in the midgut, before migrating to the hindgut and converting to infective metacyclic-trypomastigotes, which will be deposited in the feces and urine of the triatomine bug. Metacyclic-trypomastigotes gain access to mammalian host through various routes such as insect bit wounds, contact with insect feces or urine through the mucosal membrane, or ingesting the insect vector, as can be the case with animals. Once inside the mammalian host, *T. cruzi* metacyclic-trypomastigotes can infect virtually any

nucleated cell in the host. Shortly after *T. cruzi* enters a host cell, it escapes from the parasitophorous vacuole, a membrane bound acidic compartment that encapsulates *T. cruzi* in the invasion process. In the cytoplasm of host cells, *T. cruzi* (metacyclic-)trypomastigotes discard their flagella, partly via an asymmetric cellular division (Kurup and Tarleton 2014), and convert into oval-shaped amastigotes, a proliferative form which begins to divide in the host cell cytoplasm. About 4-5 days later, descendent amastigotes completely occupy the cytoplasm. Amastigotes will then convert into trypomastigotes, a non-proliferative but infectious stage, which are then released from the destroyed host cell and continue to either infect adjacent host cells or enter the circulation system, this enables that parasites to migrate to other parts of the host, to continue the infection cycle or be taken up by an insect vector via a blood meal, to continue the transmission cycle.

1.3 THE *T. CRUZI TRANS*-SIALIDASE (TCTS) GENE FAMILY

Trans-sialidase is an enzyme glycosylphosphatidylinositol(GPI)-anchored on the surface of *T. cruzi* that catalyzes a transglycosylation reaction, transferring sialic acids from sialylated donors (mostly host cell surface molecules) to the terminal galactose on the parasite surface, as well as the reverse transfer reaction (Vandekerckhove, Schenkman et al. 1992, Ferrerogarcia, Trombetta et al. 1993). *T. cruzi trans*-sialidase (TcTS) is more efficient in transferring sialic acid than hydrolyzing it (Schenkman, Decarvalho et al. 1992).

T. cruzi is unable to synthesize sialic acids *de novo*, thus the transferase activity of *trans*-sialidase is the only mechanism for parasite surface sialylation. Parasite surface sialylation protects *T. cruzi* from lysis by the alternative pathway of the complement system in the host (Kipnis, Krettli et al. 1985, Tomlinson, Decarvalho et al. 1992). Though full sialylation of

parasite surface is not a primary requirement for invasion (Schenkman, Vandekerckhove et al. 1993), *trans*-sialidase activity inhibition was shown to reduce infection of cultured cells (Agusti, Paris et al. 2004, Carvalho, Sola-Penna et al. 2010), and higher levels of sialylation of mammalian cells enhances *T. cruzi* infection (Schenkman, Vandekerckhove et al. 1993). Enzymatically inactive members of the *trans*-sialidase family were shown to bind substrates on the host cell surface without catalyzing sialic acid transfer (Todeschini, Dias et al. 2004, Dias, Fajardo et al. 2008), potentially mediating parasite attachment to host cells and blocking host cell apoptosis via cell signaling (Dias, Fajardo et al. 2008).

TcTS protein consists of 3 domains: (1) the N-terminal catalytic domain, which folds in to a six-bladed beta propeller, similar to microbial sialidases. The catalytic domain is responsible for the transfer of sialic acid and hydrolysis activity (Buschiazzo, Amaya et al. 2002). A single mutation in the catalytic domain (position 342, from tyrosine to histidine), abolishes its enzymatic activity (Uemura, Schenkman et al. 1992), while retaining the ability to bind to sialic acids (Cremona, Campetella et al. 1999), (2) the second domain is the C-terminal lectin-like domain, which is capable of binding to nerve growth factor receptor (Tropomyosin receptor kinase A, TrkA) on host cells (Chuenkova and PereiraPerrin 2005). Binding to the TrkA receptor facilitates *T. cruzi* invasion of neural, glial and epithelial cells (de Melo-Jorge and PereiraPerrin 2007, Weinkauf, Salvador et al. 2011), (3) the third domain is the C-terminal extension of lectin-like domain formed by tandem arrangement of variable number of units of 12 amino acid repeats known as shed acute phase antigen (SAPA). SAPA repeats have been shown to increase the half-life of TcTS when released from the parasite surface.

TcTS is encoded by a gene family consisting of 3209 members, a recent update from 1430, based on previous work of the thesis author (Weatherly, Peng et al. 2016). A majority of

the 3209 TcTS genes do not encode functional protein, 1718 are pseudogenes characterized by at least one coding-debilitating mutations or frameshifts, 2516 are partial genes characterized by a lack of a signal peptide or GPI anchor which marks the canonical start and end of TcTS genes (Weatherly, Peng et al. 2016). 437 are full length, potential protein coding TcTS, and only 15 are predicted to code enzymatically active TcTS protein.

Extensive evidence of recombination among members of the TcTS gene family has been documented by the thesis author, 1255 (40%) out of 3209 TcTS genes have participated at least once in recombination, generating 787 TcTS with at least one mosaic piece (Weatherly, Peng et al. 2016). More than 200 full length, protein coding TcTS are mosaic, suggesting possible flow of genetic variation from silent members to expressed members of the TcTS gene family.

TcTS are highly expressed in the trypomastigote stage in which its activity is 10- to 20-fold higher than those found in epimastigotes (Chaves, Briones et al. 1993, Abuin, Freitas et al. 1999). Several lines of evidence indicate a concomitant expression of *trans*-sialidase from multiple alleles (Millar, Wleklinski-Lee et al. 1999, Atwood, Weatherly et al. 2005, Freitas, dos Santos et al. 2011); a *T. cruzi* proteome study confirms at least 246 TcTS genes are expressed concurrently, including both full and truncated genes, 30 of the 50 highest abundance peptides are from TcTS, signifying the abundance of TcTS proteins (Atwood, Weatherly et al. 2005). The expression of this large repertoire of similar (TcTS) genes and possibly varying genes at this scale is unprecedented.

1.4 HOST CONTROL OF AND CD8+ T CELL RESPONSE TO *T. CRUZI* INFECTION

T. cruzi cycles between extracellular and intracellular stages inside mammalian hosts and the host mounts both a humoral and cellular immune response to *T. cruzi* parasites. Both immune

responses are critical to controlling *T. cruzi* infection but are often insufficient to clear the infection. The humoral immune response contributes to protection against *T. cruzi* via production of lytic antibodies(Cordeiro, Martins-Filho et al. 2001) as well as via other antigen-specific antibody responses including activation of the cellular immune response (Gupta and Garg 2010). *T. cruzi*-infected mice with defects in antibody production are unable to control the infection and ultimately succumb (Kumar and Tarleton 1998).

The cellular immune response is of equal importance in controlling *T. cruzi* infection in mice (Kumar and Tarleton 1998). CD4+ T cells coordinate the immune-effector mechanism through secretion of cytokines. CD4+ T cells also regulate antibody class switching, activation of professional phagocytes to kill intracellular parasites and activation of the CD8+ T cell response. Despite multi-faceted contributions of CD4+ T cell to *T. cruzi* infection control, there is clear evidence to suggest that the CD8+ T cell response plays a critical role in the control of *T. cruzi* infection. Mice that lack CD8+ T cells succumb to *T. cruzi* infection very quickly (Tarleton 1990, Tarleton, Koller et al. 1992, Tarleton, Sun et al. 1994). In the absence of CD4+ T cells, a robust CD8+ T cell response can still develop in response to *T. cruzi* infection, and displays similar effector function as in CD4+ T cell sufficient mice (Padilla, Xu et al. 2007).

CD8+ T cells are regarded as one of the most crucial immune cell subsets that controls *T. cruzi* infection. The anti-*T. cruzi* CD8+ T cells display effective *in vivo* cytolytic activity (Bustamante, Bixby et al. 2008). The CD8+ T cell response against *T. cruzi* displays a similar kinetic compared to most viral and bacterial infections (Bixby and Tarleton 2008, Bustamante, Bixby et al. 2008). However, the onset of CD8+ T cell response is delayed in *T. cruzi* infection. The initial CD8+ T cell activation is not evident until 9-11 days after initial *T. cruzi* infection (Padilla, Bustamante et al. 2009, Padilla, Simpson et al. 2009), a significant delay compared to

viral and bacterial infection models (*e.g.* LCMV and *Listeria monocytogenes)*, and to more closely related protozoan parasite *Plasmodium yoelii*. In these infection models, pathogen CD8+ T cell response can be detected at 3-6 days post infection. Development of CD8+ T cell response to *T. cruzi* infection is also delayed, with the peak level of *T. cruzi* specific CD8+ T cells at 3 weeks post initial infection (Padilla, Simpson et al. 2009), while for other infection models, the highest level of CD8+ T cells is often obtained within 5-9 days post initial infection.

Several potential reasons were ruled out for delayed onset and development of the *T. cruzi* specific CD8+ T cell response. Coinfection with *Listeria monocytogenes* and *T. cruzi* does not prevent the host from generating a strong CD8+ T cell response to *L. monocytogenes*at 6 days post infection, suggesting that there is no immunosuppressive effect associated with *T. cruzi* infection that was exerted on the generation of the *L. monocytogenes*-specific CD8+ T cell response (Padilla, Simpson et al. 2009). It is not the failure in the trafficking of parasites from the site of infection to draining lymph nodes that prevents the timely development of the CD8+ T cell response (Giddings, Eickhoff et al. 2006). Increasing the initial infection dose of *T. cruzi* to unnatural levels can only marginally speed up the generation of CD8+ T cell response (Tzelepis, Persechini et al. 2007) (Padilla, unpublished results). The silent invader hypothesis is favored; it states that *T. cruzi* provides little or no ligands for host pattern recognition receptors (Toll-like receptors, TLR), resulting in insufficient induction of innate immune responses and subsequent adaptive responses (Tarleton 2007). Kurup et al. tested this hypothesis by generating transgenic *T. cruzi* lines expressing bacterial pathogen-associated molecular patterns (PAMPs). Mice infected with these transgenic *T. cruzi* lines developed a *T. cruzi*-specific CD8+ T response 2-3 days earlier and the response peaked at around a similar time (2 weeks) when compared to mice infected with wildtype controls. A higher level of *T. cruzi*-specific CD8+ T cells was maintained

throughout the course of infection when PAMPs were perpetually expressed on *T. cruzi* (Kurup and Tarleton 2013). This indicates that the lack of recognizable PAMPs is responsible, but only partially, for delayed onset and peak of the *T. cruzi* specific CD8+ T cell response.

One additional abnormality regarding the CD8+ T cell response generated against *T. cruzi*, at least in a C57Bl/6 mouse infection model, is highly focused to a few dominant epitopes. At the peak of infection, more than 30% of the total CD8+ T cell response is directed against a few peptides (TSKB18 and TSKB20 etc) derived from TcTS proteins (Martin, Weatherly et al. 2006, Padilla, Bustamante et al. 2009). These dominant *T. cruzi* specific CD8+ T cells have full range of functions (cytolytic and IFNy induction *etc*), are maintained in mice during the chronic phase of infection (Martin, Weatherly et al. 2006), and contribute to control of *T. cruzi* infection (Rosenberg, Martin et al. 2010). TcTS protein peptides are also CD8+ T cell targets in human, as these are the only reported peptides recognized directly by *ex vivo* PBMCs from chronically infected humans (Martin, Weatherly et al. 2006). However, susceptibility to *T. cruzi* and parasite clearance rate was largely unchanged when the dominant *T. cruzi trans*-sialidase specific CD8+ T cells were eliminated by either repetitive intravenous administration of peptides (Rosenberg, Martin et al. 2010) or by the creation of transgenic mice that centrally tolerate dominant *trans*-sialidase epitopes (Rosenberg, Zhang et al. 2016). The unexplained delay of the *T. cruzi* specific T cell response and the mysterious, unknown purpose of the vastly expanded *trans*-sialidase gene family begs for further hypothesis testing which formulates this doctoral research.

1.5 THE CRISPR/CAS9 GENOME EDITING TECHNIQUE

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas (CRISPR-associated) genes are components of adaptive immunity in bacteria and archaea. The

CRISPR/Cas system enables these organisms to respond to and eliminate invading foreign nucleic acids (Barrangou, Fremaux et al. 2007, Fineran and Charpentier 2012, Wiedenheft, Sternberg et al. 2012). The type II CRISPR system is the most studied CRISPR system. In this system, foreign DNA is cut into small fragments and incorporated into the CRISPR locus between CRISPR repeats within the bacterial host genome. The CRISPR loci are transcribed and processed into small RNAs: CRISPR RNAs (crRNAs), each harboring a fragment from invading DNA. A Cas protein, Cas9, forms a ribonucleoprotein (RNP) complex with a crRNA and a second RNA, the transactivating RNA (tracrRNA). Each RNP can recognize a specific foreign DNA coded by its crRNA. Upon recognition of foreign DNA, Cas9 destroys foreign DNA by cleaving both DNA strands at the recognition site with its nuclease domains.

Type II CRISPR systems from *S. pyogenes* and other bacteria species have been adapted for inducing site-specific double-stranded breaks (DSB) and subsequent editing of genome sequence through repairing of the DSB (Gasiunas, Barrangou et al. 2012, Jinek, Chylinski et al. 2012, Esvelt, Mali et al. 2013, Hou, Zhang et al. 2013, Ran, Cong et al. 2015). The easy-to-use and most widely used format of this system consists of the Cas9 nuclease and a single guide RNA (sgRNA). sgRNA is a fusion of crRNA and tracrRNA. The gRNA target sequence (often referred to as gRNA sequence) is coded in the first ~20 nucleotides at the 5' end of gRNA, corresponding to the foreign DNA fragment in crRNA. The target sequence guides the Cas9-gRNA complex to a specific site of a (genome) using DNA-RNA complementarity base-paring. If a protospacer adjacent motif (PAM) is found directly following the target site, a double stranded break is introduced at the target site by the Cas9 nuclease.

Cas9-induced DSB can be repaired by one of the three different pathways: nonhomologous end-joining (NHEJ), homology directed repair (HDR), and micro-homology

mediated end-joining (MMEJ). NHEJ is an error-prone repair pathway that operates in many higher eukaryotic cells. This repair process doesn't require a DNA donor template and results in small insertions or deletions when DSBs are joined back together. NHEJ is often used to knockout gene's ability to encode functional protein. The HDR repair pathway is often used to precisely edit genomic sequences, introducing point mutations or insert desired sequences such as epitope tags, through exogenously supplied DNA donor templates. The MMEJ pathway uses regions with 5-25 bp of microhomology sequence pair that flanks a DSB to rejoin DNA ends. One strand of each end of the DSB is degraded (by 5' end resection), revealing suitable microhomologies on each side of DSB that could anneal to each other. DNA synthesis and ligation processes fill the gaps and rejoin the break, resulting in a deletion of the region between the microhomology and retention of a single microhomology sequence (McVey and Lee 2008, Glover, Jun et al. 2011).

Due to its ease of use, relatively rapid and high effectiveness in editing genome sequences, CRISPR/Cas9 system has been tested and used to alter genes in species from most evolutionary clades, such as bacteria (Jiang, Bikard et al. 2013), yeast (DiCarlo, Norville et al. 2013), rice (Shan, Wang et al. 2013), zebra fish (Hwang, Fu et al. 2013), frogs (Nakayama, Fish et al. 2013), fruit flies (Yu, Ren et al. 2013), mice (Wang, Yang et al. 2013) and rabbits (Yang, Xu et al. 2014). CRISPR/Cas9 system has been exploited beyond sequence editing, due to its easily re-programmable guide RNA that allows Cas9 to bind to specific sites in the genome. Effector molecules that can interact with certain sites of the genome are fused with a deactivated version of Cas9 (dCas9) without nuclease activity. Researchers can easily repress or activate genes using dCas9 fused with transcription activators or repressors (Gilbert, Larson et al. 2013, Chavez, Scheiman et al. 2015, Konermann, Brigham et al. 2015), purify or visualize genomic

regions using dCas9 fused with certain tags (Chen, Gilbert et al. 2013, Ma, Tu et al. 2016, Liu, Zhang et al. 2017), and modify epigenetic status of genomic regions using dCas9 fused with histone modifiers (Hilton, D'Ippolito et al. 2015, Klann, Black et al. 2017).

However, CRISPR/Cas9 system has its own limitations. Cas9 occasionally tolerates mismatches between gRNA and target DNA, allowing sequence editing at unwanted sites, creating off-target modifications. Moreover, not every site in the genome can be targeted. Cas9 from *S. pyogenes* (SpCas9) has a PAM motif of "NGG" (IUPAC nucleotide code) requirement that occurs once in 16 nucleotides assuming random nucleotide composition. "NNGRRT" (IUPAC nucleotide code) PAM motif is required by Cas9 from *S. aureus* (SaCas9) (Ran, Cong et al. 2015), an ortholog about ¾ the size of SpCas9.

1.6 MAJOR QUESTIONS

The substantial delay in the onset and development of the anti-*T. cruzi* CD8+ T cell response suggests an act of immune evasion that could only be partially explained by the lack/concealment of PAMPs on *T. cruzi*. In this thesis work, we aim to test the hypothesis that the TcTS gene family repertoire plays a role in the immune evasion process.

The TcTS enzymatic activity is provided by gene products of only a dozen genes, so the purpose of the maintenance of 3000+ copies of TcTS which lack enzymatic activity is unknown. Concomitant expression of hundreds of enzymatically inactive *trans*-sialidase is not justified by the lectin-like binding functions of the TcTS proteins. The expansion of the *trans*-sialidase gene family is unprecedented, only rivaled by a close relative of *T. cruzi*: *Trypanosoma brucei*, which maintains a repertoire of up to 2000+ surface anchored variant surface glycoprotein (VSG) genes (Cross, Kim et al. 2014). *T. brucei* expresses one VSG at a given time and switches to a different

VSG occasionally (Horn 2014, Mugnier, Cross et al. 2015). This has proven to be a successful immune evasion technique (Horn 2014). One common characteristic shared between VSG and TcTS gene families is frequent intra-family recombination (Jackson, Berry et al. 2012, Horn 2014, Mugnier, Cross et al. 2015), which produces chimeric, variant counterpart genes, which could possibly contribute to immune evasion by supplying new epitope variants.

It is hypothesized that the TcTS gene family may also contribute to immune evasion. The TcTS gene repertoires are distinct among different *T. cruzi* stains and they elicit a different TcTS-specific immune profile, indicating that immune pressure may have been acting to select and shape the different *trans*-sialidase gene repertoires in the different stains. This supports the hypothesis that the *trans*-sialidase gene family aids immune evasion. The most promising hypothetical mechanism by which *trans*-sialidase gene family may aid immune evasion is through antigen flooding. *T. cruzi* expresses hundreds of *trans*-sialidase, which contain hundreds if not thousands of CD8+ T cell epitopes (Martin, Weatherly et al. 2006). The immune system has a fixed, maximum capacity for processing and presenting epitopes from pathogens, as well as certain kinetics for epitopes recognition by effector cells. As a result, there could be multiple consequences of TcTS antigen flooding, including: (1) The host antigen presentation pathway could be largely occupied by *trans*-sialidase antigens, and as a result, immune responses could be diverted from more conserved parasite proteins which might elicit immune responses that are more effective in recognizing and eliminating parasites. (2) The activation of immune effector cells may be delayed by competition of a large, sheer volume of variant *trans*-sialidase antigens, which reduces the effective concentration of the presentation of each antigen on antigen-presenting cells (APC). APCs have limited physical space for contact between APC and immune effector cells such as T-cells. Many epitopes could be below the threshold necessary to

efficiently activate CD8+ T cells. (3) Further, altered peptide ligand (APL) antagonism, which causes T-cells to be refractive to activation and performing effector functions, may also contribute to immune evasion. Scanning the *T. cruzi* genome revealed at least 400 APLs (highly similar epitopes) of a single *trans*-sialidase epitope (Martin and Tarleton 2004). Binding of the T-cell receptor to a number of the 400 APLs will lead to antagonist activity.

The observation of a slow-developing, high-frequency, persistent, but dispensable CD8+ T cell response against *trans*-sialidase present in both infected humans and mice (Martin, Weatherly et al. 2006) are consistent with the presence of a class of decoy antigens that aid immune evasion. Moreover, *trans*-sialidase genes could be undergoing recombination and generating new epitopes, in addition to the existing abundance and extensive variation of epitopes. These phenomena may prove difficult for the immune system to adapt and effectively recognize *T. cruzi* parasites.

To unravel the role that TcTS genes play in immune evasion, we aim to knock down the expression of a large number of TcTS genes. This could be easily accomplished using RNAi, targeting highly homologous regions in *trans*-sialidase genes. Because *T. cruzi* lacks a functional RNAi pathway, we propose to utilize the CRISPR/Cas9 system to conduct thousands of genome edits in the *T. cruzi* genome to achieve knockdown of TcTS expression. Like other genome engineering tools such as: Meganucleases, zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases, the CRISPR/Cas9 system can edit multiple sites of interest given that it matches the target sequence, making it possible to knockdown the TcTS gene family in *T. cruzi*. We chose the CRISPR/Cas9 system over other genome engineering tools to perform the TcTS gene family knockdown because Meganucleases, zinc finger nucleases (ZFNs) and transcription activator–like effector nucleases (TALENS) have greater limitations.

Meganucleases are engineered versions of naturally occurring restriction enzymes. The engineering process involved to produce Meganucleases is especially challenging because the DNA recognition and cleavage functions of these enzymes are intertwined in the same domain (Smith, Grizot et al. 2006, Silva, Poirot et al. 2011). By contrast, ZFNs and TALENs have separate DNA recognition domains and a FokI cleavage domain. However, modifying zinc finger domains to recognize user-defined sequences has remained challenging, mainly due to the context-dependent effects between individual finger domains in the zinc finger domain array. In contrast to the zinc finger domains, TALE repeat domains have fewer context-dependent effects and can be assembled in a modular fashion to recognize any user-defined DNA sequence (Reyon, Tsai et al. 2012), with each TALE repeat domain recognizing one nucleotide in a strict one-to-one code fashion. However, the highly-repetitive nature of TALE repeats makes the molecular assembly process challenging. Highly repetitive DNA is unstable, thus, maintaining, subcloning and propagating DNA sequences coding the TALE repeat domains require non-standard molecular techniques, which oftentimes result in suboptimum outcomes. The author of this thesis had attempted and failed to edit an integrated GFP gene in the *T. cruzi* genome using a pre-assembled TALEN. The unstable nature of highly-repetitive DNA, despite non-standard molecular techniques applied and associated suboptimum performance, have made it impractical to assemble a few dozen of the TALEN required and express them all in *T. cruzi* to conduct *trans*-sialidase gene family knock down.

1.7 SUMMARY

Although the CRISPR/Cas9 system can perform genome editing in many species across the kingdoms of life, it has not been reported to edit *T. cruzi* genomes prior to this thesis work. In

Chapter 2, we show for the first time that the CRISPR/Cas9 system can efficiently edit the *T. cruzi* genome. We further demonstrate that the CRISPR/Cas9 system can edit multiple gene members of a mid-sized (65 copies) gene family using 3 gRNAs. In Chapter 3, we describe a computation tool that automates the design of gRNAs, and a key feature, which is the capability to enumerate, for each gRNA, all genomic hits that reside in the target gene family. This allows searching for gRNAs that target more members in a large gene family, such as the *trans*-sialidase gene family, which could significantly reduce the number of gRNAs required for targeting thousands of TcTS genes. In Chapter 4, we first identified 2560 TcTS in a clonal line of *T. cruzi* Brazil strain, which is an optimum strain to conduct TcTS knockdown in. Next, we describe a computation pipeline to parse all gRNAs found in the TcTS genes, and using it, we computed an optimum combination of 21 gRNAs and associated DNA repair templates that could be used to conduct TcTS knockdown. With the sequential delivery of 21 gRNA with SaCas9 and DNA repair templates into *T. cruzi* parasites, we obtained parasite lines that showed significant decrease of surface TcTS levels. We could detect the intended genome editing events in TcTS genes using PCR followed by amplicon sequencing.

1.8 REFERENCES

(WHO), W. H. O. (2010). " Control and prevention of Chagas disease in Europe. Report of a WHO Informal Consultation (jointly organized by WHO headquarters and the WHO Regional Office for Europe Geneva, Switzerland." Geneva: WHO; 2010. Report No: WHO/HTM/NTD/IDM/2010.1.

Abuin, G., L. H. G. Freitas, W. Colli, M. J. M. Alves and S. Schenkman (1999). "Expression of *trans*-sialidase and 85-kDa glycoprotein genes in *Trypanosoma cruzi* is differentially regulated at the post-transcriptional level by labile protein factors." Journal of Biological Chemistry **274**(19): 13041-13047.

Agusti, R., G. Paris, L. Ratier, A. C. C. Frasch and R. M. de Lederkremer (2004). "Lactose derivatives are inhibitors of *Trypanosoma cruzi trans*-sialidase activity toward conventional substrates in vitro and in vivo." Glycobiology **14**(7): 659-670.

Atwood, J. A., D. B. Weatherly, T. A. Minning, B. Bundy, C. Cavola, F. R. Opperdoes, R. Orlando and R. L. Tarleton (2005). "The *Trypanosoma cruzi* proteome." Science **309**(5733): 473-476.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Bern, C., S. Kjos, M. J. Yabsley and S. P. Montgomery (2011). "*Trypanosoma cruzi* and Chagas' Disease in the United States." Clin Microbiol Rev **24**(4): 655-681.

Bixby, L. M. and R. L. Tarleton (2008). "Stable CD8+ T cell memory during persistent *Trypanosoma cruzi* infection." J Immunol **181**(4): 2644-2650.

Buschiazzo, A., M. F. Amaya, M. L. Cremona, A. C. Frasch and P. M. Alzari (2002). "The crystal structure and mode of action of *trans*-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis." Mol Cell **10**(4): 757-768.

Bustamante, J. M., L. M. Bixby and R. L. Tarleton (2008). "Drug-induced cure drives conversion to a stable and protective CD8+ T central memory response in chronic Chagas disease." Nat Med **14**(5): 542-550.

Camargo, M. E., E. L. Segura, I. G. Kagan, J. M. Souza, R. Carvalheiro Jda, J. F. Yanovsky and M. C. Guimaraes (1986). "Three years of collaboration on the standardization of Chagas' disease serodiagnosis in the Americas: an appraisal." Bull Pan Am Health Organ **20**(3): 233-244.

Carvalho, S. T., M. Sola-Penna, I. A. Oliveira, S. Pita, A. S. Goncalves, B. C. Neves, F. R. Sousa, L. Freire-de-Lima, M. Kurogochi, H. Hinou, S. I. Nishimura, L. Mendonca-Previato, J. O. Previato and A. R. Todeschini (2010). "A new class of mechanism-based inhibitors for *Trypanosoma cruzi trans*-sialidase and their influence on parasite virulence." Glycobiology **20**(8): 1034-1045.

Chaves, L. B., M. R. S. Briones and S. Schenkman (1993). "*Trans*-Sialidase from Trypanosoma-Cruzi Epimastigotes Is Expressed at the Stationary-Phase and Is Different from the Enzyme Expressed in Trypomastigotes." Molecular and Biochemical Parasitology **61**(1): 97-106.

Chavez, A., J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, P. R. I. E, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins and G. M. Church (2015). "Highly efficient Cas9-mediated transcriptional programming." Nat Methods **12**(4): 326-328.

Chen, B., L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi and B. Huang (2013). "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system." Cell **155**(7): 1479-1491.

Chuenkova, M. V. and M. PereiraPerrin (2005). "A synthetic peptide modeled on PDNF, Chagas' disease parasite neurotrophic factor, promotes survival and differentiation of neuronal cells through TrkA receptor." Biochemistry **44**(48): 15685-15694.

Cohen, J. E. and R. E. Gurtler (2001). "Modeling household transmission of American trypanosomiasis." Science **293**(5530): 694-698.

Cordeiro, F. D., O. A. Martins-Filho, M. O. D. Rocha, S. J. Adad, R. Correa-Oliveira and A. J. Romanha (2001). "Anti-*Trypanosoma cruzi* immunoglobulin G1 can be a useful tool for diagnosis and prognosis of human Chagas' disease." Clinical and Diagnostic Laboratory Immunology **8**(1): 112-118.

Cremona, M. L., O. Campetella, D. O. Sanchez and A. C. C. Frasch (1999). "Enzymically inactive members of the *trans*-sialidase family from *Trypanosoma cruzi* display beta-galactose binding activity." Glycobiology **9**(6): 581-587.

Cross, G. A., H. S. Kim and B. Wickstead (2014). "Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427." Mol Biochem Parasitol **195**(1): 59-73.

da Silveira, J. F., E. S. Umezawa and A. O. Luquetti (2001). "Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis." Trends Parasitol **17**(6): 286-291.

de Melo-Jorge, M. and M. PereiraPerrin (2007). "The chagas' disease parasite *Trypanosoma cruzi* exploits nerve growth factor receptor TrkA to infect mammalian." Cell Host & Microbe **1**(4): 251-261.

Dias, W. B., F. D. Fajardo, A. V. Graca-Souza, L. Freire-de-Lima, F. Vieira, M. F. Girard, B. Bouteille, J. O. Previato, L. Mendonca-Previato and A. R. Todeschini (2008). "Endothelial cell signalling induced by *trans*-sialidase from *Trypanosoma cruzi*." Cell Microbiol **10**(1): 88-99.

DiCarlo, J. E., J. E. Norville, P. Mali, X. Rios, J. Aach and G. M. Church (2013). "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems." Nucleic Acids Res **41**(7): 4336-4343.

Esvelt, K. M., P. Mali, J. L. Braff, M. Moosburner, S. J. Yaung and G. M. Church (2013). "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing." Nat Methods **10**(11): 1116-1121.

Ferrerogarcia, M. A., S. E. Trombetta, D. O. Sanchez, A. Reglero, A. C. C. Frasch and A. J. Parodi (1993). "The Action of Trypanosoma-Cruzi *Trans*-Sialidase on Glycolipids and Glycoproteins." European Journal of Biochemistry **213**(2): 765-771.

Fineran, P. C. and E. Charpentier (2012). "Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information." Virology **434**(2): 202-209.

Freitas, L. M., S. L. dos Santos, G. F. Rodrigues-Luiz, T. A. O. Mendes, T. S. Rodrigues, R. T. Gazzinelli, S. M. R. Teixeira, R. T. Fujiwara and D. C. Bartholomeu (2011). "Genomic Analyses, Gene Expression and Antigenic Profile of the *Trans*-Sialidase Superfamily of *Trypanosoma cruzi* Reveal an Undetected Level of Complexity." Plos One **6**(10).

Gascon, J., C. Bern and M. J. Pinazo (2010). "Chagas disease in Spain, the United States and other non-endemic countries." Acta Trop **115**(1-2): 22-27.

Gasiunas, G., R. Barrangou, P. Horvath and V. Siksnys (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." Proc Natl Acad Sci U S A **109**(39): E2579-2586.

Giddings, O. K., C. S. Eickhoff, T. J. Smith, L. A. Bryant and D. F. Hoft (2006). "Anatomical route of invasion and protective mucosal immunity in *Trypanosoma cruzi* conjunctival infection." Infect Immun **74**(10): 5549-5560.

Gilbert, L. A., M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman and L. S. Qi (2013). "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes." Cell **154**(2): 442-451.

Glover, L., J. Jun and D. Horn (2011). "Microhomology-mediated deletion and gene conversion in African trypanosomes." Nucleic Acids Res **39**(4): 1372-1380.

Gupta, S. and N. J. Garg (2010). "Prophylactic Efficacy of TcVac2 against *Trypanosoma cruzi* in Mice." Plos Neglected Tropical Diseases **4**(8).

Hilton, I. B., A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy and C. A. Gersbach (2015). "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers." Nat Biotechnol **33**(5): 510-517.

Horn, D. (2014). "Antigenic variation in African trypanosomes." Mol Biochem Parasitol **195**(2): 123-129.

Hou, Z., Y. Zhang, N. E. Propson, S. E. Howden, L. F. Chu, E. J. Sontheimer and J. A. Thomson (2013). "Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis." Proc Natl Acad Sci U S A **110**(39): 15644-15649.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh and J. K. Joung (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nat Biotechnol **31**(3): 227-229.

Imai, K., T. Maeda, Y. Sayama, M. Osa, K. Mikita, I. Kurane, Y. Miyahira, A. Kawana and S. Miura (2015). "Chronic Chagas disease with advanced cardiac complications in Japan: Case report and literature review." Parasitol Int **64**(5): 240-242.

Jackson, A. P., A. Berry, M. Aslett, H. C. Allison, P. Burton, J. Vavrova-Anderson, R. Brown, H. Browne, N. Corton, H. Hauser, J. Gamble, R. Gilderthorp, L. Marcello, J. McQuillan, T. D. Otto, M. A. Quail, M. J. Sanders, A. van Tonder, M. L. Ginger, M. C. Field, J. D. Barry, C. Hertz-Fowler and M. Berriman (2012). "Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species." Proc Natl Acad Sci U S A **109**(9): 3416-3421.

Jiang, W., D. Bikard, D. Cox, F. Zhang and L. A. Marraffini (2013). "RNA-guided editing of bacterial genomes using CRISPR-Cas systems." Nat Biotechnol **31**(3): 233-239.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.

Kipnis, T. L., A. U. Krettli and W. Dias da Silva (1985). "Transformation of trypomastigote forms of *Trypanosoma cruzi* into activators of alternative complement pathway by immune IgG fragments." Scand J Immunol **22(**2): 217-226.

Klann, T. S., J. B. Black, M. Chellappan, A. Safi, L. Song, I. B. Hilton, G. E. Crawford, T. E. Reddy and C. A. Gersbach (2017). "CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome." Nat Biotechnol **35**(6): 561-568.

Konermann, S., M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki and F. Zhang (2015). "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex." Nature **517**(7536): 583-588.

Kumar, S. and R. L. Tarleton (1998). "The relative contribution of antibody production and CD8(+) T cell function to immune control of *Trypanosoma cruzi*." Parasite Immunology **20**(5): 207-216.

Kurup, S. P. and R. L. Tarleton (2013). "Perpetual expression of PAMPs necessary for optimal immune control and clearance of a persistent pathogen." Nat Commun **4**: 2616.

Kurup, S. P. and R. L. Tarleton (2014). "The *Trypanosoma cruzi* Flagellum Is Discarded via Asymmetric Cell Division following Invasion and Provides Early Targets for Protective CD8(+) T Cells." Cell Host & Microbe **16**(4): 439-449.

Liu, X., Y. Zhang, Y. Chen, M. Li, F. Zhou, K. Li, H. Cao, M. Ni, Y. Liu, Z. Gu, K. E. Dickerson, S. Xie, G. C. Hon, Z. Xuan, M. Q. Zhang, Z. Shao and J. Xu (2017). "In Situ Capture of Chromatin Interactions by Biotinylated dCas9." Cell **170**(5): 1028-1043 e1019.

Ma, H., L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald and T. Pederson (2016). "Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow." Nat Biotechnol **34**(5): 528-530.

Martin, D. and R. Tarleton (2004). "Generation, specificity, and function of CD8+ T cells in *Trypanosoma cruzi* infection." Immunol Rev **201**: 304-317.

Martin, D. L., D. B. Weatherly, S. A. Laucella, M. A. Cabinian, M. T. Crim, S. Sullivan, M. Heiges, S. H. Craven, C. S. Rosenberg, M. H. Collins, A. Sette, M. Postan and R. L. Tarleton (2006). "CD8+ T-Cell responses to *Trypanosoma cruzi* are highly focused on strain-variant *trans*-sialidase epitopes." PLoS Pathog **2**(8): e77.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Millar, A. E., M. Wleklinski-Lee and S. J. Kahn (1999). "The surface protein superfamily of *Trypanosoma cruzi* stimulates a polarized Th1 response that becomes anergic." J Immunol **162**(10): 6092-6099.

Mugnier, M. R., G. A. Cross and F. N. Papavasiliou (2015). "The in vivo dynamics of antigenic variation in Trypanosoma brucei." Science **347**(6229): 1470-1473.

Nakayama, T., M. B. Fish, M. Fisher, J. Oomen-Hajagos, G. H. Thomsen and R. M. Grainger (2013). "Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in Xenopus tropicalis." Genesis **51**(12): 835-843.

Padilla, A., D. Xu, D. Martin and R. Tarleton (2007). "Limited role for CD4(+) T-cell help in the initial priming of *Trypanosoma cruzi*-specific CD8(+) T cells." Infection and Immunity **75**(1): 231-235.

Padilla, A. M., J. M. Bustamante and R. L. Tarleton (2009). "CD8+ T cells in *Trypanosoma cruzi* infection." Curr Opin Immunol **21**(4): 385-390.

Padilla, A. M., L. J. Simpson and R. L. Tarleton (2009). "Insufficient TLR activation contributes to the slow development of CD8+ T cell responses in *Trypanosoma cruzi* infection." J Immunol **183**(2): 1245-1252.

Ran, F. A., L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. V. Koonin, P. A. Sharp and F. Zhang (2015). "In vivo genome editing using Staphylococcus aureus Cas9." Nature **520**(7546): 186-191.

Reyon, D., S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander and J. K. Joung (2012). "FLASH assembly of TALENs for high-throughput genome editing." Nat Biotechnol **30**(5): 460-465.

Rosenberg, C. S., D. L. Martin and R. L. Tarleton (2010). "CD8+ T cells specific for immunodominant *trans*-sialidase epitopes contribute to control of *Trypanosoma cruzi* infection but are not required for resistance." J Immunol **185**(1): 560-568.

Rosenberg, C. S., W. Zhang, J. M. Bustamante and R. L. Tarleton (2016). "Long-Term Immunity to *Trypanosoma cruzi* in the Absence of Immunodominant *trans*-Sialidase-Specific CD8+ T Cells." Infect Immun **84**(9): 2627-2638.

Schenkman, R. P. F., F. Vandekerckhove and S. Schenkman (1993). "Mammalian-Cell Sialic-Acid Enhances Invasion by Trypanosoma-Cruzi." Infection and Immunity **61**(3): 898-902.

Schenkman, S., L. P. Decarvalho and V. Nussenzweig (1992). "Trypanosoma-Cruzi *Trans*-Sialidase and Neuraminidase Activities Can Be Mediated by the Same Enzymes." Journal of Experimental Medicine **175**(2): 567-575.

Shan, Q., Y. Wang, J. Li, Y. Zhang, K. Chen, Z. Liang, K. Zhang, J. Liu, J. J. Xi, J. L. Qiu and C. Gao (2013). "Targeted genome modification of crop plants using a CRISPR-Cas system." Nat Biotechnol **31**(8): 686-688.

Silva, G., L. Poirot, R. Galetto, J. Smith, G. Montoya, P. Duchateau and F. Paques (2011). "Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy." Curr Gene Ther **11**(1): 11-27.

Smith, J., S. Grizot, S. Arnould, A. Duclert, J. C. Epinat, P. Chames, J. Prieto, P. Redondo, F. J. Blanco, J. Bravo, G. Montoya, F. Paques and P. Duchateau (2006). "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences." Nucleic Acids Res **34**(22): e149.

Tarleton, R. L. (1990). "Depletion of Cd8+ T-Cells Increases Susceptibility and Reverses Vaccine-Induced Immunity in Mice Infected with Trypanosoma-Cruzi." Journal of Immunology **144**(2): 717-724.

Tarleton, R. L. (2007). "Immune system recognition of *Trypanosoma cruzi*." Curr Opin Immunol **19**(4): 430-434.

Tarleton, R. L., B. H. Koller, A. Latour and M. Postan (1992). "Susceptibility of Beta-2-Microglobulin-Deficient Mice to Trypanosoma-Cruzi Infection." Nature **356**(6367): 338-340.

Tarleton, R. L., J. Sun, L. Zhang and M. Postan (1994). "Depletion of T-cell subpopulations results in exacerbation of myocarditis and parasitism in experimental Chagas' disease." Infect Immun **62**(5): 1820-1829.

Todeschini, A. R., W. B. Dias, M. F. Girard, J. M. Wieruszeski, L. Mendonca-Previato and J. O. Previato (2004). "Enzymatically inactive *trans*-sialidase from *Trypanosoma cruzi* binds sialyl and beta-galactopyranosyl residues in a sequential ordered mechanism." J Biol Chem **279**(7): 5323-5328.

Tomlinson, S., L. P. Decarvalho, F. Vandekerckhove and V. Nussenzweig (1992). "Resialylation of Sialidase-Treated Sheep and Human Erythrocytes by Trypanosoma-Cruzi *Trans*-Sialidase - Restoration of Complement Resistance of Desialylated Sheep Erythrocytes." Glycobiology **2**(6): 549-551.

Tzelepis, F., P. M. Persechini and M. M. Rodrigues (2007). "Modulation of CD4(+) T cell-dependent specific cytotoxic CD8(+) T cells differentiation and proliferation by the timing of increase in the pathogen load." PLoS One **2**(4): e393.

Uemura, H., S. Schenkman, V. Nussenzweig and D. Eichinger (1992). "Only Some Members of a Gene Family in Trypanosoma-Cruzi Encode Proteins That Express Both *Trans*-Sialidase and Neuraminidase Activities." Embo Journal **11**(11): 3837-3844.

Urbina, J. A. (2015). "Recent clinical trials for the etiological treatment of chronic chagas disease: advances, challenges and perspectives." J Eukaryot Microbiol **62**(1): 149-156.

Vandekerckhove, F., S. Schenkman, L. P. Decarvalho, S. Tomlinson, M. Kiso, M. Yoshida, A. Hasegawa and V. Nussenzweig (1992). "Substrate-Specificity of the Trypanosoma-Cruzi *Trans*-Sialidase." Glycobiology **2**(6): 541-548.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.

Weatherly, D. B., D. Peng and R. L. Tarleton (2016). "Recombination-driven generation of the largest pathogen repository of antigen variants in the protozoan *Trypanosoma cruzi*." BMC Genomics **17**(1): 729.

Weinkauf, C., R. Salvador and M. PereiraPerrin (2011). "Neurotrophin Receptor TrkC Is an Entry Receptor for *Trypanosoma cruzi* in Neural, Glial, and Epithelial Cells." Infection and Immunity **79**(10): 4081-4087.

Wiedenheft, B., S. H. Sternberg and J. A. Doudna (2012). "RNA-guided genetic silencing systems in bacteria and archaea." Nature **482**(7385): 331-338.

Yang, D., J. Xu, T. Zhu, J. Fan, L. Lai, J. Zhang and Y. E. Chen (2014). "Effective gene targeting in rabbits using RNA-guided Cas9 nucleases." J Mol Cell Biol **6**(1): 97-99.

Yu, Z., M. Ren, Z. Wang, B. Zhang, Y. S. Rong, R. Jiao and G. Gao (2013). "Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila." Genetics **195**(1): 289-291.

CHAPTER 2

CRISPR-CAS9 MEDIATED SINGLE GENE AND GENE FAMILY DISRUPTION IN *TRYPANOSOMA CRUZI*

Duo Peng, Samarchith P. Kurup, Phil Y. Yao, Todd A. Minning, and Rick L. Tarleton. 2015. Mbio. vol. 6 issue 1, doi:10.1128/mBio.02097-14. Reprinted here with permission of the publisher

2.1 ABSTRACT

Trypanosoma cruzi is a protozoan parasite of humans and animals, affecting 10-20 million people and innumerable animals primarily in the Americas. Despite being the largest cause of infection-induced heart disease worldwide, even amongst the neglected tropical diseases (NTDs), *T. cruzi* is considered one of the least well-understood and understudied. The genetic complexity of *T. cruzi* as well as the limited set of efficient techniques for genome engineering contribute significantly to the relative lack of progress in and understanding this pathogen. Here we adapted the CRISPR-Cas9 system for the genetic engineering of *T. cruzi*, demonstrating rapid and efficient knockout of multiple endogenous genes, including essential genes. We observed that in the absence of a template, repair of the Cas9-induced double-stranded breaks (DSB) in *T. cruzi* occurs exclusively by microhomology-mediated end-joining (MMEJ) with variable sized deletions. When a template for DNA repair is provided, DSB repair by homologous recombination is achieved at an efficiency several orders of magnitude higher than in the absence of CRISPR-Cas9-induced DSB. We also demonstrate the high multiplexing capacity of CRISPR-Cas9 in *T. cruzi* by knocking-down expression of an enzyme gene family consisting of 65 members, resulting in a significant reduction of enzymatic product with no apparent off-target mutations. Lastly, we showed that Cas9 can mediate disruption of its own coding sequence, rescuing a growth defect in stable Cas9-expressing parasites. These results establish a powerful new tool for the analysis of gene functions in *T. cruzi*, enabling the study of essential genes and their function and analysis of the many large families of related genes that occupy a substantial portion of the *T. cruzi* genome.

2.2 IMPORTANCE

Trypanosoma cruzi, the causative agent of human Chagas disease is the leading worldwide cause of infectious myocarditis. Diagnostics for the infection are relatively poor, treatment options are limited and of variable effectiveness, and suitable vaccines are nonexistent. The *T. cruzi* genome is replete with genes of unknown function and greatly expanded families of genes of 100s of members. The absence of facile genetic engineering tools, including RNAi in *T. cruzi* has prevented elucidation of gene and gene family function and the development of better infection prevention and control measures. In this study, we demonstrate the CRISPR-Cas9 system as a versatile and powerful tool for genome manipulations in *T. cruzi*, bringing new opportunities for unraveling the functions of previously uncharacterized genes and how this human pathogen engages its large gene families encoding surface proteins to interact with human and animal hosts.

2.3 INTRODUCTION

The protozoan parasite *Trypanosoma cruzi* is the causative agent of Chagas disease, the highest impact infectious disease of the Americas with 10-20 million humans and innumerable animals affected. The study of *T. cruzi* and Chagas disease is particularly challenging, in part due to the complexity and unique characteristics of its genome and the relative paucity of tools to manipulate these characteristics and thus determine their importance for parasite persistence and pathogenicity. In addition to the substantial number of genes lacking homologues in other eukaryotes, the *T. cruzi* genome also contains an unprecedented number of gene families, in some cases with thousands of members (Weston, Patel et al. 1999, El-Sayed, Myler et al. 2005, De Pablos and Osuna 2012). Among the largest of these genes families are those encoding *trans*-sialidase-like, mucins and mucin-associated proteins that are expressed on the parasite surface and thus directly interact with the insect and animal hosts – including as immunological targets.

Although the methods to express exogenous or over-express endogenous genes and to delete genes in *T. cruzi* have been useful, these methods are laborious and time-consuming. For example, current gene knockout (KO) strategies utilize spontaneous homologous recombination of a DNA cassette containing a drug selection marker generally flanked by ~500 bp of coding sequence (CDS) or untranslated regions (UTR) of the target gene. In addition to the fact that homologous recombination in *T. cruzi* has a very low efficiency, this approach is limited to a single allele KO per drug selection marker (Xu, Brandan et al. 2009) and the drug selection process is slow, requiring at least 1 month per allele. Collectively, the success rate for generating null mutants in *T. cruzi* is low and the limited number of drug selectable markers restricts the number of manipulations that can be attempted in a single organism. These constraints, in combination with the absence of a functional RNAi system in *T. cruzi*, makes virtually

unapproachable the manipulation of multi-gene families and the determination of how gene family size is generated, maintained and contributes to parasite success in *T. cruzi*.

RNA guided nucleases utilizing clustered, regularly interspaced, short palindromic repeats– CRISPR-associated (CRISPR-Cas) has enabled rapid, targeted modification of a wide range of genomes (Sander and Joung 2014). The system has proven especially useful because of its relative ease and high efficiency as well as the ability to achieve multiple modifications in a single organism/cell (Niu, Shen et al. 2014). The specificity and targeted genome editing by CRISPR-Cas9 is achieved by a guide RNA that directs the Cas9 protein to genome locations by RNA-DNA hybridization, introducing a double-stranded break (DSB). In most species, the repair of the DSBs occurs by a non-homologous end joining (NHEJ) pathway, creating insertions or deletions (indels), or if in the presence of an appropriate DNA template, by homologous recombination.

Here we report the use of the CRISPR-Cas9 system in *T. cruzi* to knockout target genes and to enhance gene insertion by homologous recombination. Gene disruption is highly efficient, with up to 70% of the population exhibiting a mutant phenotype, and rapid, with decreased protein levels evident as early as 2 days after transfection. Because of these qualities, observation of the impact of disruption of essential genes was possible, with an efficiency rivaling that of RNAi. In the absence of a template, repair of the Cas9-induced DSB in *T. cruzi* occurs exclusively by microhomology-mediated end-joining (MMEJ) with variable sized deletions, depending on the location of homologous regions. This latter finding confirms the apparent absence of NHEJ and the dominance of MMEJ repair mechanisms in kinetoplastids (Burton, McBride et al. 2007, Glover, Jun et al. 2011). Lastly, we provide proof-of-concept that CRISPR-Cas9 system can be multiplexed to knockout multiple genes in a large (>50 member)

gene family, with no apparent off-target mutations. These results establish a powerful new tool for genome manipulation in *T. cruzi* and open the door to greater understanding of the roles of essential genes and large gene families in the biology of this human pathogen and its interactions with its animal hosts.

2.4 RESULTS

High frequency Cas9-sgRNA-mediated gene disruption in *T. cruzi*

To determine the ability of sgRNA-guided Cas9 to disrupt genes in *T. cruzi*, we first stably expressed both eGFP and Cas9 in *T. cruzi* using separate pTrex backbone (Vazquez and Levin 1999) plasmids under G418 and blasticidin drug selection, respectively (Fig. 2.1A). Transfection of epimastigotes of *T. cruzi* with single guide RNA (sgRNA) previously shown to mediate eGFP disruption in human cell lines (Fu, Sander et al. 2014) resulted in rapid and highly efficient reduction in GPF expression. Each of the three sgRNA induced loss of GFP in ~50-60% of parasites as early as day 2 after transfection (Fig. 2.1B), No reduction in eGFP expression was observed in epimastigotes transfected with 80bp human 18S rRNA as a control. GFP-targeted sgRNAs were also very efficient at disrupting gene expression when electroporated into trypomastigotes of *T. cruzi*, with vero cells infected with recently transfected trypomastigotes showing a mixture of GFP-positive and GFP negative parasites 5 days after transfection/infection (Fig. 2.1D).

Although the Cas9-mediated mutation of GFP was highly efficient, 40% or more of the population of parasites transfected with sgRNA showed no change in GFP expression levels. Simultaneous (Fig. 2.S1) or serial (Fig. 2.S2) transfection with multiple GFP-directed guides only modestly increased the frequency of eGFP mutations as compared to a single guide.

Additionally, increasing the concentration of sgRNA to >10 µg/107 parasites failed to impact the frequency of eGFP mutants (Fig. 2.S3). To determine if this "resistance" to Cas9-mediated mutation might be linked to Cas9 protein levels, we FACS-sorted the GFP-positive and -negative parasites following transfection with the eGFP sgRNA and measured Cas9 levels by ELISA. On average, the GFP-positive parasites have significantly lower levels of Cas9 protein compared to parasites in which GFP was disrupted by sgRNA transfection (Fig. 2.1C), suggesting that the less than 100% efficiency of gene disruption in this system is due to low and variable levels of Cas9 expression.

2.4.1 SGRNA-GUIDED CAS9 MUTATION OF ENDOGENOUS GENES

In order to validate the use of CRISPR-Cas9 to mutate endogenous genes, we designed sgRNA targeting a number of *T. cruzi* genes using a custom sgRNA design tool that selects sgRNA based in part on the absence of predicted off-targets (available at http://grna.ctegd.uga.edu). Transfection of *T. cruzi* epimastigotes or trypomastigotes with sgRNA targeting the multi-copy α-tubulin genes, resulted in parasites with misshapen and enlarged cell bodies and multiple flagella (Fig. 2.2 A,B). Similar defects in cytokinesis and cell shape had been previously reported following RNAi-mediated knockdown of α-tubulin in Trypanosoma brucei (Ngo, Tschudi et al. 1998).

To better estimate the efficiency of endogenous gene knockout using sgRNA-guided Cas9, we performed transfections of sgRNA targeting single-locus genes encoding histidine ammonia-lyase (HAL), an enzyme in the histidine metabolism pathway whose enzymatic activity could be easily quantified, and the putative fatty acid transporter (FATP) gene, whose protein actively could be monitored by uptake of BODIPY®-labeled fatty acids. Epimastigotes transfected with the HAL sgRNA exhibited a 60% decrease in HAL activity at day 4 post-

transfection compared to parasites transfected with the control 18s RNA (Fig. 2.2C). The epimastigotes were cloned by limiting dilution and assayed for HAL activity. Of the four clones tested, three exhibited no HAL activity, suggesting KO of both HAL alleles, and one (B6) demonstrating HAL activity similar to epimastigotes transfected with control 18s RNA. sgRNA-guided Cas9 targeting of FATP resulted in a 37% decrease in fatty acid uptake rate (Fig. 2.2D). Our previous attempts to generate null mutants in FATP by conventional knockout strategies (Xu, Brandan et al. 2009) have failed, indicating that null mutation of FATP in *T. cruzi* was probably lethal. Indeed, examination of FA import activity in FATP-sgRNA parasites at 2 weeks post-transfection showed a near normal FA uptake rate, suggesting the loss of the FATP-mutant population and survival of only the non-mutated WT population. Thus, the high efficiency of the CRISPR-Cas9 system in *T. cruzi* allows for the study of loss of function over time due to the disruption of essential genes.

2.4.2 MMEJ-MEDIATED REPAIR OF CRISPR-CAS9-INDUCED DSB IN *T. CRUZI*

Double-stranded DNA breaks induced by guided nucleases such as Cas9 are generally repaired in one of 2 ways: by error prone NHEJ, resulting in indels of varying sizes or by homology-directed repair which allows precise editing - from point mutations to large indels - depending on the available template. Sequencing of the eGFP gene from eGFP-negative clones produced by eGFP-sequence-guided Cas9 DSB showed a consistent 33bp deletion at the sgRNA targeting site (Fig. 2.3). Although the deletion junctions occur at slightly different positions, they all fall between a pair of homologous sequences of 10 bp flanking the cut site (Fig. 2.3, red highlight). This pattern is consistent with a microhomology-mediated ending joining (MMEJ) pathway repair of DSB in *T. cruzi*. Sequencing of the GFP gene from clones of the eGFP sgRNA-transfected population which remained GFP-positive (Fig. 2.3) showed an intact GFP

sequence, again consistent with insufficient Cas9-mediated DSB in some parasites (data not shown).

2.4.3 CAS9-FACILITATED HOMOLOGOUS RECOMBINATION

Template-mediated repair of sgRNA-guided Cas9 cuts has been used to facilitate homology directed repair (HDR) (Dickinson, Ward et al. 2013, Schwank, Koo et al. 2013, Auer, Duroure et al. 2014, Bottcher, Hollmann et al. 2014, Ghorbal, Gorman et al. 2014). In the presence of template DNA with sufficient homologous flanking sequences, homologous recombination can be induced in *T. cruzi* and has been extensively used for deletion of specific genes (Xu, Brandan et al. 2009). However, this process appears to have very low efficiency and requires ~30 days of drug selection to obtain stable recombinants. To determine if sgRNA-guided Cas9 cuts could be used to achieve higher rates of homologous recombination in *T. cruzi*, we conducted a fluorescence marker swap assay. *T. cruzi* epimastigotes harboring an eGFP expression cassette were co-transfected with both a sgRNA targeting eGFP and a tdTomato expression cassette with 5' and 3' homology to the eGFP insert (Fig. 2.4A). As expected from the high rate of mutation observed with eGFP and other endogenous genes, the sgRNA transfection resulted in a predominant loss of eGFP fluorescence (Fig. 2.4B). However, the sgRNA targeting eGFP at position 100bp and 152bp respectively yielded a 0.11% and 0.069% rate of fluorescence marker swap. In stark contrast, in the absence of the eGFP-targeted sgRNA (but in the presence of the tdTomato template), homologous recombination was below the level of detection by this assay.

Cas9 nickase (Cas9n), a mutant form of Cas9 which cuts only a single strand of the dsDNA has been successfully used in other organisms to favor gene repair by HDR over the production of mutations commonly produced by NHEJ repair of DSB (Ran, Hsu et al. 2013).

However, transfection of neither a single or double sgRNA (at eGFP position 313 and 339) into Cas9 nickase expressing epimastigotes of *T. cruzi* produced detectable replacement of eGFP with the supplied tdTomato template (Fig. 2.S4, S5).

2.4.4 MUTATION OF A MULTIGENE FAMILY IN *T. CRUZI*

In addition to the time required to KO multiple alleles in *T. cruzi* using conventional homologous recombination with drug selectable markers (Xu, Brandan et al. 2009), the genetic manipulation of *T. cruzi* is also challenging because of the high proportion of genes within moderate to large families (El-Sayed, Myler et al. 2005). The success with the apparent disruption of multiple α-tubulin genes (*T. cruzi* is thought to have 10-18 loci for α-tubulin) prompted us to test if RNA-guided Cas9 could disrupt larger gene families in *T. cruzi*. For this purpose, we selected the beta galactofuranosyl glycosyltransferase (β-GalGT) family of 65 annotated genes (Table 2.S3). Members of this gene family share an average of 93.1% nucleotide sequence homology, making it possible to target the whole gene family with as few as 3 sgRNAs (Fig 2.5A). We conducted sequential transfections of the 3 guides, assessing surface β-galactosyl residues after each transfection using a fluorescently-labeled peanut agglutinin (PNA) lectin specific for Gal-β(1-3)-GalNAc. Flow cytometric analysis demonstrated progressive reduction in surface beta-galactosyl residues with each delivery of an additional sgRNA (Fig. 2.5B).

To further assess the genome-wide efficiency of β-GalGT mutation and to determine potential off-target mutations induced by Cas9 armed with multiple sgRNA, we conducted whole-genome sequencing of the uncloned parasite population that had received 3 sgRNA and compared this to Cas9-expressing cells not receiving sgRNAs. For guide sites 1, 2 and 3, novel junctions indicative of gene deletions were detected in 35, 64, 27 reads, respectively (Fig. 2.5C).

These reads accounted for 31, 30 and 23 percent of the total reads that maps to corresponding target regions, indicating that collectively 63% of the β-GalGT genes have a deletion in at least one target site. This calculation matches closely to the observed loss of surface galactose residues of approximately 58% as assayed by lectin staining (Fig. 2.5B). As with previous single-copy gene targeting by Cas9-sgRNA in *T. cruzi*, the mutations induced were all deletions (of 101, 14, and 162 bp for sites 1, 2 and 3, respectively) and were all associated with regions of microhomology, again supporting MMEJ as the mechanism of repair of DSB in *T. cruzi* (Fig. 2.5C).

Potential off-target regions are identified in the genome as sequences that matched to sgRNA targeting sequence with less than 6 mismatches but excluding any mismatching in the PAM motif since this motif is required for Cas9 DSB (Cho, Kim et al. 2014). By this criterion, all 3 guides have a total of 2 potential off-target sites in the *T. cruzi* genome and with 100x and 80x coverage in these regions, no mutations in these sites were observed. As deep sequencing read mappers tend to have low tolerance for indels, we used custom Perl scripts to search unmapped reads for indels in the 2 possible off-target sites and again found no indels supported by >2 similar reads. Thus we conclude the multiple genes can be simultaneously mutated in the *T. cruzi* genome without production of detectable off-target mutations.

We have previously observed that *T. cruzi* lines expressing selected exogenous proteins occasionally have altered growth kinetics. Examination of *T. cruzi* epimastigotes with stable Cas9 expression indicated a substantial increase in doubling time (Fig. 2.6A). To attempt to rescue WT growth in these Cas9-expressing lines, we designed a sgRNA targeting Cas9 and used these to disrupt the Cas9 gene, rapidly returning these lines to wild-type growth (Fig. 2.6A). Further evidence of the disruption of the Cas9 gene by Cas9-sgRNA-guided Cas9 protein was

obtained by the failure of eGFP-sgRNA to alter eGFP expression in the lines transfected with the Cas9 sgRNA (Fig. 2.6B). Endogenous *T. cruzi* genes and the Cas9 gene can also be simultaneously disrupted. Cotransfection of HAL-sgRNA and Cas9-sgRNA resulted in clones devoid of HAL activity and with defective Cas9 (Fig. 2.6C). Thus, Cas9 can mediate disruption of its own coding gene, making it possible to perform genome modifications in *T. cruzi* lines expressing Cas9 and simultaneously "killing" Cas9 expression.

2.5 DISCUSSION

Understanding the complexities of host:pathogen interactions is greatly facilitated by the ability to manipulate host and pathogen genomes – by gene disruption or by insertion of genes with new or enhanced functions. While manipulation of the *T. cruzi* genome has been possible for some time, the processes to achieve modifications are not rapid, easy or routine. We have previously reported on a multisite Gateway-approach for the more facile production of constructs for gene disruption in *T. cruzi* (Xu, Brandan et al. 2009). However because of the low rate of homologous recombination and the relatively high resistance of *T. cruzi* to antibiotic selection, production of null mutants of single-copy genes using this approach requires a minimum of several months and the use of multiple antibiotic resistance genes. The limited number of available resistance genes makes knockout of more than 1 or 2 genes unmanageable using this approach. The latter limitation is a particular problem in *T. cruzi* which has over 100 gene families with 4 or more members and several families containing 100's of genes. And unlike its closest genetic relatives, the African trypanosomes, *T. cruzi* lacks the machinery for inhibitory RNAs (DaRocha, Otsu et al. 2004, Kolev, Tschudi et al. 2011), making suppression of gene function by RNAi unfeasible.

The CRISPR-Cas system has rapidly transformed the speed and ease of gene manipulation in multiple species (Sander and Joung 2014). By adapting the CRISPR-Cas9 for use in *T. cruzi*, we were able to quickly and efficiently disrupt endogenous single-copy and multi-copy genes, as well as exogenous genes. Because of the high rate of gene mutation using CRISPR-Cas9 in *T. cruzi* – routinely 60-70% double/multi-allelic mutations - the impact of gene disruption can be observed within days when appropriate assays are available (e.g. fluorescence for GFP, enzyme activity for HAL and fatty acid uptake for FATP) instead of months as in the case of conventional knockouts. For genes lacking assays that can be applied to a parasite population, the principal limiting factor with CRISPR-Cas-mediated mutation is the time required to generate sufficient numbers of parasites from clones in order to confirm the mutations/phenotypes.

The high efficiency of CRISPR-Cas-induced mutations in *T. cruzi* also means that null mutants in essential genes can be generated and monitored over time as protein activity is lost. Our previous studies suggested that genes involved in fatty acid uptake and beta oxidation (Xu, Brandan et al. 2009) were essential in *T. cruzi*. We further support that conclusion here with the observation that targeting FATP for mutation results in a population of parasites with decreased FA uptake soon after transfection and a return to normal FA uptake by 2 weeks post-transfection, presumably due to the death of the null mutants in the population.

We also show that CRISPR-Cas9 could greatly facilitate HR between a supplied template and specific locations in the genome. By introducing a DSB with RNA-guided Cas9, we were able to replace a genomic eGFP sequence with a larger tdTomato sequence at a frequency several orders of magnitude greater than in the absence of the DSB. This increased frequency makes feasible a number of additional approaches to manipulation of *T. cruzi*, including the

tagging of genes in their endogenous loci and rapid isolation of gene replacement mutants by flow sorting based on a knocked-in of a fluorescent protein.

Although in the presence of an appropriate template, HR is observed at a detectable frequency following the introduction of RNA-guided DSB, mutations resulting from error-prone end-joining appear to be the preferred mechanism of repairing these DSB in *T. cruzi*. For multiple analyzed mutations, we find strong evidence for microhomology-mediated end joining (MMEJ) as the mechanism of repair of DSB in *T. cruzi*. MMEJ has been considered a "backup" pathway for DSB repair in most species but appears to be the primary mechanism in trypanosomes wherein nonhomologous end-joining (NHEJ) is absent (Glover, Jun et al. 2011). MMEJ generally uses 5-20 bp regions of microhomology to repair DSB, in the process leaving deletions between the microhomology regions (McVey and Lee 2008). As such, the deletion size should be predictable, once the specific rules for MMEJ are defined (e.g., the relative contribution of microhomology length and distance from the cut site to the repair) and these data can be employed to more efficiently design sgRNAs (Bae, Kweon et al. 2014). There is also increasing appreciation for the role of MMEJ in DSB repair in species where NHEJ dominates (Ottaviani, LeCain et al. 2014) and emerging information on the putative components of the MMEJ repair machinery (Crespan, Czabany et al. 2012, Ottaviani, LeCain et al. 2014). Because NHEJ is naturally absent in *T. cruzi* and other kinetoplastids, these organisms should be especially useful for the further dissection of MMEJ mechanisms. Additionally, knowledge of the components uniquely involved in MMEJ could provide insights into how to minimize this error prone repair pathway and potentially enhance HR, further optimizing in situ genome editing in *T. cruzi*.

The potential for off-target sequence disruption is always a concern when using genome editing tools, including CRISPR/Cas (Fu, Foden et al. 2013). However because of the compactness and low complexity of the *T. cruzi* genome, and with the aid of sgRNA finder program we developed for *T. cruzi*, we were able to select highly gene-specific sgRNA with low sequence identity (<15bp matches in the 20 bp guide) to other sites in the *T. cruzi* genome. Assuring that any potential off-target sites also lacked NGG motifs needed for Cas9 nuclease activity also contributes to minimize off-target effects. This specificity and the lack of off-target effects were confirmed by analysis of deep sequencing data from parasites exposed to multiple sgRNAs.

A number of studies (Cong, Ran et al. 2013, Jao, Wente et al. 2013, Wang, Yang et al. 2013) have documented the ability to target more than 1 gene for mutation using the CRISPR/Cas system. We show here that the potential to multiplex CRISPR/Cas to more than 50 genes using a small number of sgRNAs. This experiment is only possible in organisms like *T. cruzi* that contain large numbers of closely related genes. *T. cruzi* genome contains a remarkable number of moderate (>20), large (>50), and very large (>500) gene families, among these are ones encoding the thousands of mucins, mucin-associated proteins and *trans*-sialidase-like proteins that form a large part of the interface between the parasite and mammalian hosts. Modest expansion of gene families in some organisms has been linked to the potential to succeed in variable environments (Tamate, Kawata et al. 2014). However, the size of the *trans*-sialidase gene family, its variability in composition among different parasite isolates and its targeting by host immune responses all argue that the family may have expanded from the few copies identified in other kinetoplastids to its current size in *T. cruzi*, in part as an immune evasion mechanism. Testing of this hypothesis has not previously been possible due to the lack of

functional RNAi or other knockdown/knockout system capable of regulating the expression of 100's of genes. The demonstration here of the ability to knockdown a gene family of 65 members may pave the way for studying these much larger gene families *T. cruzi*.

CRISPR/Cas has revolutionized genome editing in multiple species and we show here the similar promise for use in the kinetoplastid parasite *T. cruzi*. As already noted, the exclusive dependence on MMEJ for DSB repair and the enormous multiplexing capabilities are observations unique to the CRISPR/Cas system in *T. cruzi* at the current time. Cas9 expression in *T. cruzi* does come at the price of decreased growth potential, a consequence also noted in yeast when Cas9 was highly expressed (Ryan, Skerker et al. 2014), presumably due to Cas9's ability to bind to DNA at PAM motifs without sgRNA. But this effect was reversible upon mutation of the Cas9 gene (using the CRISPR/Cas system itself), providing the potential to simultaneously mutate endogenous *T. cruzi* genes and Cas9 using a mixture of sgRNAs. Hopefully, this study is only the beginning of the exploitation of the CRISPR/Cas system in kinetoplastids.

2.6 MATERIALS AND METHODS

2.6.1 GROWTH, TRANSFECTION AND CLONING OF *T. CRUZI*

Epimastigotes of the CL strain of *T. cruzi* were cultured at 26 °C in supplemented liver digested-neutralized tryptose (LDNT) medium as described previously (Xu, Brandan et al. 2009). Unless otherwise indicated, 5 × 106 early-log epimastigotes or recently egressed trypomastigotes were resuspended in 100 µl room temperature Human T Cell Nucleofector™ Solution (Amaxa AG, Cologne, Germany) and 20 µg sgRNA in a total volume of 15 µl and electroporated using program "U-33" in an AMAXA nucleofector device. For sequential transfections, parasites were allowed 5 days recovery time between transfection. For tests of

homology recombination an additional 10ug of template DNA was added. The electroporated parasites were cultured in 25 cm2 cell culture flasks (Corning Incorporated, Lowell, MA, USA) with 10 ml LDNT medium. To generate *T. cruzi* lines stably expressing eGFP and Cas9, 10 μg linearized pTrex-n-eGFP-Neo plasmid was transfected into epimastigotes using protocol described above. 250 μg/ml G418 was added at 24 hrs post-transfection and drug concentration was maintained for 4 weeks (at which point parasites transfected with no DNA were no longer viable), eGFP positive parasites was then sorted by a MoFlow (Dako-Cytomation, Denmark) cell sorter. Sorted eGFP positive parasites were transfected with pTrex-b-NLS-hSpCas9, 25 μg/ml blasticidin was added 24 hrs post transfection and drug concentration was maintained for 4 weeks.

The doubling time of epimastigotes was calculated by fitting an exponential curve to density data from day 1 to 14 of culture using least squares fitting at http://www.doubling-time.com/compute.php?lang=en.

2.6.2 PLASMID CONSTRUCTION

The *T. cruzi* pTrex-n-eGFP and pTrex-b-NLS-hSpCas9 plasmids were constructed by subcloning the coding sequence of hSpCas9 from pX330 (Cong, Ran et al. 2013) or eGFP [Genbank accession: JQ693016.1 , 633 to 1352 bp], respectively into multiple cloning site of the pTrex plasmid (Vazquez and Levin 1999) containing neomycin phosphotransferase gene (pTrex-n) or blasticidin-S deaminase gene (pTrex-b).

2.6.3 SGRNA PREPARATION

sgRNA targeting sequences were designed using a custom sgRNA design tool (available at sgRNA.ctegd.uga.edu) which 1) identifies all potential 20 bp sequences containing an 'NGG'

PAM site within the query sequence, 2) predicts all potential off-target sites for each sgRNA, including those with 5 or fewer mismatches to the sgRNA, and 3) indicates the microhomology pairs flanking the sequences targeted by the identified sgRNA. 4) predicts targeting efficiency using a positional specific nucleotide composition scoring matrix (Doench, Hartenian et al. 2014). A list of ranked sgRNA targeting sequences is returned based on minimal off-targets and minimal flanking distance of microhomology pairs and maximum length of microhomology sequence. For guide design, sgRNAs that are severely self-complementary, potentially preventing its hybridization with target DNA were eliminated by RNA secondary structure predictions obtained at: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.

sgRNA were in vitro transcribed using the MEGAshortscript T7 Kit (Ambion, Life Technologies) according to the manufacturer's instructions. DNA templates for sgRNA in vitro transcription were generated by using PCR to amplify sgRNA scaffold sequence from plasmid pX330 using 5' primers containing T7 promoter sequence and above designed 20bp target sequence (Table 2.S1, S2). A 80bp fragment of human 18S rRNA transcribed from pTRI-RNA 18S control template supplied with MEGAshortscript kit (Life Technologies) and was used as control RNA.

2.6.4 FLOW CYTOMETRY AND FLUORESCENCE MICROSCOPY

Flow cytometric analysis was performed on a CyAn flow cytometer (Beckman Coulter) and data were collected by Summit v4.3 software (Beckman Coulter). For single-cell cloning, drug selected lines were deposited into a 96-well plate at a density of 1 cell/well using a MoFlow (Dako-Cytomation, Denmark) cell sorter and cultured in 200 µl LDNT supplemented with G418 or blasticidin. Each population from an individual well was considered an individual clone.

Fluorescence microscopy was performed to determine the presence of GFP protein in intracellular and amastigote stage of *T. cruzi* post transfected of sgRNA targeting eGFP,

modifying the protocol described previously (Agrawal, van Dooren et al. 2009). Images were acquired with an Applied Precision Delta Vision microscope, images were deconvolved and adjusted for contrast using its Softworx software (Applied Precision).

2.6.5 ELISA ANALYSIS

To determine the relative abundance of Cas9 protein in epimastigotes, serial dilutions of whole-cell lysate (cells lysed by 4 freeze-thaw cycles) were assayed with anti-FLAG M2 monoclonal antibody (1:1000) and anti-α-tubulin monoclonal antibody (1:500) as a loading standard.

2.6.6 HISTIDINE AMMONIA-LYASE ASSAY.

Epimastigote stage parasites (1 × 106) were disrupted by freeze-thawing and the lysate was incubated in 100mM Tris-HCl (pH 9.0) and 50 mM MgCl2 buffer for 30 minutes at 25°C before 100mM histidine was added. HAL activity was determined by the rate of urocanic acid formation measured by absorbance at 277nm after addition of histidine (Tabor and Mehler 1955).

2.6.7 BODIPY LABELED FATTY ACID UPTAKE ASSAY

The uptake of fatty acids by *T. cruzi* epimastigotes was measured using the QBT™ Fatty Acid Uptake Assay kit (Molecular Devices). Briefly, 1 × 107 *T. cruzi* epimastigotes were pelleted from LDNT media and resuspended in 10ul phosphate buffered saline (PBS) pH 7.0, then 200ul of reconstituted QBT loading solution was added and the parasites immediately analyzed by a CyAn flow cytometer (Beckman Coulter). Flow cytometry data was collected continuously for 200 seconds. To calculate the rate of fatty acid uptake rate using Summit 4.3

(Beckman Coulter), the continuous data was first converted into discreet data by binning fluorescence intensity data into 4 sec intervals and determining the mean fluorescence intensity (MFI) for each bin. Then MFIs were plotted against center point time-stamp of each bin and the slope of the trend line used to calculate the uptake rate.

2.6.8 GALACTOSIDASE TREATMENT AND PNA STAINING

T. cruzi epimastigotes (1 × 107) were washed twice in PBS, then incubated at 37ºC for 1.5hr with 10 U of β-Galactosidase (Grade VIII from Escherichia coli, Sigma-Aldrich) in PBS pH 7.3 containing 50mM Tris-HCl, 10mM 2-mercaptoethanol and 10mM MgCl2. To assess surface galactose residues, parasites were incubated with 10µg/ml PNA Alexa Fluor® 647 conjugate (Life Technologies) in PBS at 37ºC for 10 min and analyzed by a CyAn flow cytometer (Beckman Coulter).

2.6.9 WHOLE GENOME SEQUENCING

Genomic DNA was isolated from *T. cruzi* epimastigotes as previously described (Minning, Weatherly et al. 2011). The DNA samples were enzymatically sheared to 150 bp mean fragment size using the Ion Shear DNA fragmentation kit (Life Technologies), and ligated to barcoded adapters using the Ion Xpress™ Plus Fragment Library Kit (Life Technologies) per the manufacturer's instructions. Adapter-ligated DNA was size-selected using the Pippen Prep (Sage Biosciences), quantified by BioAnalyzer (Agilent) and Ion Library Quantitation Kit (Life Technologies), and the libraries were pooled equally and used to template Ion Sphere Particles using the Ion PI™ Template OT2 200 Kit v2 (Life Technologies), per the manufacturer's instructions. Templated beads were then loaded onto a P1 Ion Proton sequencing chip and sequenced on the Ion Proton (Life Technologies). Sequencing reads were mapped to *T. cruzi*

strain CL Brener reference genome (TritrypDB version 5) using Tmap 4.2 (Life Technologies) with non-default parameters: --score-match 4 --pen-mismatch 5. To assess mutations at potential off-target sites, a custom Perl script was written to identify novel junctions supported by unmapped reads, briefly, all unmapped reads was first BLASTed to reference genome using blastn-short program in suite BLAST 2.2.29+ using non-default parameters "-reward 2 -penalty -3 –e-value 0.001". Reads that have greater than 95% overall sequence coverage (coverage gap is less than 10bp) and best hit for each alignment segment coming from 2 or more different locations in the genome are considered reads that supported a novel junction.

2.7 ACKNOWLEDGMENTS

We thank Dr. Boris Striepen and Dr Michael Terns for productive discussions and valuable suggestions, Gretchen Cooley for technical assistance, and all the members of Tarleton Research Group for helpful suggestions throughout this study. We also thank Julie Nelson of the Center for Tropical and Emerging Global Diseases Flow Cytometry Facility at the University of Georgia and Dr. Muthugapatti Kandasamy of the Biomedical Microscopy Core at the University of Georgia for their assistance. This work was supported by NIH grant R01 AI-089952 to RLT.

2.8 REFERENCES

Agrawal, S., G. G. van Dooren, W. L. Beatty and B. Striepen (2009). "Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins." J Biol Chem **284**(48): 33683-33691.

Auer, T. O., K. Duroure, A. De Cian, J. P. Concordet and F. Del Bene (2014). "Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair." Genome Res **24**(1): 142-153.

Bae, S., J. Kweon, H. S. Kim and J. S. Kim (2014). "Microhomology-based choice of Cas9 nuclease target sites." Nat Methods **11**(7): 705-706.

Bottcher, R., M. Hollmann, K. Merk, V. Nitschko, C. Obermaier, J. Philippou-Massier, I. Wieland, U. Gaul and K. Forstemann (2014). "Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells." Nucleic Acids Res **42**(11): e89.

Burton, P., D. J. McBride, J. M. Wilkes, J. D. Barry and R. McCulloch (2007). "Ku heterodimer-independent end joining in Trypanosoma brucei cell extracts relies upon sequence microhomology." Eukaryot Cell **6**(10): 1773-1781.

Cho, S. W., S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae and J. S. Kim (2014). "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases." Genome Res **24**(1): 132-141.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Crespan, E., T. Czabany, G. Maga and U. Hubscher (2012). "Microhomology-mediated DNA strand annealing and elongation by human DNA polymerases lambda and beta on normal and repetitive DNA sequences." Nucleic Acids Res **40**(12): 5577-5590.

DaRocha, W. D., K. Otsu, S. M. Teixeira and J. E. Donelson (2004). "Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*." Mol Biochem Parasitol **133**(2): 175-186.

De Pablos, L. M. and A. Osuna (2012). "Multigene families in *Trypanosoma cruzi* and their role in infectivity." Infect Immun **80**(7): 2258-2264.

Dickinson, D. J., J. D. Ward, D. J. Reiner and B. Goldstein (2013). "Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination." Nat Methods **10**(10): 1028-1034.

Doench, J. G., E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier and D. E. Root (2014). "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation." Nat Biotechnol.

El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, D. Nilsson, G. Aggarwal, A. N. Tran, E. Ghedin, E. A. Worthey, A. L. Delcher, G. Blandin, S. J. Westenberger, E. Caler, G. C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D. A. Campbell, M. Carrington, J. Crabtree, H. Darban, J. F. da Silveira, P. de Jong, K. Edwards, P. T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A. C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, S. Kluge, H. Koo, D. Lacerda, M. J. Levin, H. Lorenzi, T. Louie, C. R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J. C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J. L. Ramirez, J. Rinta, L. Robertson, S. L. Salzberg, D. O. Sanchez, A.

Seyler, R. Sharma, J. Shetty, A. J. Simpson, E. Sisk, M. T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P. N. Ward, B. Wickstead, J. Wortman, O. White, C. M. Fraser, K. D. Stuart and B. Andersson (2005). "The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease." Science **309**(5733): 409-415.

Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung and J. D. Sander (2013). "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells." Nat Biotechnol **31**(9): 822-826.

Fu, Y., J. D. Sander, D. Reyon, V. M. Cascio and J. K. Joung (2014). "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." Nat Biotechnol **32**(3): 279-284.

Ghorbal, M., M. Gorman, C. R. Macpherson, R. M. Martins, A. Scherf and J. J. Lopez-Rubio (2014). "Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system." Nat Biotechnol **32**(8): 819-821.

Glover, L., J. Jun and D. Horn (2011). "Microhomology-mediated deletion and gene conversion in African trypanosomes." Nucleic Acids Res **39**(4): 1372-1380.

Jao, L. E., S. R. Wente and W. Chen (2013). "Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system." Proc Natl Acad Sci U S A **110**(34): 13904-13909.

Kolev, N. G., C. Tschudi and E. Ullu (2011). "RNA interference in protozoan parasites: achievements and challenges." Eukaryot Cell **10**(9): 1156-1163.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Minning, T. A., D. B. Weatherly, S. Flibotte and R. L. Tarleton (2011). "Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization." BMC Genomics **12**: 139.

Ngo, H., C. Tschudi, K. Gull and E. Ullu (1998). "Double-stranded RNA induces mRNA degradation in Trypanosoma brucei." Proc Natl Acad Sci U S A **95**(25): 14687-14692.

Niu, Y., B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A. P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji and J. Sha (2014). "Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos." Cell **156**(4): 836-843.

Ottaviani, D., M. LeCain and D. Sheer (2014). "The role of microhomology in genomic structural variation." Trends Genet **30**(3): 85-94.

Ran, F. A., P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang and F. Zhang (2013). "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity." Cell **154**(6): 1380-1389.

Ryan, O. W., J. M. Skerker, M. J. Maurer, X. Li, J. C. Tsai, S. Poddar, M. E. Lee, W. DeLoache, J. E. Dueber, A. P. Arkin and J. H. Cate (2014). "Selection of chromosomal DNA libraries using a multiplex CRISPR system." Elife **3**.

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." Nat Biotechnol **32**(4): 347-355.

Schwank, G., B. K. Koo, V. Sasselli, J. F. Dekkers, I. Heo, T. Demircan, N. Sasaki, S. Boymans, E. Cuppen, C. K. van der Ent, E. E. Nieuwenhuis, J. M. Beekman and H. Clevers (2013). "Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients." Cell Stem Cell **13**(6): 653-658.

Tabor, H. and A. Mehler (1955). "[29] Histidase and urocanase." Methods in enzymology **2**: 228-233.

Tamate, S. C., M. Kawata and T. Makino (2014). "Contribution of nonohnologous duplicated genes to high habitat variability in mammals." Mol Biol Evol **31**(7): 1779-1786.

Vazquez, M. P. and M. J. Levin (1999). "Functional analysis of the intergenic regions of TcP2beta gene loci allowed the construction of an improved *Trypanosoma cruzi* expression vector." Gene **239**(2): 217-225.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.

Weston, D., B. Patel and W. C. Van Voorhis (1999). "Virulence in *Trypanosoma cruzi* infection correlates with the expression of a distinct family of sialidase superfamily genes." Mol Biochem Parasitol **98**(1): 105-116.

Xu, D., C. P. Brandan, M. A. Basombrio and R. L. Tarleton (2009). "Evaluation of high efficiency gene knockout strategies for *Trypanosoma cruzi*." BMC Microbiol **9**: 90.

2.9 FIGURES AND TABLES

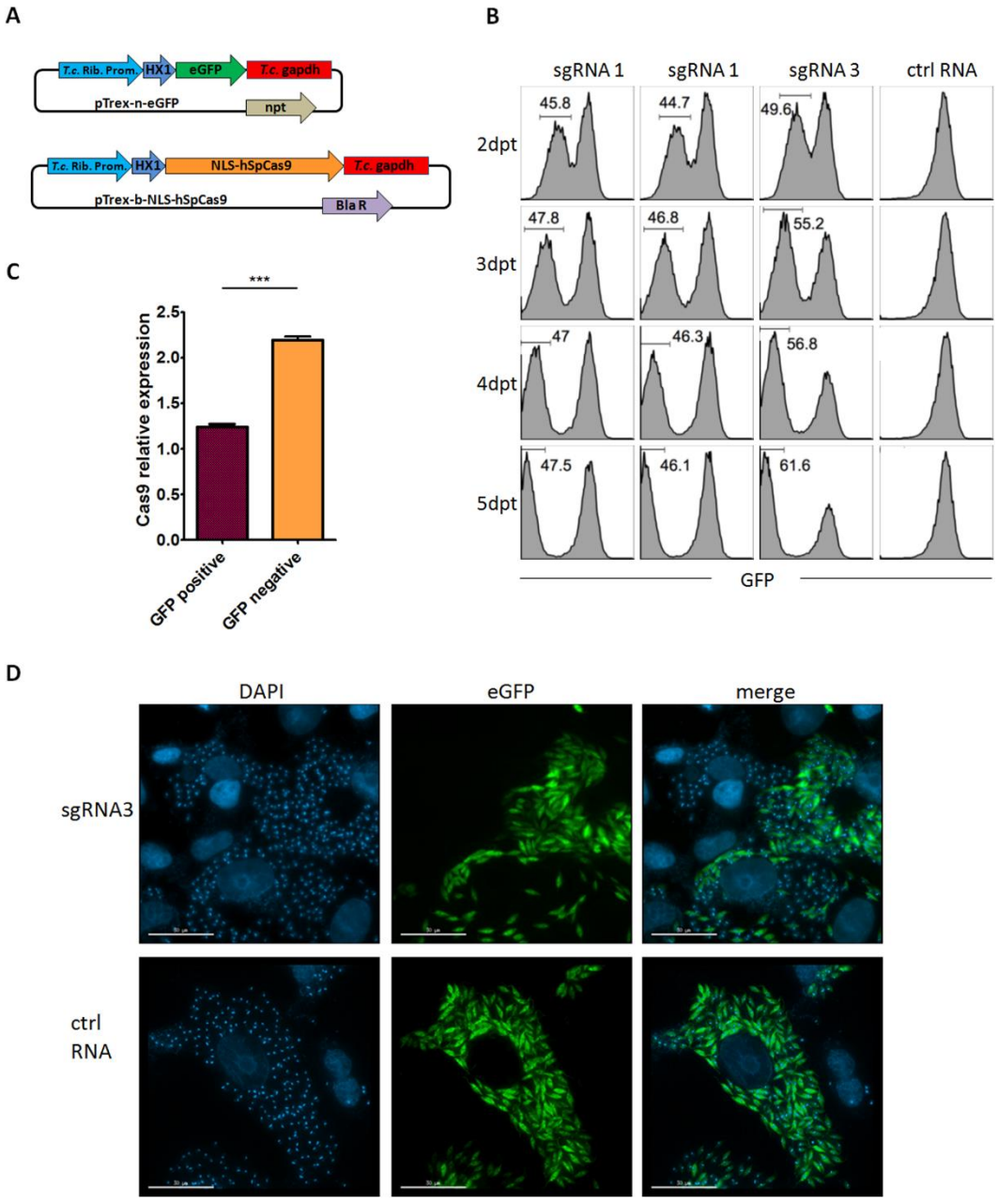


Figure 2.1. Cas9-mediated eGFP disruption in *T. cruzi* epimastigotes and trypomastigotes. (A) Design of constructs for stable expression of eGFP and nuclear localized Cas9 in *T. cruzi* (B) Flow cytometric analysis of *T. cruzi* eGFP- and Cas9-expressing epimastigotes transfected with eGFP sgRNA. Disruption of eGFP is evident

as early as 2 days post-transfection (dpt) and progressive loss of GFP signal is observed over time. (C) ELISA analysis of Cas9 expression in GFP-positive (eGFP-intact parasites (sequence confirmed)) and GFP-negative (Cas9-inudced KO) parasites sorted following eGFP disruption. Expression level of Cas9 was normalized to alpha tubulin expression in corresponding sample. (D) eGFP- and Cas9-expressing *T. cruzi* trypomastigotes transfected with eGFP sgRNA or control RNA, then used to infect vero cells and imaged 5 days later. Comparison of DAPI-stained nuclei and kinetoplasts with GFP florescence demonstrates a mixture of GFP-expressing and non-expressing parasites in the eGFP-targeting sgRNA transfected group and uniform GFP expression in parasites transfected with control RNA

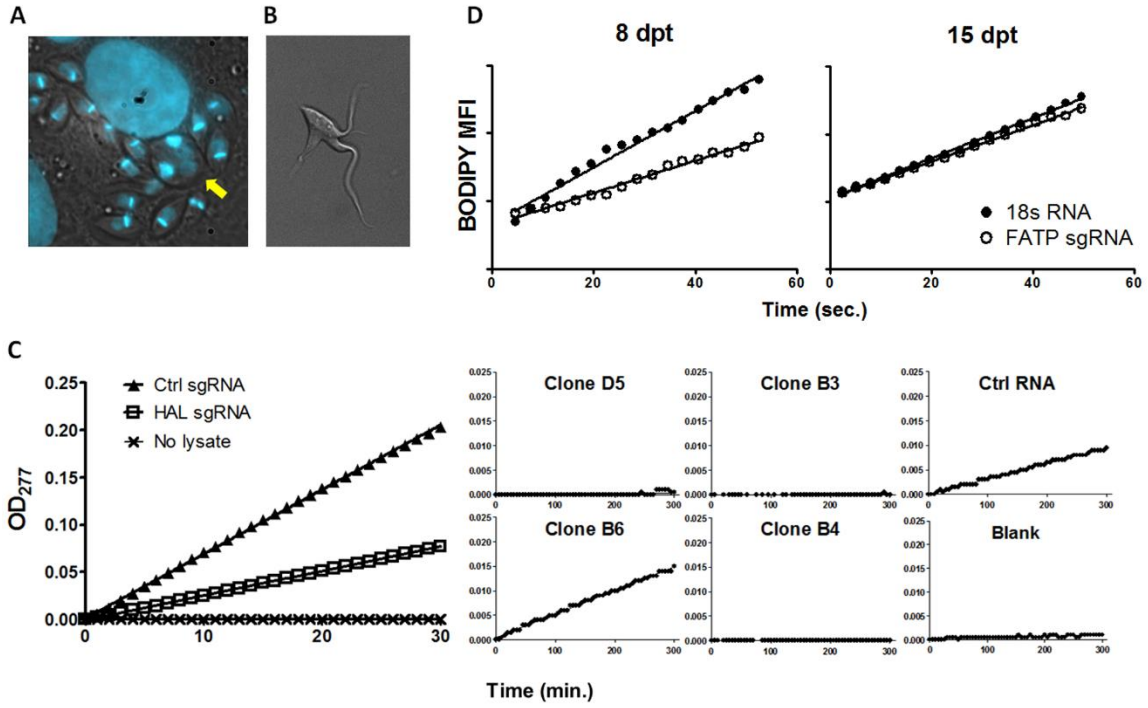


Figure 2.2. Disruption of endogenous *T. cruzi* genes by Cas9-mediated mutation. (A) DAPI stained Cas9-expressing *T. cruzi t*rypomastigotes transfected with α-tubulin-targeting sgRNA, then used to infected vero cells and imaged 4 days later. Note the enlarged intracellular amastigote with multiple nuclei (arrow). (B) Swollen and multi-flagellated Cas9-expressing *T. cruzi* epimastigotes transfected with α-tubulin-targeting sgRNA. (C) HAL activity determined by rate of urocanic acid formation (measured at OD_{277}) in epimastigote lysate 4 days post transfection of HAL gene targeting sgRNA or ctrl RNA. Left, HAL activity of the transfected population; right, HAL activity of individual clonal lines. Epimastigotes transfected with the HAL sgRNA exhibited a 60% decrease in HAL activity at day 4 post-transfection compared to parasites transfected with the control 18s RNA. Three of the four clones tested exhibited no HAL activity. (D) BODIPY-labeled fatty acid analog uptake by epimastigotes at day 8 and 15 post-transfection with FATP gene-targeting sgRNA or ctrl RNA. sgRNA-guided Cas9 targeting of FATP resulted in a 37% decrease in the fatty acid uptake rate at day 8, but at 2 weeks post-transfection, near normal FA uptake was observed, suggesting the loss of the FATP-mutant population and survival of only the non-mutated WT population.

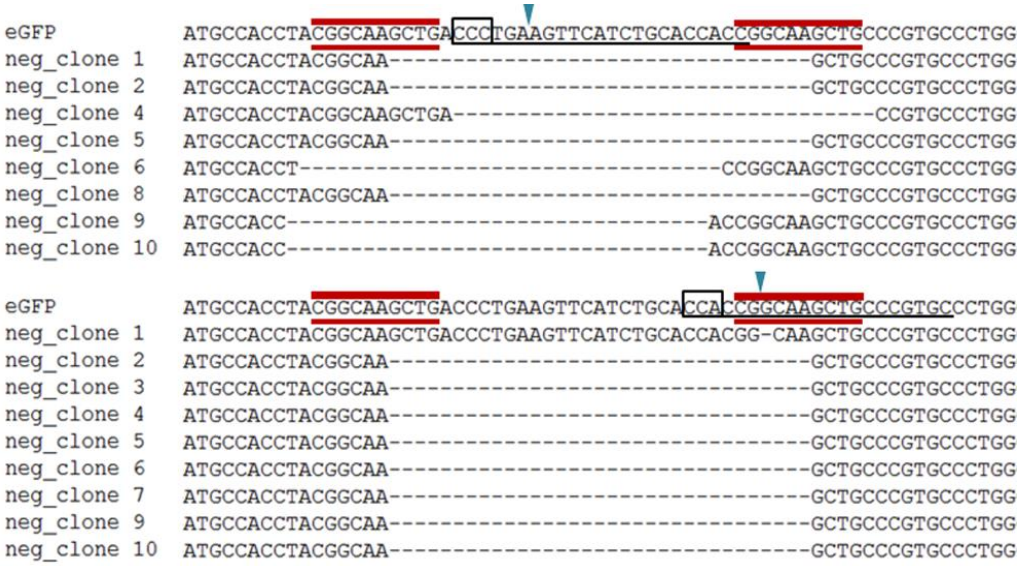


Figure 2.3. Sequencing of eGFP gene in GFP-disrupted clones sorted from parasite populations transfected with eGFP-targeting sgRNA 3 (top) and sgRNA 1 (bottom). underlined is the sgRNA targeting sequence, boxed is the PAM motif (on reverse strand), red mark-up shows microhomology sequences flanking the sgRNA targeting sequence, double-stranded cut sites are indicated with blue arrow. The sequence of negative clone 1 in the sgRNA-1 (bottom) set shows a single base deletion in this region, likely due to a spontaneous mutation in GFP. All other sequences showed a 33bp deletion, though deletion junctions occur at slightly different positions, they all fall between a pair of homologous sequences of 10 bp flanking the cut site, as predicted by repair via a microhomology-mediated ending joining (MMEJ) pathway.

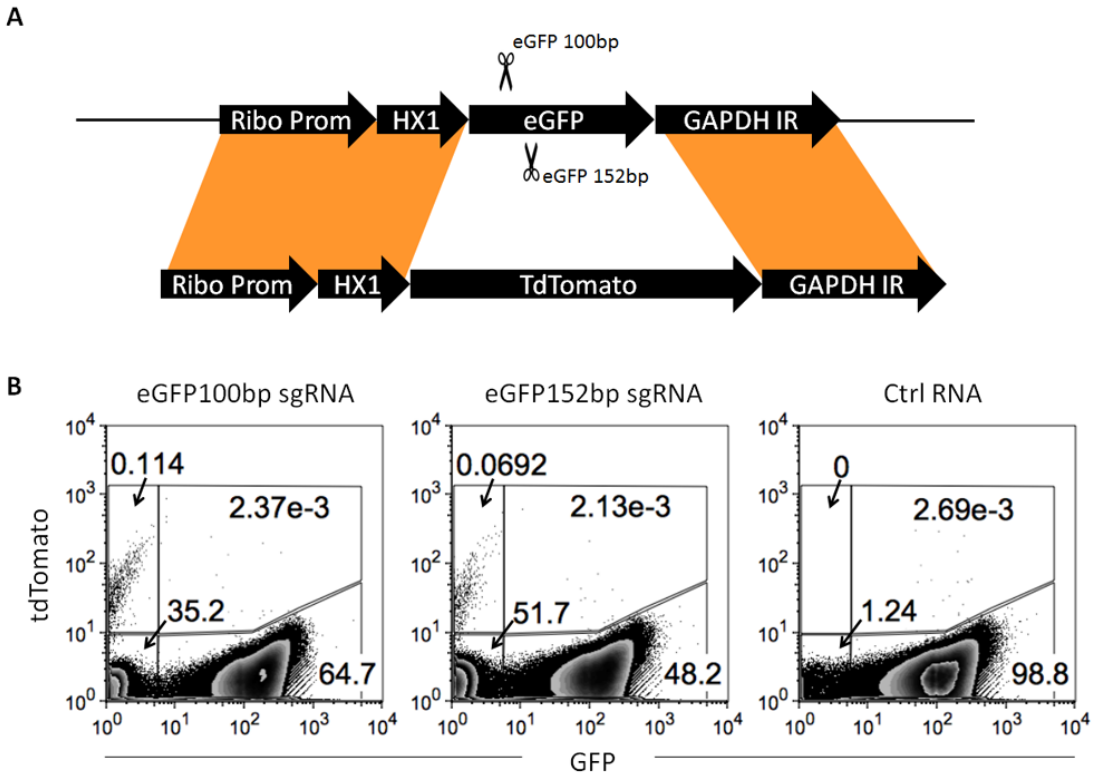


Figure 2.4. Homologous recombination-mediated replacement of the eGFP gene with a tdTomato gene facilitated by Cas9-induced DSB in eGFP. (A) Schematic diagram of fluorescence maker swap assay design.*T. cruzi* epimastigotes harboring an eGFP expression cassette were co-transfected with both a sgRNA targeting eGFP at position 100bp or 152bp and a tdTomato expression cassette with 5' and 3' homology to the eGFP insert. (B) Flow cytometric analysis of tdTomato-eGFP fluorescence maker swap assay performed in Cas9 expressing epimastigotes 5 days after transfection with the indicated eGFP-targeting sgRNAs. Numbers indicate percentage of cells in each quadrant. The sgRNA targeting eGFP at position 100bp and 152bp respectively yielded a 0.11% and 0.069% rate of fluorescence marker swap, indicated by the loss of GFP and gain of tdTomato. In the absence of the eGFP-targeted sgRNA (but in the presence of the tdTomato template), homologous recombination was below the level of detection by this assay.

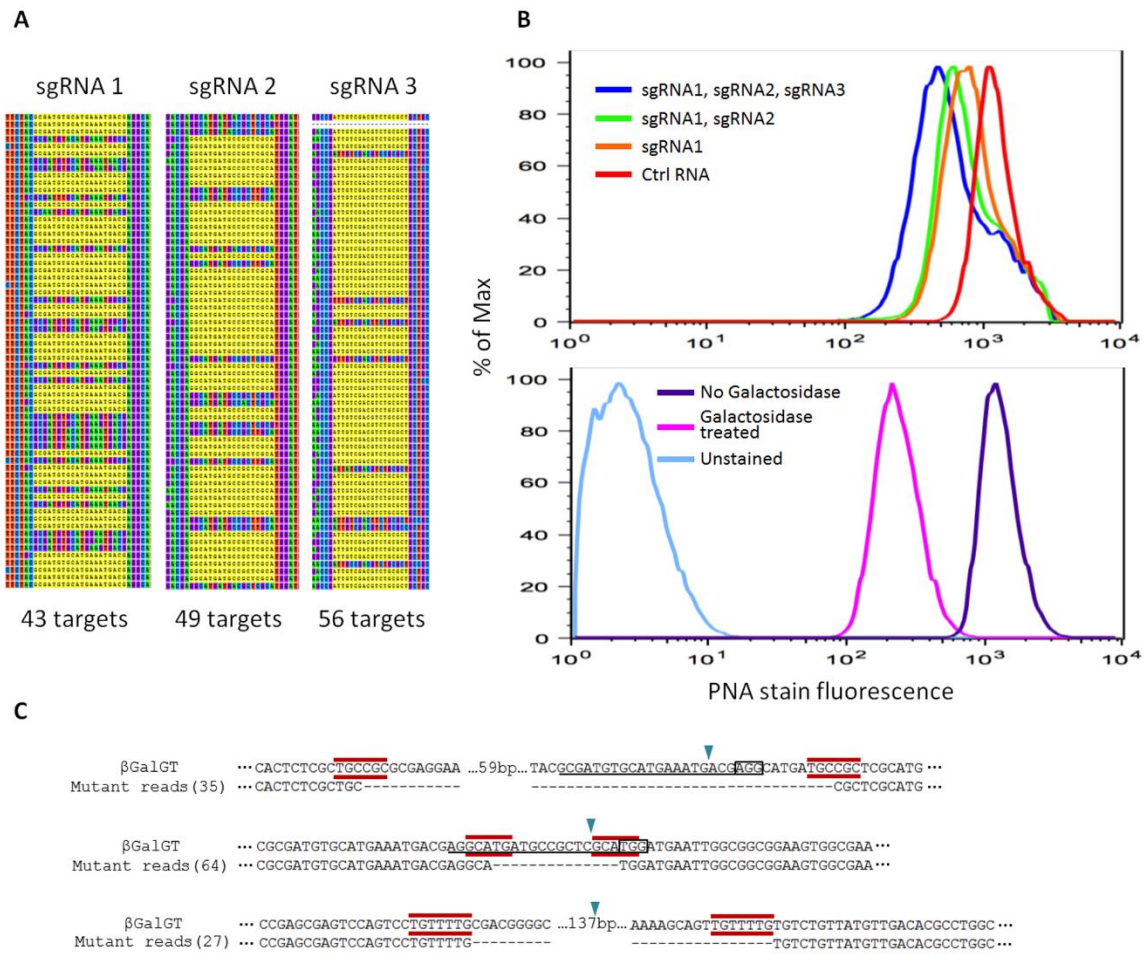


Figure 2.5. Knockdown of β galactofuranosyl glycosyltransferase (β-GalGT) activity by Cas9-mediated gene family mutation. (A) sgRNA targeting strategies for β-GalGTgene family. DNA alignment at sgRNA targeting regions are shown, highlighted in yellow are sgRNA targeting sequences. sgRNA 1,2 and 3 targets 43,49 and 56 out of 65 genes in the gene family, each gene is targeted by at least one sgRNA. Complete gene identifiers, genome location and annotations are presented in Table S3. (B) Flow cytometric histograms of PNA-stained *T. cruzi* epimastigotes transfected sequentially with one, two or three sgRNA targeting the β-GalGT gene family (top) or treated with β-galactosidase (bottom). PNA binds specifically to β-galactosyl residues and indicates a progressive reduction in surface β-galactosyl residues with each additional sgRNA transfection. Biochemical removal of surface β-galactosyl residues using galactosidase is used as a control for the PNA staining. (C) Genome sequencing reads support novel junctions in the 3 target regions, from top: target site 1, target site 2 and target site 3. Boxed is the PAM motif (on reverse strand), red mark-up shows microhomology sequences flanking the sgRNA targeting sequence and double-stranded cut sits are indicated with blue arrow.

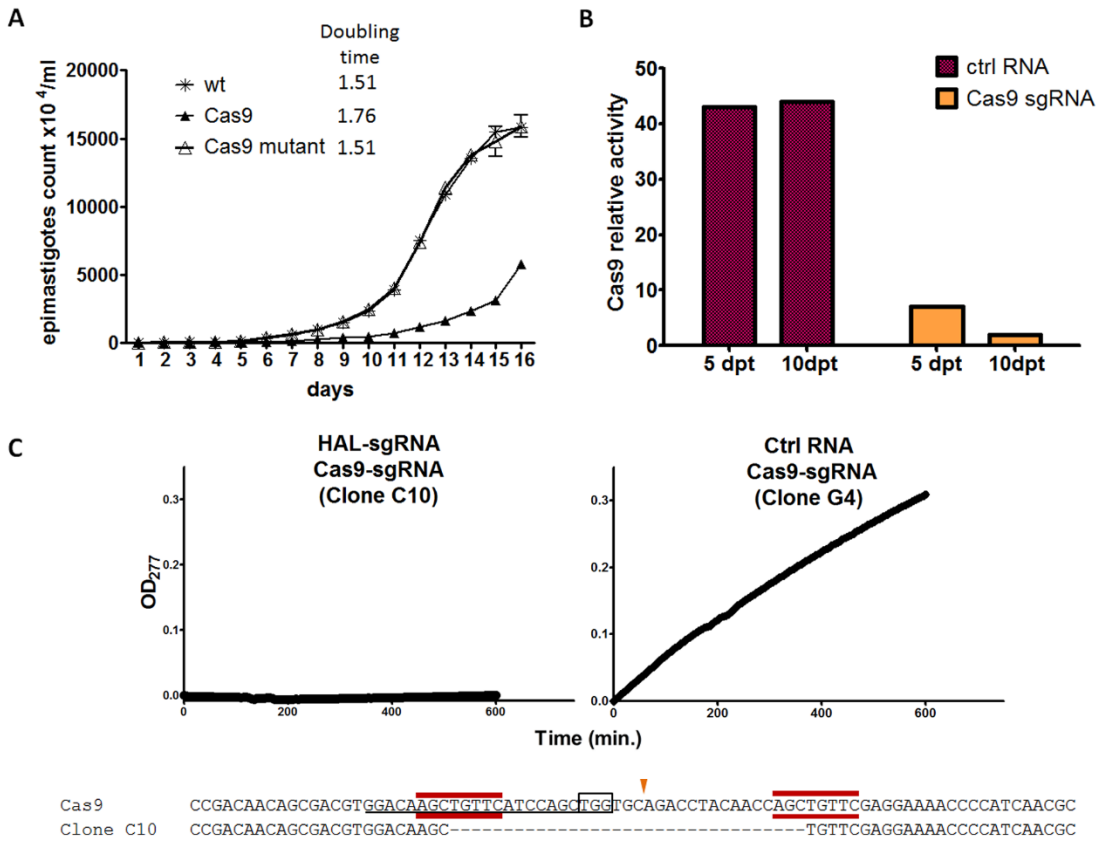


Figure 2.6. Cas9 expression negatively impacts growth of *T. cruzi* epimastigotes but could mediate disruption of its own coding sequence. (A) Growth curve of eGFP-epimastigotes that do not express Cas9, expresses Cas9, or have Cas9 knocked-out by transfecting Cas9-targeting sgRNA. Parasites with stable Cas9 expression have a significant greater doubling time compared to WT or Cas9-disrupted parasites. (B) 83% and 95% of Cas9 activity was abolished on day 5 and day 10 post-transfection of Cas9-targeting sgRNA, as determined by eGFP disruption assay, demonstrating that Cas9 could mediate disruption of its own coding sequence. (C) HAL activity determined by rate of urocanic acid formation (measured at OD_{277}) in lysates of clonal epimastigotes lines from parasites cotransfected with Cas9 gene targeting sgRNA and HAL gene targeting sgRNA or ctrl RNA. Sequencing of Cas9 gene in clone C10, which lacks HAL activity confirmed the disruption of Cas9, demonstrating knockout of both Cas9 gene and HAL gene in parallel.

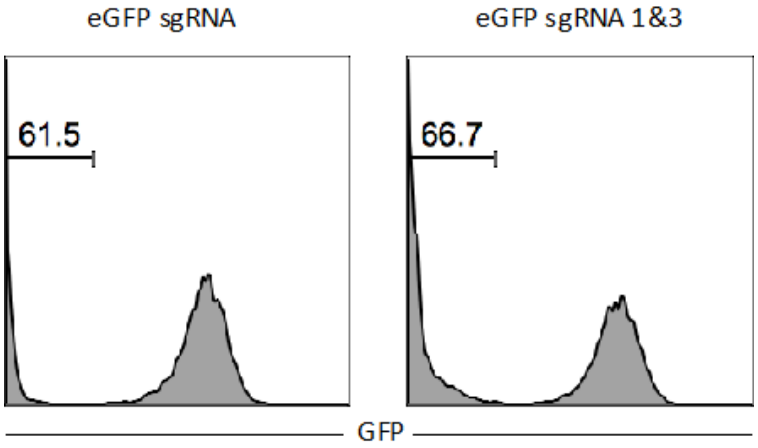


Figure 2.S1. Simultaneously transfection of two eGFP-targeting sgRNAs resulted in a modest increase of eGFP disruption rate compared to transfection of single eGFP-targeting sgRNA. Equal amounts of total sgRNA was used for each transfection.

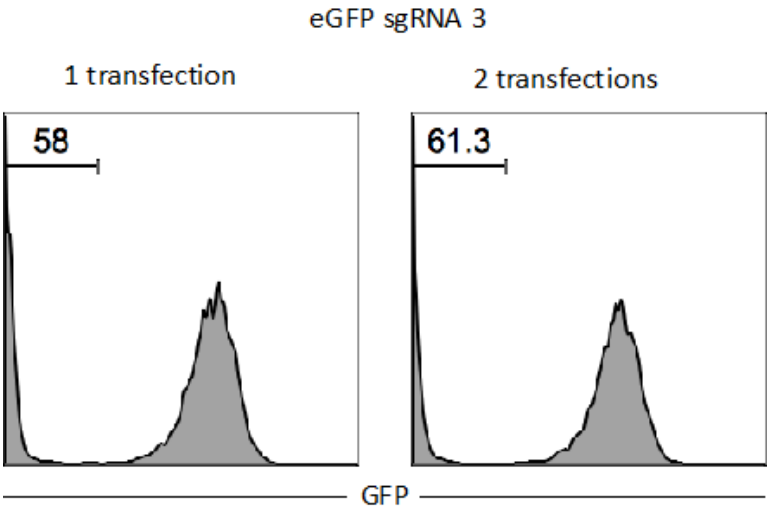


Figure 2.S2. Sequential transfection of the same sgRNA resulted in a marginal increase in the frequency of eGFP disruption.

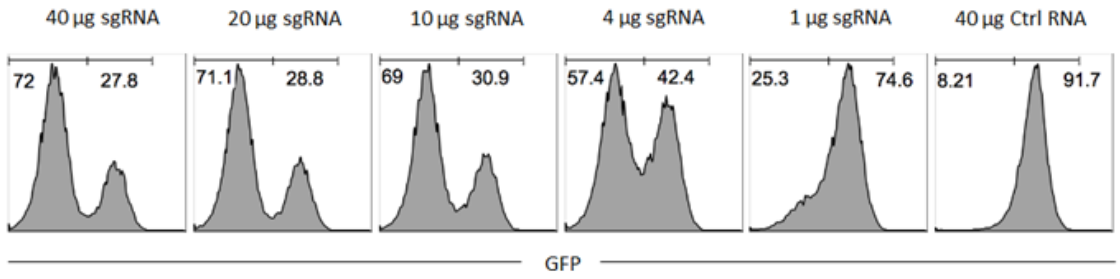


Figure 2.S3. Impact of quantity of transfected sgRNA on eGFP disruption rate. Bulk of eGFP disruption rate was achieved with 10µg of sgRNA with only marginal increases of eGFP disruption achieved by doubling and quadrupling sgRNA concentration.

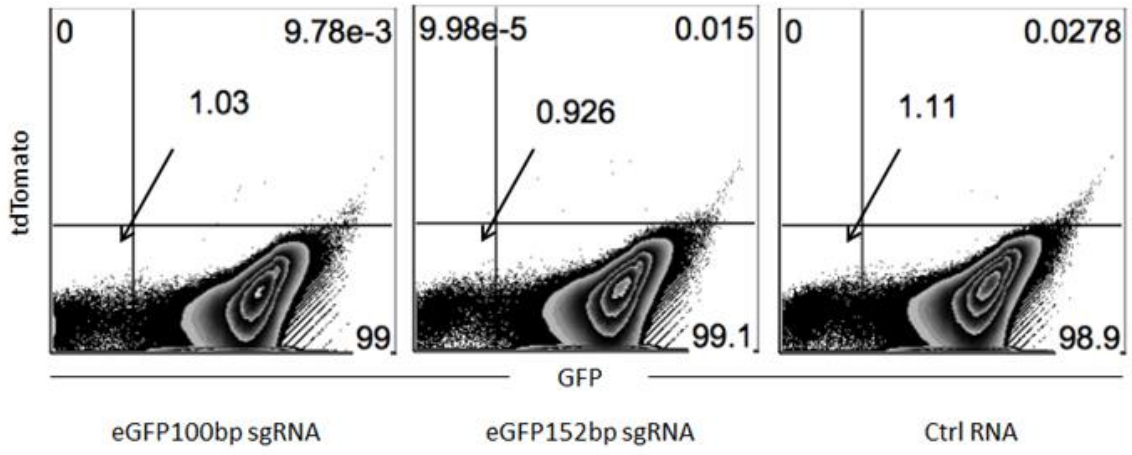


Figure 2.S4. Flow cytometric analysis of tdTomato-eGFP fluorescence maker swap assay performed in Cas9-nickase expressing epimastigotes. eGFP was targeted by a single sgRNA at position 100bp or 152bp. Numbers indicate percentage of cells in each quadrant. All groups were co-transfected with tdTomato expression cassette template.

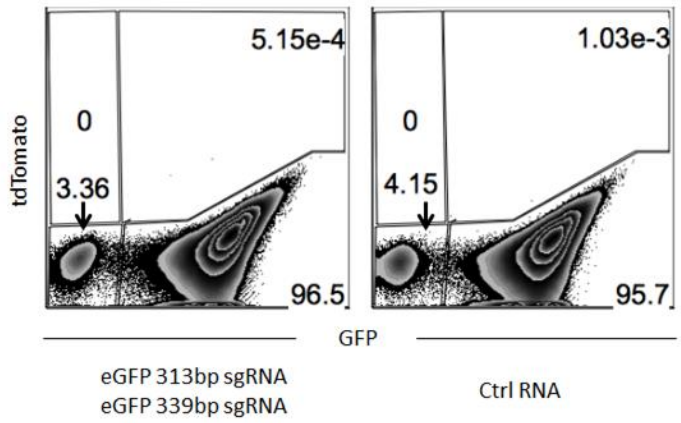


Figure 2.S5. Flow cytometric analysis of tdTomato-eGFP fluorescence maker swap assay performed in Cas9-nickase expressing epimastigotes. eGFP was targeted in a "double-nick" fashion by a pair of closely positioned sgRNA on opposite strands, likely result in a double-stranded breakage due to closely positioned nick on both strands. Numbers indicate percentage of cells in each quadrant.

Table 2.S1. Primers used in this study

Primers	Sequence (5' to 3')	Procedure
XbaI-eGFP-F	AAAAATCTAGAatGGTGAGCAAGGGCGAGGA	To clone eGFP into pTrex-n
XhoI-eGFP-R	AAAAACTCGAGTTACTTGTACAGCTCGTCCATGCCG	
XbaI-hSpCas9-F	GGGGTCTAGAATGGACTATAAGGACCACGACG	To clone NLS-hSpCas9 into pTrex-b
XhoI-hSpCas9-R	GGGGCTCGAGTTACTTTTTCTTTTTTGCCTGGCCGGCCTTTTTCGTG	
px330a1680F	TTCCTGGTGGAAGAGGATAAG	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a1844R	GGCCAGGGCCAGATAGAT	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a2359F	ACCTGACCCTGCTGAAAGC	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a2956F	TGACCGAGGGAATGAGAAAG	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a3560F	CGCCATTAAGAAGGGCATC	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a4026F	CTGAACGCCAAGCTGATTAC	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
px330a4576F	GGGAGATCGTGTGGGATAAG	To sequence NLS-hSpCas9 in pTrex-b-NLS-hSpCas9 construct
T7- eGFP sgRNA1-scaffold-F	GGAGGCCGGAGAATTGTAATACGACTCACTATAGGGAGAGCGAGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- eGFP sgRNA2-scaffold-F	GGAGGCCGGAGAATTGTAATACGACTCACTATAGGGAGAGCACGGGCAGCTTGCCGGGTTTTAGAGCTAGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- eGFP sgRNA3-scaffold-F	GGAGGCCGGAGAATTGTAATACGACTCACTATAGGGAGAGGTGGTGCAGATGAACTTCAGTTTTAGAGCTAGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template

T7- HAL sgRNA1-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGACGCCTCG TCCGAAATTTCGAGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- HAL sgRNA2-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGCCCGAT GTACTGTACGCTTGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- HAL sgRNA3-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGGGCGCC ACCATCGAAATTTGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- FATP sgRNA-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGATTGTGCC GTGCTGCAATTGCGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- α-tubulin sgRNA-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGTGCTGG GAGCTGTTCTGTCGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- β-GalGT sgRNA1-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGCGATGT GCATGAAATGACGGTTTTAGAGC TAGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- β-GalGT sgRNA2-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGAGGCAT GATGCCGCTCGCAGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- β-GalGT sgRNA3-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGCAGCGC AGACGTCGACAATGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- Cas9 sgRNA-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGGACAAG CTGTTCATCCAGCGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- eGFP-339F sgRNA-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGCCGTCG TCCTTGAAGAAGAGTTTTAGAGCT AGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template
T7- eGFP-313R sgRNA-scaffold-F	GGAGGCCGGAGAATTGTAATACG ACTCACTATAGGGAGAGAAGTTC GAGGGCGACACCCGTTTTAGAGC TAGAAATAGCAAG	Paired with primer sgRNA-scaffold-R to produce PCR product used for sgRNA IVT template

sgRNA-scaffold-R	AAAAAAGCACCGACTCGGTGCCACTT	Reverse primer to produce PCR product used for sgRNA IVT template
eGFP-4F	GTGAGCAAGGGCGAGGAG	To sequence the sgRNA targeting region of eGFP gene in eGFP disrupted mutants
eGFP-325R	TCTTGTAGTTGCCGTCGTCC	To sequence the sgRNA targeting region of eGFP gene in eGFP disrupted mutants

Table 2.S2. List of sgRNA targeting sequences used in this study

sgRNA name	targeted gene	sgRNA targeting sequence (PAM included)
eGFP sgRNA1	eGFP	GCGAGGGCGATGCCACCTACGG
eGFP sgRNA2	eGFP	GCACGGGCAGCTTGCCGGTGG
eGFP sgRNA3	eGFP	GGTGGTGCAGATGAACTTCAGGG
HAL sgRNA1	HAL (Tc00.1047053506247.220)	CGCCTCGTCCGAAATTTCGATGG
HAL sgRNA2	HAL (Tc00.1047053506247.220)	GCCCGATGTACTGTACGCTTTGG
HAL sgRNA3	HAL (Tc00.1047053506247.220)	GGGCGCCACCATCGAAATTTCGG
FATP sgRNA	FATP (TcCLB.511907.110 and TcCLB.506799.10)	TTGTGCCGTGCTGCAATTGCTGG
α-tubulin sgRNA	α-tubulin	GTGCTGGGAGCTGTTCTGTCTGG
β-GalGT sgRNA1	β-GalGT	GCGATGTGCATGAAATGACGAGG
β-GalGT sgRNA2	β-GalGT	GAGGCATGATGCCGCTCGCATGG
β-GalGT sgRNA3	β-GalGT	GCAGCGCAGACGTCGACAATCGG
Cas9 sgRNA	Cas9	GGACAAGCTGTTCATCCAGCTGG
eGFP 313R sgRNA	eGFP	GCCGTCGTCCTTGAAGAAGATGG
eGFP 339F sgRNA	eGFP	GAAGTTCGAGGGCGACACCCTGG

Table 2.S3. List of annotated β-GalGT genes being targeted in this study

Gene identifier	Genome location	Gene annotation (Tritrypdb.org)
TcCLB.420293.40	TcChr8-S from 386,470 to 387,609	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.463279.9	Tcruzi_23428 from 3 to 437	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase
TcCLB.503447.60	TcChr31-S from 18,705 to 19,541	beta galactofuranosyl glycosyltransferase
TcCLB.503739.10	Tcruzi_4728 from 1,295 to 2,587	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.503861.50	TcChr12-P from 514,422 to 515,447	beta galactofuranosyl glycosyltransferase
TcCLB.504115.30	TcChr25-P from 712,908 to 714,236	beta galactofuranosyl glycosyltransferase
TcCLB.504189.10	TcChr38-P from 629,613 to 630,644	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.504733.10	Tcruzi_6377 from 2,118 to 3,407	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.504801.10	Tcruzi_6402 from 372 to 1,667	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.505207.20	TcChr41-P from 7,179 to 8,051	beta galactofuranosyl glycosyltransferase
TcCLB.505267.30	TcChr31-P from 21,510 to 22,535	beta galactofuranosyl glycosyltransferase
TcCLB.505365.50	TcChr24-S from 758,242 to 759,579	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.505915.40	Tcruzi_6735 from 7,097 to 8,080	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.505975.30	Tcruzi_6760 from 9,450 to 10,772	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506129.60	TcChr6-S from 357,490 to 358,515	beta galactofuranosyl glycosyltransferase
TcCLB.506183.70	TcChr26-S from 756,494 to 757,786	beta galactofuranosyl glycosyltransferase
TcCLB.506217.20	Tcruzi_6856 from 2,508 to 3,806	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506245.150	Tcruzi_6868 from 44,061 to 44,639	beta galactofuranosyl glycosyltransferase
TcCLB.506331.80	TcChr40-P from 13,993 to 15,477	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506341.30	TcChr20-S from 9,113 to 10,441	beta galactofuranosyl glycosyltransferase

TcCLB.506345.100	TcChr22-S from 675,411 to 676,820	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506393.30	TcChr14-P from 563,083 to 564,108	beta galactofuranosyl glycosyltransferase
TcCLB.506471.40	TcChr33-P from 1,011,264 to 1,012,274	beta galactofuranosyl glycosyltransferase
TcCLB.506507.30	TcChr20-P from 9,139 to 10,164	beta galactofuranosyl glycosyltransferase
TcCLB.506595.60	TcChr33-P from 23,128 to 24,153	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506683.140	TcChr12-S from 33,132 to 34,337	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506813.130	Tcruzi_5397 from 27,746 to 28,729	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.506843.10	Tcruzi_7113 from 3,096 to 4,178	beta galactofuranosyl glycosyltransferase
TcCLB.507121.40	TcChr7-S from 16,753 to 18,414	beta galactofuranosyl glycosyltransferase
TcCLB.507167.200	TcChr28-P from 798,196 to 799,221	beta galactofuranosyl glycosyltransferase
TcCLB.507591.20	TcChr31-P from 499,826 to 501,154	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.507643.90	TcChr22-P from 678,716 to 680,461	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.507875.60	TcChr34-S from 48,892 to 50,094	beta galactofuranosyl glycosyltransferase
TcCLB.508071.30	Tcruzi_7589 from 5,527 to 6,765	beta galactofuranosyl glycosyltransferase
TcCLB.508325.220	TcChr13-S from 527,258 to 528,607	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.508521.30	TcChr35-P from 36,408 to 37,655	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.508581.30	Tcruzi_7788 from 6,895 to 7,920	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.508607.60	TcChr35-S from 1,161,753 to 1,162,823	beta galactofuranosyl glycosyltransferase
TcCLB.508677.50	TcChr22-P from 219,167 to 220,192	beta galactofuranosyl glycosyltransferase
TcCLB.508831.180	TcChr30-S from 756,296 to 757,321	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.509033.20	TcChr32-P from 96,606 to 97,865	beta galactofuranosyl glycosyltransferase

TcCLB.509223.60	TcChr28-P from 736,117 to 737,142	beta galactofuranosyl glycosyltransferase
TcCLB.509295.50	TcChr28-P from 756,600 to 757,625	beta galactofuranosyl glycosyltransferase
TcCLB.509581.20	TcChr14-S from 574,410 to 575,534	beta galactofuranosyl glycosyltransferase
TcCLB.509777.40	TcChr21-S from 365,342 to 366,532	beta galactofuranosyl glycosyltransferase
TcCLB.509815.50	TcChr16-S from 592,030 to 593,112	beta galactofuranosyl glycosyltransferase
TcCLB.509875.50	TcChr26-P from 769,970 to 771,292	beta galactofuranosyl glycosyltransferase
TcCLB.509931.10	TcChr8-P from 11,241 to 12,017	beta galactofuranosyl glycosyltransferase
TcCLB.509959.40	TcChr31-S from 922,113 to 923,195	beta galactofuranosyl glycosyltransferase
TcCLB.510233.30	Tcruzi_8339 from 3,353 to 3,997	beta galactofuranosyl glycosyltransferase
TcCLB.510483.340	TcChr38-P from 1,129,573 to 1,130,553	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.510497.160	Tcruzi_8411 from 32,582 to 33,127	beta galactofuranosyl glycosyltransferase
TcCLB.510643.110	TcChr16-P from 622,099 to 623,124	beta galactofuranosyl glycosyltransferase
TcCLB.510713.50	TcChr38-P from 978,140 to 979,120	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.510843.10	Tcruzi_8519 from 905 to 1,822	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.510847.20	TcChr19-S from 40,661 to 41,686	beta galactofuranosyl glycosyltransferase
TcCLB.510853.50	TcChr19-S from 97,180 to 97,860	beta galactofuranosyl glycosyltransferase
TcCLB.511077.20	Tcruzi_8591 from 3,506 to 4,486	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.511671.30	TcChr30-P from 760,724 to 761,959	beta galactofuranosyl glycosyltransferase
TcCLB.511771.30	TcChr17-P from 77,236 to 78,324	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.511861.60	TcChr15-P from 128,833 to 129,858	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.507479.51	TcChr21-S from 58,023 to 59,048	beta galactofuranosyl glycosyltransferase

TcCLB.507643.15	TcChr22-P from 709,505 to 710,677	beta galactofuranosyl glycosyltransferase
TcCLB.506723.41	TcChr35-P from 7,895 to 9,187	beta galactofuranosyl glycosyltransferase (pseudogene)
TcCLB.508521.112	TcChr35-P from 54,974 to 55,858	beta galactofuranosyl glycosyltransferase (pseudogene), putative

CHAPTER 3

EUPAGDT: A WEB TOOL TAILORED TO DESIGN CRISPR GUIDE RNAS FOR EUKARYOTIC PATHOGENS

Duo Peng and Rick Tarleton. 2015. Microbial Genomics. Volume 1 issue 4, doi:10.1099 mgen.0.000033. Reprinted here with permission of the publisher

3.1 ABSTRACT

Recent development of CRISPR-Cas9 genome editing has enabled highly efficient and versatile manipulation of a variety of organisms and adaptation of CRISPR-Cas9 system to eukaryotic pathogens has opened new avenues for studying these otherwise hard to manipulate organisms. Here we describe a webtool, Eukaryotic Pathogen gRNA Design Tool (EuPaGDT; available at http://grna.ctegd.uga.edu), which identifies gRNA in input gene(s) to guide users in arriving at well-informed and appropriate guide RNA (gRNA) design for many eukaryotic pathogens. Flexibility in gRNA design, accommodating unique eukaryotic pathogen (gene and genome) attributes and high throughput gRNA design are the main features that distinguish EuPaGDT from other gRNA design tools. In addition to employing an array of known principles to score and rank gRNAs, EuPaGDT implements an effective on-target search algorithm to identify gRNA targeting multi-gene families, which are highly-represented in these pathogens and play important roles in host:pathogen interactions. EuPaGDT also identifies and scores microhomology sequences flanking each gRNA targeted cut-site; these sites are often essential for the microhomology-mediated end joining (MMEJ) process used for double stranded break repair in these organisms. EuPaGDT also assists users in designing single-stranded oligonucleotide for homology directed repair (HDR). In batch processing mode, EuPaGDT could process genome scale sequences, enabling preparation of gRNA library for large-scale screening projects.

3.2 IMPACT STATEMENT

Like other communities, the eukaryotic pathogen research community has begun to adopt the CRISPR-Cas9 system for engineering genomes of eukaryotic parasites at an unprecedented scale and ease. Our EuPaGDT webtool enables researchers to arrive at well-informed gRNA and homology-directed repair (HDR) template design for CRISPR-Cas9 experiments in eukaryotic pathogens. Using an array of known principles, EuPaGDT characterizes all potential gRNAs in a user-input gene, allowing users to quickly identify refined gRNA in a gene sequence and greatly increases the chances of successful CRISPR-Cas9 experiments. EuPaGDT also features a batch processing mode, which can identify gRNA at a whole genome scale allowing the in-silico preparation of gRNA libraries for large-scale screening projects. EuPaGDT is currently available for 25 eukaryotic pathogen genomes, covering major lineages of eukaryotic pathogens and popular model organisms. Users can also upload custom genomes or request default genomes to be added to EuPaGDT. At this time, the EuPaGDT server has been running for 5 months, and has processed over 1340 requests from 654 users originating from 30 countries.

3.3 INTRODUCTION

RNA-guided Cas9 nuclease has enabled rapid, targeted modification of a wide range of genomes, including those of eukaryotic pathogens (Shen, Brown et al. 2014, Sidik, Hackett et al. 2014, Zheng, Jia et al. 2014, Peng, Kurup et al. 2015, Sollelis, Ghorbal et al. 2015), which are the causative agents of some of the most devastating and intractable diseases of humans. The CRISPR-Cas9 system is likely to be a particularly important tool for the study of gene function in pathogens that lack functional RNAi pathways, such as Plasmodium sp. (Ghorbal, Gorman et al. 2014, Lee and Fidock 2014, Wagner, Platt et al. 2014, Zhang, Xiao et al. 2014), the causative agent of malaria, and *Trypanosoma cruzi*, the agent of Chagas disease (Peng, Kurup et al. 2015).

The CRISPR-Cas9 system has proven especially useful because of its relative ease and high efficiency as well as the ability to achieve multiple modifications in a single organism/cell. This latter property is particularly useful for modifying members of multigene families that are common in these pathogens (Peng, Kurup et al. 2015).

Current gRNA design tools for CRISPR-Cas9 system have limitations when applied to gRNA design for eukaryotic pathogens. For example, most parasite genomes exhibit great nucleotide sequence divergence (even at within-specie level) and harbor large, rapidly-evolving gene families that are important players of host-pathogen interaction. In order to harness CRISPR-Cas9's multiplexing power to edit gene families, a gRNA design tool must handle multiple "on-target" hits (gene family members) and discriminate them from true off target sequences. To address these problems, we have developed a web tool, EuPaGDT, tailored to designing gRNA for eukaryotic pathogens.

3.4 EUKARYOTIC PATHOGEN GRNA DESIGN TOOL (EUPAGDT)

EuPaGDT identifies all possible gRNAs in input sequence, and then calculates a ranked list of those gRNAs based on (1) on-target and off-target hit(s) in the selected or uploaded pathogen genome, (2) predicted gRNA activity, and (3) identified microhomology pairs flanking the gRNA targeted cut site. (Fig. 3.1) illustrates a workflow for a non-batch job request.

EuPaGDT can identify on-target hits in the genome for each gRNA, including those intended for editing multi-gene families. The on-target hit feature has the advantage of also identifying additional, sometimes unannotated copies of target genes. This feature is particularly important in incomplete and poorly annotated pathogen genomes, and/or in incompletely annotated gene family sets that can number in the thousands. To find all on-target genome-hits,

the program compares the homology between sequences flanking the identified gRNA with all genome-target flanking sequences. If flanking homology regions meet the programmed identity criteria, then the genome-target will qualify as a potential on-target hit. Key parameters governing flanking homology comparison, such as sequence length and threshold for alignment identity and coverage, can be adjusted by the user to enable effective searching of on-targets in a spectrum of highly conserved to more divergent gene families. Identified similar sequences that do not meet these on-target criteria are automatically assigned to the off-target list. All genomic hits are annotated to aid user in determining whether off-targets are within coding regions and to verify genomic loci annotations for each on-target site (Fig. 3.2a and b).

A target score is assigned to each identified gRNA. The target score reflects how well a gRNA can target on-target sites while avoiding off-targets and is calculated as follows:

$$target\ score = \frac{on\ target\ index}{Maximum\ on\ target\ index} - \frac{off\ target\ index}{on\ target\ index}$$

$$in\ which: \quad on\ target\ index$$
$$= number\ of\ perfectly\ matching\ on\ target\ hits$$
$$+ 0.5 \times (number\ of\ imperfectly\ matching\ on\ target\ hits)$$

$$off\ target\ index$$
$$= number\ of\ perfectly\ matching\ off\ target\ hits$$
$$+ 0.5 \times (number\ of\ imperfectly\ matching\ off\ target\ hits)$$

The first term of the target score equation is a fraction that reflects how well the current gRNA hits on-targets compared to theoretical maximum target index. Maximum on-target index represents the maximum on-target number that gRNAs in a given input sequence can have, for

example, a single-locus 2-allele gene will have 2 as the "maximum on target index", and a gene with 5 copies in the genome would have 5 as the "maximum on target index". The first term of the equation has a maximum value of 1. The second term of the formula evaluates how well the current gRNA avoids hitting off-target sites. Briefly, if a gRNA has far more on-targets than off-targets, the second term would be a small fraction, which is deducted from first term; in contrast, the second term will be >1 if a gRNA has more off-targets than on-targets, rendering a negative target score. The target score equation as a whole works equally well for single locus genes and large gene families.

To further assess the potential utility of each gRNA identified, we have adopted a gRNA activity prediction scoring matrix empirically determined by Doench et al. (Doench, Hartenian et al. 2014). The scoring matrix scores gRNA based on the nucleotide composition of gRNA's 20nt targeting sequence as well as 4nt, 3nt up- and downstream respectively. Though the gRNA activity prediction scoring matrix is developed from data obtain in mammalian cells, we've found in *T. cruzi* that higher activity score (>0.5) correlates strongly with successful gRNA function (n>10).

EuPaGDT is extremely flexible, allowing users to tune many parameters governing the design of gRNA and characterization of on-/off-targets. Recent engineering of CRISPR-Cas9 nucleases has identified several smaller Cas9 orthologues as well as ones with altered PAM preferences (Kleinstiver, Prew et al. 2015). The expanding list of available PAMs greatly increases the number of potential gRNA in a given nucleotide sequence, empowering researchers to edit genomes with higher precision. EuPaGDT by default is programmed to use "NGG" as an on-target PAM and "NAG" and "NGA" as off-target PAM motifs for the most widely-used SpCas9. However, users can specify multiple, custom on-target PAM motifs. For example,

using the standard International Union of Pure and Applied Chemistry (IUPAC) code (Cornish-Bowden 1985) to specify degenerate PAM motif(s) of 3-10 bp, users can input a degenerative PAM sequence "NNGRRT" recognized by Cas9 from Staphylococcus aureus (SaCas9) plus a variant PAM sequence "NGA" used by Cas9 from Streptococcus pyrogenes (SpCas9), in addition to using the classical PAM motif "NGG" for SpCas9. EuPaGDT also allows users to specify multiple off-target PAM motifs. For example, SpCas9 can recognize the "NGA" PAM motif at a low level (Kleinstiver, Prew et al. 2015); off-targets using such alternative PAMs can be evaluated individually and integrated into the selection process for gRNAs if desired.

Users can refer to the "table of on-/off-targets" pages available from main result output page for each job request to visually inspect each on-target/off-target and their corresponding PAM motifs, as well as the alignment of gRNA with the target sequence. EuPaGDT provides other customized parameters such as gRNA length (as shorter gRNAs are shown to have fewer off-targets (Fu, Sander et al. 2014)) and on-target and off-target search parameters can be custom specified to accommodate users unique needs. For example, (1) searching for on-targets in a fast evolving multigene family might require relaxation of on-target searching criteria, lowering the "identity" and "coverage" parameter values in small decrements to determine if gRNA on-target numbers increase with a steady number of off-targets; and (2) selecting a shorter gRNA length might require a corresponding decrease in the "maximum number of mismatches" allowed in the off-target parameter setting to ensure accurate off-target evaluation.

Microhomology-mediated end-joining (MMEJ) has been shown to repair double-stranded breaks (DSB) generated by CRISPR-Cas9 system in mammalian cells (Wang, Yang et al. 2013, Bae, Kweon et al. 2014) and eukaryotic parasites (Peng, Kurup et al. 2015), often resulting in local sequence deletions which disrupt target genes. Analyzing pairs of microhomology

sequences that flank gRNA can help predict size of MMEJ-mediated gene deletions. EuPaGDT identifies gRNA-flanking microhomology pairs of length 5-20bp within 500bp of each gRNA. Although MMEJ-mediated DSB repair is known to be involved in DSB repair in trypanosomes, to date, the specific rules governing microhomology length and proximity to DSB with respect to the efficiency of MMEJ DSB repair are not known (Glover, Jun et al. 2011, Peng, Kurup et al. 2015). Therefore, EuPaGDT assigns each gRNA is scored on a scale of 0-1 reflecting length of microhomology pairs and their proximity to the gRNA-directed cut site, with a score of 1 for an ideal pair of microhomology (>20 in length, and immediately flanking the gRNA cut site).

In addition to relying on MMEJ-induced deletions to mutate specific sequences, single-stranded oligonucleotides (ssODN) bearing homology arms bracketing gRNA-guided DSB can be used to introduce modifications at specific target positions by homology directed repair (HDR; (Wang, Yang et al. 2013, Wu, Liang et al. 2013, Yang, Guell et al. 2013). Using such repair templates may also obviate undesired changes in the targeted gene (e.g. chromosome translocation or large scale deletions; (Brunet, Simsek et al. 2009, Lee, Kim et al. 2010, Cho, Kim et al. 2014). To aid researchers in designing oligonucleotide repair templates, EuPaGDT will by default generate an archetype ssODN sequence for each gRNA. Each ssODN repair donor has 30bp of homology arms flanking gRNA's predicted DSB site and an 11bp sequence consisting of 3 stop codons in 3 reading frames that will be inserted into targeted sequence at the DSB site. Our default insertion site is at the DSB site (3bp upstream of PAM motif in the case of SpCas9) because editing efficiency is highest when the desired nucleotide change is proximal to DSB site (Yang, Guell et al. 2013, Bialk, Rivera-Torres et al. 2015). Users can specify the length of the homology arms, and also enter desired custom insertion-sequence in place of the

default 3 frame stop codon sequence. Further custom changes such as sequence deletion, nucleotide changes, or insertion adjacent to DSB site can be easily made at user's discretion.

Additional quality control is performed by EuPaGDT to ensure that gRNA will be transcribed efficiently. EuPaGDT checks each identified gRNA for the presence of DNA motifs that may inhibit or terminate RNA polymerase transcription (Bogenhagen and Brown 1981). EuPaGDT will also remind users to add a leading "G" for efficient initiation of transcription when gRNA does not start with "G".

EuPaGDT ranks all gRNAs found in an input sequence based on their total score which is calculated by unweighted averaging of respective target score, activity-prediction score and microhomology-pair scoring. Our repeated test runs using a variety of input gene sequences show that the ranking process performs well in placing gRNAs with more desirable traits closer to the top of the list. Users can rapidly identify desirable gRNAs using the ranked list and further choose gRNA of interest based on specific usage, e.g. targeting a specific region of the input sequence.

For each request, EuPaGDT produces a summary page, including a schematic diagram showing each gRNA's position and strand in input sequence, and a ranked list of such gRNAs along with concise results of each characterization step (Fig. 3.3). At a glance, users can easily grasp essential information of each gRNA. Additional detailed information (such as on-/off-targets summary (Fig. 3.2a and b), archetype ssODNs (Fig. 3.4), and summary of microhomology pairs (Fig. 3.5)) from each characterization step is also available to the user to allow for selection of gRNA suitable to the project.

EuPaGDT also features a "batch mode", which can process a list of genes and return a user-defined number of top-ranking gRNAs for each gene. Additionally, a multi-threaded batch

mode is available to process genome-scale gene lists for genome-wide screening projects. EuPaGDT is currently available for 25 eukaryotic genomes, covering major lineages of eukaryotic pathogens as well as several popular model organisms. Users can also upload custom genomes or request default genomes to be available in EuPaGDT.

Informed and well-scrutinized gRNA design is instrumental to successful CRISPR-Cas9-mediated genome editing or gene expression manipulation. Here we described a gRNA design web tool, EuPaGDT, tailored to eukaryotic pathogen gRNA design. By characterizing potential gRNAs with an array of currently accepted principles, EuPaGDT can facilitate researchers of eukaryotic pathogens to arrive at proper gRNA design for CRISPR-Cas9 experiments and thus bring new understanding for these otherwise difficult to manipulate pathogens.

3.5 CONCLUSION

We developed EuPaGDT, a web-based tool, to assist the eukaryotic pathogen research community in designing gRNA and ssODN HDR templates for CRISPR-Cas9 experiments.

3.6 REFERENCES

Bae, S., J. Kweon, H. S. Kim and J. S. Kim (2014). "Microhomology-based choice of Cas9 nuclease target sites." Nature Methods **11**(7): 705-706.

Bialk, P., N. Rivera-Torres, B. Strouse and E. B. Kmiec (2015). "Regulation of Gene Editing Activity Directed by Single-Stranded Oligonucleotides and CRISPR/Cas9 Systems." PLoS One **10**(6): e0129308.

Bogenhagen, D. F. and D. D. Brown (1981). "Nucleotide sequences in Xenopus 5S DNA required for transcription termination." Cell **24**(1): 261-270.

Brunet, E., D. Simsek, M. Tomishima, R. DeKelver, V. M. Choi, P. Gregory, F. Urnov, D. M. Weinstock and M. Jasin (2009). "Chromosomal translocations induced at specified loci in human stem cells." Proc Natl Acad Sci U S A **106**(26): 10620-10625.

Cho, S. W., S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae and J. S. Kim (2014). "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases." Genome Res **24**(1): 132-141.

Cornish-Bowden, A. (1985). "IUPAC-IUB SYMBOLS FOR NUCLEOTIDE NOMENCLATURE." Nucl. Acids Res. **13**: 3021-3030.

Doench, J. G., E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier and D. E. Root (2014). "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation." Nat Biotechnol **32**(12): 1262-1267.

Fu, Y. F., J. D. Sander, D. Reyon, V. M. Cascio and J. K. Joung (2014). "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." Nature Biotechnology **32**(3): 279-284.

Ghorbal, M., M. Gorman, C. R. Macpherson, R. M. Martins, A. Scherf and J. J. Lopez-Rubio (2014). "Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system." Nat Biotechnol **32**(8): 819-821.

Glover, L., J. Jun and D. Horn (2011). "Microhomology-mediated deletion and gene conversion in African trypanosomes." Nucleic Acids Res **39**(4): 1372-1380.

Kleinstiver, B. P., M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng, A. P. Gonzales, Z. Li, R. T. Peterson, J. J. Yeh, M. J. Aryee and J. K. Joung (2015). "Engineered CRISPR-Cas9 nucleases with altered PAM specificities." Nature.

Lee, H. J., E. Kim and J. S. Kim (2010). "Targeted chromosomal deletions in human cells using zinc finger nucleases." Genome Res **20**(1): 81-89.

Lee, M. C. and D. A. Fidock (2014). "CRISPR-mediated genome editing of Plasmodium falciparum malaria parasites." Genome Med **6**(8): 63.

Peng, D., S. P. Kurup, P. Y. Yao, T. A. Minning and R. L. Tarleton (2015). "CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*." MBio **6**(1): e02097-02014.

Shen, B., K. M. Brown, T. D. Lee and L. D. Sibley (2014). "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9." MBio **5**(3): e01114-01114.

Sidik, S. M., C. G. Hackett, F. Tran, N. J. Westwood and S. Lourido (2014). "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9." Plos One **9**(6).

Sollelis, L., M. Ghorbal, C. R. MacPherson, R. M. Martins, N. Kuk, L. Crobu, P. Bastien, A. Scherf, J. J. Lopez-Rubio and Y. Sterkers (2015). "First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites." Cell Microbiol.

Wagner, J. C., R. J. Platt, S. J. Goldfless, F. Zhang and J. C. Niles (2014). "Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum." Nat Methods **11**(9): 915-918.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell **153**(4): 910-918.

Wu, Y., D. Liang, Y. Wang, M. Bai, W. Tang, S. Bao, Z. Yan, D. Li and J. Li (2013). "Correction of a genetic disease in mouse via use of CRISPR-Cas9." Cell Stem Cell **13**(6): 659-662.

Yang, L., M. Guell, S. Byrne, J. L. Yang, A. De Los Angeles, P. Mali, J. Aach, C. Kim-Kiselak, A. W. Briggs, X. Rios, P. Y. Huang, G. Daley and G. Church (2013). "Optimization of scarless human stem cell genome editing." Nucleic Acids Res **41**(19): 9049-9061.

Zhang, C., B. Xiao, Y. Jiang, Y. Zhao, Z. Li, H. Gao, Y. Ling, J. Wei, S. Li, M. Lu, X. Z. Su, H. Cui and J. Yuan (2014). "Efficient editing of malaria parasite genome using the CRISPR/Cas9 system." MBio **5**(4): e01414-01414.

Zheng, J., H. Jia and Y. Zheng (2014). "Knockout of leucine aminopeptidase in Toxoplasma gondii using CRISPR/Cas9." Int J Parasitol.

3.7 FIGURES AND TABLES

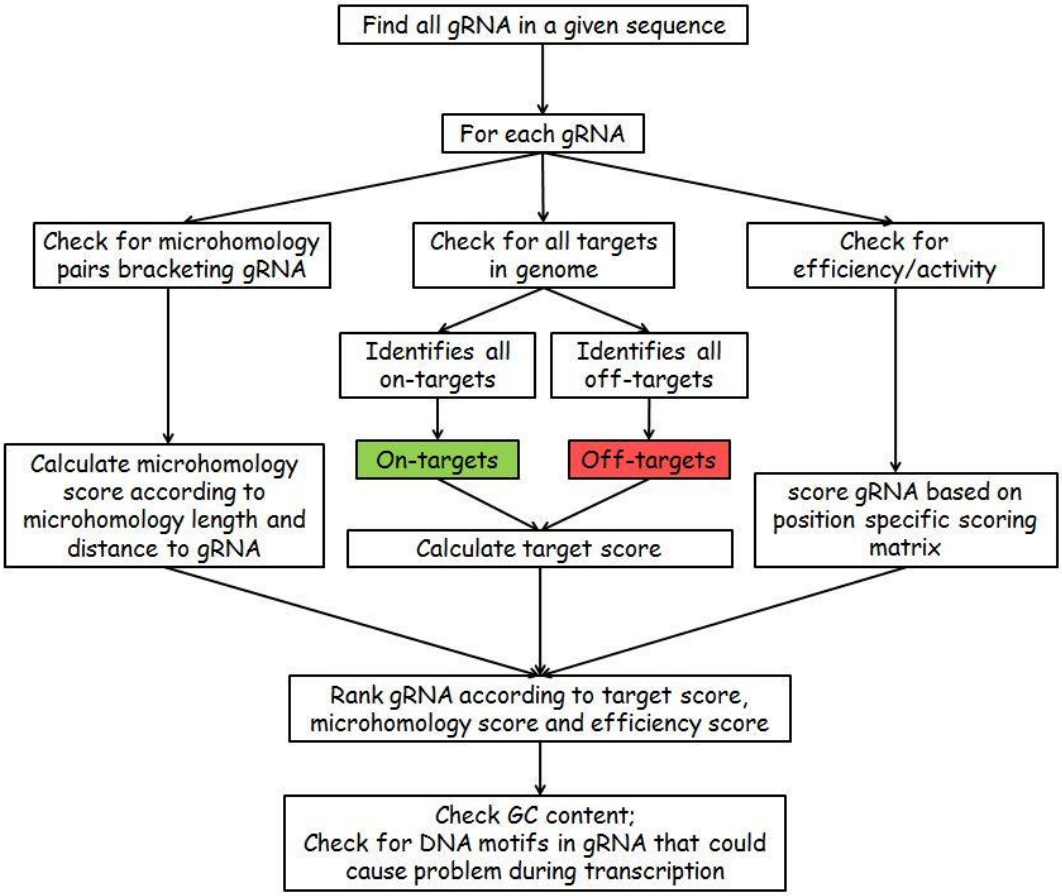


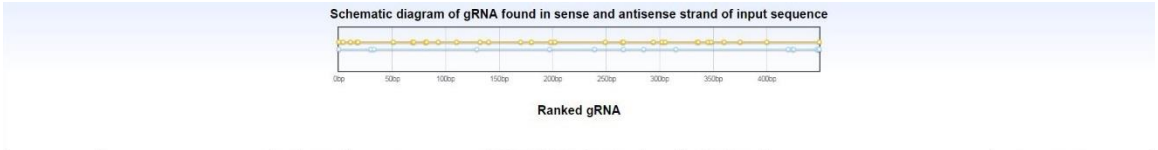
Figure 3.1. Example workflow of a non-batch job request.

gRNA_name(GeneName_gRNAstartPosition)	gRNA_sequence	gRNA_match_start	gRNA_match_end	alignment_identity	genome annotation	match_chromosome	chromosome_match_start	chromosome_match_end
TcCLB.506799.10-FATP_265	GCTATGGCAGAGATCACACA|TGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204139	204161
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204353	204375
TcCLB.506799.10-FATP_266	CTATGGCAGAGATCACACAT|GGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204140	204162
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204354	204376
TcCLB.506799.10-FATP_294	CGTAAGCATGACACTTCTCA|AGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204168	204190
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204382	204404
TcCLB.506799.10-FATP_302	TGACACTTCTCAAGGCATTT|AGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204176	204198
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204390	204412
TcCLB.506799.10-FATP_305	CACTTCTCAAGGCATTTAGG|AGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204179	204201
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204393	204415
TcCLB.506799.10-FATP_335	AATTTGTCGCTTGTGCGCAG|CGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204209	204231
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204423	204445
TcCLB.506799.10-FATP_336	ATTTGTCGCTTGTGCGCAGC|GGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204210	204232
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204424	204446
TcCLB.506799.10-FATP_345	TTGTGCGCAGCGGGAACATC|AGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204219	204241
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204433	204455
TcCLB.506799.10-FATP_348	TGCGCAGCGGGAACATCAGG|TGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204222	204244
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204436	204458
TcCLB.506799.10-FATP_360	ACATCAGGTGGTGTTTGAGA|TGG							
		1	23	100.00	TcCLB.511907.110: fatty+acid+transporter+protein-like%2C+putative	TcChr26-P	204234	204256
		1	23	100.00	TcCLB.506799.10: fatty+acid+transporter+protein-like%2C+putative	TcChr26-S	204448	204470

Figure 3.2. (a) Example output of genomic on-target hits and annotations for 10 gRNAs found in TcFATP gene (gene id TcCLB.506799.10) in the *T. cruzi* CL Brener genome.

gRNA_name(GeneName_gRNAstartPosition)	gRNA_sequence	gRNA_match_start	gRNA_match_end	alignment_identity	genome annotation	match_chromosome	chromosome_match_start	chromosome_match_end
TcCLB.506799.10-FATP_265	GCTATGGCAGAGATCACACA|TGG							
		9	23	100.00	TcCLB.508275.9: dynein+heavy+chain%2C+putative	TcChr30-P	712682	712696
TcCLB.506799.10-FATP_266	CTATGGCAGAGATCACACAT|GGG							
		no off-target hits						
TcCLB.506799.10-FATP_294	CGTAAGCATGACACTTCTCA|AGG							
		no off-target hits						
TcCLB.506799.10-FATP_302	TGACACTTCTCAAGGCATTT|AGG							
		7	23	94.12	TcCLB.506436.120: ATP-dependent+RNA+helicase%2C+putative	TcChr34-P	857218	857234
TcCLB.506799.10-FATP_305	CACTTCTCAAGGCATTTAGG|AGG							
		no off-target hits						
TcCLB.506799.10-FATP_335	AATTTGTCGCTTGTGCGCAG|CGG							
		no off-target hits						
TcCLB.506799.10-FATP_336	ATTTGTCGCTTGTGCGCAGC|GGG							
		no off-target hits						
TcCLB.506799.10-FATP_345	TTGTGCGCAGCGGGAACATC|AGG							
		7	23	94.12	TcCLB.511283.20: protein+transport+protein+Sec13%2C+putative	TcChr40-P	626927	626943
		7	23	94.12	TcCLB.506525.20: protein+transport+protein+Sec13%2C+putative	TcChr40-S	626266	626282
TcCLB.506799.10-FATP_348	TGCGCAGCGGGAACATCAGG|TGG							
		no off-target hits						
TcCLB.506799.10-FATP_360	ACATCAGGTGGTGTTTGAGA|TGG							
		no off-target hits						

Figure 3.2. (b) Example output of genomic off-target hits and annotations for 10 gRNAs found in TcFATP gene in the *T. cruzi* CL Brener genome



Ranked gRNA

gRNA id	gRNA sequence (PAM "NGG")	Total score	efficiency score	On-target hits (perfect-match | nonperfect-but-PAM-match)	Off-target hits (perfect-match | nonperfect-match)	Target score	GC content	Flanking microhomolgy pair score	Flanking microhomolgy pair score explanation	Potential problems during transcription
TcCLB.506799.10-FATP_110	TGCAGGCAAGAAACTCTATG|CGG	0.64	0.60	2 | 0	0 | 0	1.00	0.45	0.33	very good microhomology pair(s) found	gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_375	TGAGATGGCGACAACAGATG|CGG	0.63	0.65	2 | 0	0 | 0	1.00	0.5	0.24	good microhomology pair(s) found	gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_305	CACTTCTCAAGGCATTTAGG|AGG	0.63	0.60	2 | 0	0 | 0	1.00	0.45	0.29	good microhomology pair(s) found	gRNA contains >=3 "T(U)s" in a row, it will terminate POLIII transcription gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_294	CGTAAGCATGACACTTCTCA|AGG	0.63	0.60	2 | 0	0 | 0	1.00	0.45	0.29	good microhomology pair(s) found	gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_51	GCATTACGCATACATCACGA|CGG	0.62	0.60	2 | 0	0 | 0	1.00	0.45	0.25	good microhomology pair(s) found	No problem found
TcCLB.506799.10-FATP_335	AATTTGTCGCTTGTGCGCAG|CGG	0.61	0.65	2 | 0	0 | 0	1.00	0.5	0.19	good microhomology pair(s) found	gRNA contains >=3 "T(U)s" in a row, it will terminate POLIII transcription gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_360	ACATCAGGTGGTGTTTGAGA|TGG	0.61	0.60	2 | 0	0 | 0	1.00	0.45	0.22	good microhomology pair(s) found	gRNA contains >=3 "T(U)s" in a row, it will terminate POLIII transcription gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_170	TTGTGCCGTGCTGCAATTGC|TGG	0.60	0.61	2 | 0	0 | 0	1.00	0.55	0.20	good microhomology pair(s) found	gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_336	ATTTGTCGCTTGTGCGCAGC|GGG	0.60	0.61	2 | 0	0 | 0	1.00	0.55	0.19	good microhomology pair(s) found	gRNA contains >=3 "T(U)s" in a row, it will terminate POLIII transcription gRNA does not start with G, please manually add a leading "G"
TcCLB.506799.10-FATP_266	CTATGGCAGAGATCACACAT|GGG	0.60	0.60	2 | 0	0 | 0	1.00	0.45	0.20	good microhomology pair(s) found	gRNA does not start with G, please manually add a leading "G"

Figure 3.3. Example output of summary page of gRNA found in TcFATP gene in the *T. cruzi* CL Brener genome (only top 10 ranking gRNAs shown).

gRNA id	gRNA sequence	oligo template for stop codon insertion
TcCLB.506799.10-FATP_110	TGCAGGCAAGAAACTCTATG|CGG	+strand:TCTTTGGGTGCCGTGCAGGCAAGAAACTCT**TAGATAGATAG**ATGCGGAACGCAAAGTTAATTTGTAAGGAT
		-strand:ATCCTTACAAATTAACTTTGCGTTCCGCAT**CTATCTATCTA**AGAGTTTCTTGCCTGCACGGCACCCAAAGA
TcCLB.506799.10-FATP_294	CGTAAGCATGACACTTCTCA|AGG	+strand:CACATGGGCAGTCCGTAAGCATGACACTTC**TAGATAGATAG**TCAAGGCATTTAGGAGGCAGTCACAATTTG
		-strand:CAAATTGTGACTGCCTCCTAAATGCCTTGA**CTATCTATCTA**GAAGTGTCATGCTTACGGACTGCCCATGTG
TcCLB.506799.10-FATP_305	CACTTCTCAAGGCATTTAGG|AGG	+strand:TCCGTAAGCATGACACTTCTCAAGGCATTT**TAGATAGATAG**AGGAGGCAGTCACAATTTGTCGCTTGTGCG
		-strand:CGCACAAGCGACAAATTGTGACTGCCTCCT**CTATCTATCTA**AAATGCCTTGAGAAGTGTCATGCTTACGGA
TcCLB.506799.10-FATP_375	TGAGATGGCGACAACAGATG|CGG	+strand:ATCAGGTGGTGTTTGAGATGGCGACAACAG**TAGATAGATAG**ATGCGGCCGTCCCAAAAATGAAGCTTTGTG
		-strand:CACAAAGCTTCATTTTTGGGACGGCCGCATC**TATCTATCTA**CTGTTGTCGCCATCTCAAACACCACCTGAT
TcCLB.506799.10-FATP_51	GCATTACGCATACATCACGA|CGG	+strand:GGGTTATCATGTCGCATTACGCATACATCA**TAGATAGATAG**CGACGGCTGAGTATATAAAACTGGGATGCT
		-strand:AGCATCCCAGTTTTATATACTCAGCCGTCGC**TATCTATCTA**TGATGTATGCGTAATGCGACATGATAACCC
TcCLB.506799.10-FATP_335	AATTTGTCGCTTGTGCGCAG|CGG	+strand:AGGAGGCAGTCACAATTTGTCGCTTGTGCG**TAGATAGATAG**CAGCGGGAACATCAGGTGGTGTTTGAGATG
		-strand:CATCTCAAACACCACCTGATGTTCCCGCTG**CTATCTATCTA**CGCACAAGCGACAAATTGTGACTGCCTCCT
TcCLB.506799.10-FATP_360	ACATCAGGTGGTGTTTGAGA|TGG	+strand:GTGCGCAGCGGGAACATCAGGTGGTGTTTG**TAGATAGATAG**AGATGGCGACAACAGATGCGGCCGTCCCAA
		-strand:TTGGGACGGCCGCATCTGTTGTCGCCATCT**CTATCTATCTA**CAAACACCACCTGATGTTCCCGCTGCGCAC
TcCLB.506799.10-FATP_30_revcom	TCTTCCTACCGTATTGACGG|TGG	+strand:NNNNNNNNNNNNNNNNATGGCGACCACCG**TAGATAGATAG**TCAATACGGTAGGAAGATGGAATTGGGTTA
		-strand:TAACCCAATTCCATCTTCCTACCGTATTGA**CTATCTATCTA**CGGTGGTCGCCATNNNNNNNNNNNNNNNNN
TcCLB.506799.10-FATP_170	TTGTGCCGTGCTGCAATTGC|TGG	+strand:GCATGGAAATTGATTGTGCCGTGCTGCAAT**TAGATAGATAG**TGCTGGAAGAAGAAGGAGAACTCTATTGTT
		-strand:AACAATAGAGTTCTCCTTCTTCTTCCAGCA**CTATCTATCTA**ATTGCAGCACGGCACAATCAATTTCCATGC
TcCLB.506799.10-FATP_266	CTATGGCAGAGATCACACAT|GGG	+strand:CATGATGTTGCCGCTATGGCAGAGATCACA**TAGATAGATAG**CATGGGCAGTCCGTAAGCATGACACTTCTC
		-strand:GAGAAGTGTCATGCTTACGGACTGCCCATG**CTATCTATCTA**TGTGATCTCTGCCATAGCGGCAACATCATG

Figure 3.4. Example output of archetype ssODNs for top 10 ranking gRNA found in the TcFATP gene.

Flanking microhomologies for TcCLB.506799.10-FATP_69

color code:
best microhomolgy based on distance to gRNA
best microhomolgy based on length

Microhomolgy length	Microhomology sequence	Microhomology occurance position	gRNA position
5	ACGCA	56 134	69
5	CATCA	63 361	69
5	GCGAC	4 382	69
5	GGCGA	3 381	69
5	GCATT	51 316	69
5	TGTCG	47 339	69
5	ATCAC	64 277	69
5	TCATG	44 252	69
5	TGGGT	37 101	69
5	GGAAG	25 191	69
5	TGGCG	2 380	69
5	TAGGA	23 321	69
5	TTGGG	36 100	69
5	GTCGC	48 340	69
5	ATGGA	30 159	69
5	AGATG	28 377 390	69
5	GATGG	29 220 378	69
5	GAAGA	26 192 195	69
5	AATTG	34 164 184	69
5	ATGGC	1 268 379	69
5	ACATC	62 236 360	69
5	TGGAA	31 160 190	69
6	TGGCGA	2 380	69
6	TTGGGT	36 100	69
6	TGTCGC	47 339	69
6	GGAAGA	25 191	69
6	AGATGG	28 377	69
6	GGCGAC	3 381	69
6	ATGGAA	30 159	69
6	ATGGCG	1 379	69
6	ACATCA	62 360	69
7	ATGGCGA	1 379	69
7	TGGCGAC	2 380	69
8	ATGGCGAC	1 379	69

Figure 3.5. Example output of microhomology pairs found for a gRNA in the TcFATP gene.

CHAPTER 4

KNOCKDOWN OF THE *TRANS*-SIALIDASE GENE FAMILY IN *TRYPANOSOMA CRUZI*

4.1 SUMMARY

In this chapter, I utilized the genome-editing techniques and tools developed in Chapter 1 and 2 to knockdown the expression of the TcTS gene family repertoire in *T. cruzi*, with the aim of producing a TcTS knockdown parasite line that enable the testing of a long-standing hypothesis: the TcTS gene family participates in evasion of the immune response by *T. cruzi*. I designed a set of 21 gRNAs and associated repair templates, which in combination, target 63% of all TcTS genes and 92% of the TcTS genes capable of generating protein products. The DNA repair template for each gRNA allowed insertion of GFP11 peptide tags prior to stop codons into TcTS genes for tracking TcTS-editing events. I sequentially delivered each gRNA in complex with Cas9 along with corresponding DNA repair template into *T. cruzi* at transfection intervals of 3-5 days. Metacyclic and culture-derived trypomastigotes from the knockdown line showed a decreased surface *trans*-sialidase protein level compared to wild-type controls. GFP11 tags could be detected in the genomic DNA of the TcTS knockdown parasites. Sequence analysis of the inserts is expected to provide additional evidence of large-scale disruption of TcTS genes. Future studies will assess the ability of TcTS-depleted parasites to establish chronic infections in mice.

4.2 INTRODUCTION

The TcTS gene family in *T. cruzi* consists of thousands of genes that likely continue to be amplified and diversified through recombination(Weatherly, Peng et al. 2016). *Trans*-sialidase activity and lectin-like binding activity cannot justify the maintenance of such a sizable repertoire. It is hypothesized that the repertoire of the TcTS gene family plays a role in the immune evasion process through expression of a large number of its members and/or by generating new TcTS variants using the repertoire as a pool of genetic material. To test this hypothesis, one has to block the expression of nearly all TcTS genes, except a few TcTS bearing *trans*-sialidase enzymatic activity. To accomplish this in *T. cruzi*, which lacks an active RNAi pathway, I leveraged the CRISPR/Cas9 system to achieve thousands of genome-editing events in order to disable the coding capability of the majority of TcTS genes. I also aimed to disable the coding capability of silent TcTS genes, since frequent intra-family recombination has been documented in the TcTS gene family, and expressed TcTS gene have mosaic pieces originating from silent TcTS, indicating genetic flow from silent TcTS to expressed TcTS (Weatherly, Peng et al. 2016). Silent TcTS can contain long stretches of protein-coding sequences, and if not disrupted, could be used to restore full coding capability of expressed TcTS through recombination.

In Chapter 2, I demonstrated that the CRISPR/Cas9 system can be used to efficiently knockout *T. cruzi* genes and knockdown a medium-sized *T. cruzi* gene family (65 genes), serving as proof of feasibility in knocking down a gene family in *T. cruzi* using the CRISPR/Cas9 system. We've also demonstrated CRISPR/Cas9-mediated insertion of user-defined sequences via the homologous directed repair pathway using DNA donor repair template, making it possible to insert DNA tags into TcTS genes, for the purpose of tracking TcTS gene editing events and assessing the progress of TcTS knockdown.

In Chapter 3, I developed a gRNA design tool for eukaryotic pathogens (EuPaGDT). In addition to common functionalities, e.g. off-target searches, efficacy prediction etc, EuPaGDT can perform on-target hit search in the genome, a vital feature to have when designing gRNAs targeting a gene family. Batch mode gRNA predictions for all TcTS genes followed by analyzing all TcTS-targets of each of the designed gRNA enabled us to identify a small set of gRNAs that targeted a majority of the TcTS genes. In this Chapter, I utilized these tools to knockdown expression of the TcTS gene family, including disrupting the coding potential of silent members, and then assessed the degree and impact of TcTS knockdown.

4.3 RESULTS

4.3.1 IDENTIFICATION OF *TRANS*-SIALIDASE GENES IN *T. CRUZI* BRAZIL STRAIN CLONE A4

T. cruzi Brazil strain clone A4 was chosen to conduct the TcTS gene family knock down for 2 reasons. The genome of Brazil strain clone A4 was recently sequenced by the Tarleton lab using Pacbio single molecule real-time (SMRT) sequencing technology with long reads (77% of reads > 3000bp), which can cover whole or large parts of full length TcTS genes (2000bp-3000bp), circumventing the difficulty of resolving the origin of short reads (75-300bp) from TcTS genes sharing high nucleotide identity, as exists for the reference CL Brener genome (Weatherly, Boehlke et al. 2009).Additionally, unlike the reference CL Brener strain, the Brazil clone A4 is non-hybrid strain with identical alleles at most gene loci.

I first identified TcTS genes in the genome of Brazil strain clone A4. Using a collection of 3209 TcTS gene sequences from the CL Brener strain (Weatherly, Peng et al. 2016), to search the genome of Brazil strain clone A4 using BLAST2.2.29+, I identified 4789 TcTS-like sequences with 90% or greater nucleotide identity. To confirm that the TcTS-like sequences are

sufficiently-sialidase-like, all 4789 TcTS-like sequences were BLASTed against the *T. cruzi* CL Brener genome, and only the sequences whose best match was annotated as TcTS were retained. Using these search criteria, I identified 2560 TcTS genes in the *T. cruzi* Brazil strain clone A4.

4.3.2 IDENTIFICATION OF GRNAS TARGETING THE TCTS FAMILY

As TcTS genes are highly similar to each other, I selected a subset of TcTS genes to represent the entire TcTS gene family to design gRNAs. The subset of TcTS genes was selected based on the presence of 2 epitope groups: TSKB20 and TSKB18. TSKB20 is an 8 amino acid epitope (ANYKFTLV). CD8+ T cells specific for TSKB20 epitope can also recognize (cross-react with) other epitopes in the TSKB20 group, which includes TSKB21(ANYNFTLV) and TSKB260 (ANYKFTLL) epitope (Martin, Weatherly et al. 2006). Similarity, the TSKB18 (ANYDFTLV) epitope forms a cross-reacting group with 3 more epitopes (TSKB74 VNYDFTLV, TSKB80 ANYNFTLL and TSKB89 VNYDFTIV) (Martin, Weatherly et al. 2006).

The advantages of targeting all TcTS genes that harbor TSKB20 and TSKB18 epitope groups is that it enables the monitoring of the progress of TcTS knockdown via tracking the T cell response to these epitopes. The decrease of expression of this subset of TcTS epitopes would be expected to elicit lower numbers of CD8+ T cells recognizing the TSKB20 and TSKB18 epitope groups.

I identified 468 and 39 TcTS genes encoding TSKB20 and TSKB18 epitope groups, respectively, from the total of 2560 TcTS genes in the *T. cruzi* Brazil strain A4 clone. Next, I used the Eukaryotic Pathogen gRNA design Tool (EuPaGDT) to enumerate all gRNAs found in these TcTS genes. I identified a total of 14,937 gRNAs that are SaCas9 compatible (21nt, PAM: NNGRRT). For each gRNA, EuPaGDT lists all the family-wide on-target sites (on-target search

algorithm described in chapter 2), as well as all off-targets in the genome. Two additional filters were applied to remove gRNAs that (1) have an efficacy score lower than 0.5, and (2) target sequences in the TcTS genes located 3' to the coding sequences of the TSKB20 and TSKB18 epitope groups (such guides would not disrupt the expression of these epitopes).

Next, I developed a custom Perl script to compile all the gRNAs and conduct a search for a set of gRNAs that collectively target all the 468 TcTS genes containing the TSKB20 epitope group and the 39 TcTS that contain the TSKB18 epitope group. Either expressing or delivering a large number of gRNA poses some technical challenges. In order to minimize the number of gRNAs needed to KO a large number of TcTS genes, I developed a method similar to the Metropolis algorithm (Metropolis, Rosenbluth et al. 1953) to identify a small set of gRNAs targeting the highest number of TcTS genes. Starting with a user-provided, initial gRNA, the script evaluates the percentage of TcTS genes targeted by that gRNA, then it proposes a new gRNA to add (randomly drawn from the pool of all TcTS-targeting gRNAs) and forms a new set of gRNAs. Next, the script evaluates the percentage of TcTS genes being targeted by the new set of gRNAs. If the percentage increase is greater than the threshold that is set empirically by the user (see Material and Methods for details), the addition of the randomly drawn gRNA is accepted. The script repeatedly proposes new gRNAs to add to the set until the percentage of targeted genes approaches 100%.

To increase the search space for gRNA sets, I executed this Perl script using different initial gRNAs and different threshold parameters (see Material and Methods for details). The results ultimately converged on one set of gRNAs. In this set, 7 gRNAs can target all 39 TcTS genes that contain the TSKB18 epitope group and 14 gRNAs can target all 468 TcTS genes that

contain the TSKB20 epitope group. To summarize, I arrived at a 21-gRNA-set for SaCas9 which can target all the TcTS genes that contain the TSKB18 and TSKB20 epitope groups (Table 4.1).

Next, I assessed if the gRNA sets can target a significant number of the overall 2560 TcTS genes. Using a custom Perl script, I mapped all on-target hits of these gRNAs to the genome. A TcTS gene is considered as being targeted if it contains at least one gRNA target site. This analysis revealed that the 21-gRNA-set can target 1607 out of 2560 TcTS genes (63%). Many of the TcTS genes not being targeted are small pieces under 900bp in length, making up 73% of the untargeted genes (Fig. 4.1). Moreover, only 28 of the untargeted TcTS can produce peptides, as indicated by the presence of start codons. In contrast, 231 of the targeted TcTS can produce peptides. In summary, I designed a set of 21 gRNAs that in combination target a significant number of TcTS genes, and in theory, achieve a drastic reduction of TcTS protein production.

The reason for choosing gRNA compatible with the SaCas9 protein (a Cas9 variant) versus the more widely used SpCas9 is that delivery of SaCas9/sgRNA into *T. cruzi* is more efficient. SaCas9 has a smaller molecular weight than SpCas9, and it can be electroporated (in complex with gRNA) directly into *T. cruzi* (see Chapter 4.3 for details) (Soares Medeiros, South et al. 2017), while SpCas9 must be first expressed in *T. cruzi* and then the gRNAs are electroporated in, as needed. The problem with constitutively expressing SpCas9 is that the *T. cruzi* population expressing SpCas9 gradually loses SpCas9 activity (Peng, Kurup et al. 2015) over the course of time that is needed for separate deliveries of a set of 12 gRNAs (gRNA set for SpCas9 determined as described above).

4.3.3 DESIGN OF DNA REPAIR TEMPLATE

The rules that govern the rejoining of double-stranded break (DSB, introduced by Cas9/gRNA) in the absence of DNA repair template is not fully understood in *T. cruzi*. The rejoining of DSB without a repair template could lead to unwanted chromosomal rearrangement and large-scale sequence deletions, at least in *Leishmania* (data not shown). The co-delivery of DNA repair template with the SaCas9/gRNA complex will enable the homologous directed repair of DNA double-stranded breaks (DSB), which joins the DSB with insertion sequences in the DNA template provided. The insertion of small tag sequences will allow us to detect intended genomic events of TcTS inactivation. Because of these 2 advantages, I decided to accompany each gRNA delivery with a corresponding DNA repair template.

Using a custom Perl script, I extracted the flanking sequences of all target-sites. Next, I aligned the flanking sequences separately for each gRNA. A consensus sequence was extracted from each alignment, using a majority rule principle: if representation of any nucleotide is over 30%, but less than 70%, I incorporate a corresponding degenerate nucleotide for that position.

Each gRNA's DNA repair template consists of left and right consensus flanking sequences interposed by a universal GFP11 tag preceded by a GS linker and followed by a stop codon to terminate the translation of TcTS protein. (Fig. 4.2) The GS linker allows GFP11 to protrude out of the peptide fold and allow flexibility to be better detected by complementation with GFP1-10 protein, which restores the GFP fluorescence in GFP1-10 protein (Fig. 4.2).

Using the pipeline described above, I designed 21 DNA repair templates for 21 gRNAs (SaCas9).

4.3.4 DELIVERY OF SACAS9-GRNA INTO *T. CRUZI*

Sequential delivery of each of the 21 gRNA (SaCas9) is preferred because all 21 gRNAs will make a combined total of 6681 double-stranded breaks in the genome, which could potentially facilitate chromosomal rearrangements and/or death of the parasite. Each delivery of a gRNA will be together with SaCas9, forming a ribonucleoprotein complex, accompanied by a corresponding repair template.

The author of this thesis and colleagues in the Tarleton lab have optimized a protocol for the delivery of multiple SaCas9-gRNA ribonucleoprotein complexes into the same population of parasites (Soares Medeiros, South et al. 2017). I observed that multiple electroporation pulses during the same transfection reaction increases gene knockout efficiency. *T. cruzi* parasites receiving 2 consecutive electroporation pulses had a 102% increase in the proportion of knockout parasites compared to that receiving only 1 electroporation pulse (Fig. 4.3a). I also observed that repeated transfection of the same SaCas9/gRNA RNP in 2 days intervals can achieve close to 100% knockout efficiency (Fig. 4.3b). Since 1280 out of 1607 *trans*-sialidase targets can be targeted by more than 1 gRNA, in other words, the 21 gRNAs have significant overlap of targeted TcTS genes, it would be beneficial to deliver 21 gRNAs by separate transfections, so that many TcTS are targeted repeatedly to drive the knockout to completion. I further chose to conduct 2 electroporation pulses in each transfection reaction, to increase the gene knockout efficiency for each gRNA, but at the cost of more parasite death caused by the electroporation process. Thus, a single pool of parasites was pulsed twice with a single RNP complex, then cultured for 3-6 days until the cell number returned to the pre-transfection level before transfecting with the next RNP plus repair template.

4.3.5 ASSESSMENT OF TCTS KNOCKDOWN

4.3.5.1 PARASITE SURFACE TCTS EXPRESSION

As a first indicator of TcTS expression reduction, I assessed the surface TcTS protein levels of the *T. cruzi* population that received 21 gRNAs. As *T. cruzi* has a much higher expression of TcTS in the trypomastigote form compared to the epimastigote form, I converted *T. cruzi* from epimastigotes to metacyclic-trypomastigotes using chemically defined differentiation media TAU (Contreras, Salles et al. 1985) and TAU3AAG (Goldenberg, Contreras et al. 1987) (see Material and Methods for details). I stained for parasite surface TcTS protein using fluorescently labeled fetuin protein and performed flow cytometry analysis. Fetuin is a glycoprotein containing sialylated glycans which are capable of binding to TcTS protein. Fetuin staining of the differentiated parasites revealed 2 populations: a low fetuin-staining population matching the fetuin-staining level of epimastigotes (suggesting an incomplete epimastigotes to metacyclic conversion), and a high fetuin-staining population corresponding to metacyclic-trypomastigotes (Fig. 4.4). In the TcTS knockdown group, the fetuin staining of the high fetuin-staining population (corresponding to metacyclic-trypomastigotes) is half that of the wild type control, while the low fetuin staining population (not fully differentiated epimastigotes) has comparable fetuin staining signal with the controls.

The low fetuin-staining population in the TcTS knockdown parasites were sorted by flow cytometry and used to infect vero cells. Because of the concern that the *trans*-sialidase-low population would be inefficient at cell invasion, due to the necessity of TcTS activity for that process (Piras, Henriquez et al. 1987, Rubin-de-Celis, Uemura et al. 2006, Carvalho, Sola-Penna et al. 2010), I incubated the parasites with filtrated supernatant from late-stage cultures of infected vero cells, which contains *trans*-sialidase released from wild-type trypomastigotes

(Schenkman, Vandekerckhove et al. 1993). The resulting trypomastigotes produced by this infection cycle were predominantly low fetuin staining trypomastigotes (Fig. 4.5).

4.3.5.2 ASSESSING GFP11 TAG INSERTION

A second method to assess the knockdown of TcTS is by analyzing the GFP11 tag inserted into the TcTS genes just upstream of the truncating stop codon within the repair templates (Fig. 4.2). The GFP11 peptide complements the GFP1-10 protein fragment and restores its GFP fluorescence (Cabantous, Terwilliger et al. 2005). Unfortunately, however I was unable to detect recovery of GFP fluorescence when I expressed both GFP11 and GFP1-10 in the cytoplasm of *T. cruzi* epimastigotes (data not shown). Nevertheless, we could still use the molecular detection of the GFP11 tag insertion to assess TcTS gene disruption. For this, I adopted the nonspecific primer anchored PCR technique developed by Zhang et al. (Zhang, Liu et al. 2011). In this PCR technique, 2 forward primers in the GFP11 sequence and 2 reverse random primers were used in a nested PCR. In each PCR step, special ramp speeds, annealing temperature, and primer concentration were set to accommodate the annealing of random primers to the DNA template (see material and methods for details). The combined use of GFP11 primers and the random primers should result in amplicons consisting of GFP11 and its flanking sequences. If GFP11 sequences were inserted as I predicted, the flanking sequences should map to TcTS genes, thus providing the location of the GFP11 tag in the genome. Nonspecific primer anchored PCR using genome DNA from TcTS knockdown parasites and wildtype controls produced different amplification patterns (Fig. 4.6). I sequenced 8 single molecules for the TcTS knockdown parasites and the wildtype control amplicons and 7/8 amplicons in the TcTS knockdown sample contained the GFP11 tag and associated flanking sequences that mapped to 6 distinct TcTS genes. None of the amplicons from the wildtype sample contained the GFP11 tag.

These results demonstrate that the GFP11 tag is only present in the TcTS knockdown parasites, consistent with the expected outcome of the CRISPR/Cas9 mediated editing of the TcTS genes and (2) using GFP11 and nonspecific primers, I can specifically amplify GFP11 and its flanking sequences with only minor nonspecific amplicons.

Deep sequencing of the GFP11-flank amplicons enables us to accurately assess the TcTS knockdown at the genetic level, given sufficient length of flanking sequences were obtained to accurately map each GFP11 insertion sequence to a specific TcTS gene. We deep sequenced the regions flanking GFP11 tag in 17 clones derived from the TcTS knockdown population. The number of TcTS genes with stop insertions ranged from 283 to 790, representing 21 to 59 percent of all targetable TcTS genes. (Table 4.2).

4.3.5.3 PREPARING TCTS KNOCKDOWN *T. CRUZI* FOR MOUSE INFECTION

The knockdown of 1000's of (mostly enzymatically inactive) TcTS genes in *T. cruzi* might also decrease the production of enzymatically active TcTS required for successful host cell invasion and potentially attenuate infection in mice. To eliminate this possibility, I overexpressed the enzyme-active TcTS ts611/2 (Buschiazzo, Campetella et al. 1997) in the TcTS knockdown parasites. Ts611/2 does not encode either the TSKB20 or TSKB18 epitope groups, and thus it should not impact induction of CD8 T cell responses to these epitopes. I also overexpressed a tdTomato gene in the TcTS knockdown/Ts611/2 complemented parasites, allowing easier tracking and quantification of parasites *in vivo* during mouse infections. Finally, the TcTS knockdown/Ts611/2 tdTomato-expression parasite population was cloned by limiting dilution in order to generate single lines with uniform and definable TcTS knockdown patterns.

4.4 DISCUSSION

In this Chapter, I utilized the CRISPR/Cas9 system to disrupt the coding capability of both expressed and silent TcTS genes in *T. cruzi*. By simultaneously targeting both expressed and silent TcTS genes, I not only knocked-down TcTS expression, but also aimed at impeding the reversion of TcTS expression. My previous work has shown evidence of genetic flow from silent TcTS gene to its expressed counterparts (Weatherly, Peng et al. 2016). This flow of genetic material could mediate the regeneration of coding capability in disrupted TcTS. Parasites that have performed such "repair" of TcTS genes could lead to increase to TcTS expression and this phenotype could be selected for in a mouse infection, obscuring the TcTS knockdown effect. Selection of recombined surface protein coding genes has been observed in mice infected with *T. brucei* (Jackson, Berry et al. 2012, Horn 2014, Mugnier, Cross et al. 2015). Thus, targeting both the expressed and silent TcTS would represent a better strategy than only targeting expressed TcTS in obtaining a stable line of TcTS knockdown *T. cruzi* parasites.

The TcTS knockdown *T. cruzi* population showed a 55% decrease in surface Fetuin-binding capability. Fetuin can bind to TcTS proteins through its sialylated glycans. Since fetuin can also bind to other (non-TcTS) molecules on the parasite surface, the surface TcTS reduction is likely underestimated using the fetuin staining method. Furthermore, parasites in the TcTS knockdown population likely have a spectrum of reduction of TcTS expression. Future work will analyze the TcTS expression level of clonal lines derived from this population using fetuin staining and TcTS transcript analysis (RNAseq). A clonal line with the least TcTS expression will be identified by cross-referencing these results.

4.5 MATERIALS AND METHODS

4.5.1 IDENTIFICATION OF TCTS GENES ENCODING TSKB20 AND TSKB18 EPITOPE GROUPS

TSKB20 and TSKB18 epitope group sequences were searched in the *trans*-sialidase genes using tblastn with customized option parameters "-word_size 2" and "-evalue 100"

The TSKB20 epitope group, TSKB20: ANYKFTLV, TSKB21: ANYNFTLV, TSKB260: ANYKFTLL

The TSKB18 epitope group, TSKB18: ANYDFTLV, TSKB74: VNYDFTLV, TSKB89: VNYDFTIV, TSKB80: ANYNFTLL

4.5.2 COMPILING ALL THE GRNAS AND GENOMIC TARGETS SITES FROM EUPAGDT WEBSERVER OUTPUT

2 Perl scripts ("1.findgRNA_and_targetsites.pl" and "2.collapse_gRNA_seq.pl") were executed using Perl v5.18.2 in Ubuntu 14.04.4 LTS environment. The Perl scripts are available at https://github.com/duopeng/compile-gRNA-and-targets-for-EuPaGDT-batch-job-

4.5.3 SEARCHING FOR THE SMALLEST SET OF GRNAS THAT TARGETS A SUBSET OF TCTS GENES

A Perl script ("3.find_gRNA_combination.pl") was executed using Perl v5.18.2 in Ubuntu 14.04.4 LTS. A Linux shell script was used to repeatedly execute the Perl script under different option parameters. Top 10 gRNAs with highest number of TcTS targets was used as the initial gRNA in different execution instances. For each initial gRNA, I executed the Perl script using 9 different settings of threshold of target number increment: 1,2,4,6,10,20,30,40,50. A total

of 90 instances of Perl script execution were conducted, producing 90 sets of gRNAs. I chose the best set based on the number of gRNAs in each set, the average efficacy score.

The Perl script is available at https://github.com/duopeng/find-gRNA-set

4.5.4 EXTRACTION OF FLANKING SEQUENCES OF TARGET-SITES AND ALIGNMENT

A Perl script ("3.find_gRNA_combination.pl") was executed using Perl v5.18.2 in Ubuntu 14.04.4 LTS. The Perl script is available at https://github.com/duopeng/get-flank-of-gRNA-target. The Perl script output was viewed as DNA alignment in Geneious R10.

4.5.5 GROWTH, TRANSFECTION AND CLONING OF *T. CRUZI*

T. cruzi epimastigote was cultured and transfected as described in Chapter 2. For double-pulse transfection, parasites were electroporated twice without interval using program "U-33" in an AMAXA nucleofector device. Clonal lines of *T. cruzi* were obtained by limiting dilution. Briefly, *T. cruzi* epimastigotes were diluted to a concentration of 0.5 parasite/200 µl in LDNT medium, 200 µl of diluted culture were placed in each well of 96-well plates. The LDNT medium used in limiting dilution was spiked with 0.5% spent media (0.2 µM filtered LDNT medium supernatant from active *T. cruzi* epimastigote culture).

4.5.6 SGRNA/SACAS9 RIBONUCLEOPROTEIN PREPARATION

sgRNA was prepared by in vitro transcription and purified as described in Chapter 2. SaCas9 protein were produced by the colleagues of the Tarleton lab as described (Medeiros 2017, accepted). To prepare the sgRNA/SaCas9 ribonucleoprotein, 5 µl of gRNA (>10 µg/ul) was incubated at 90 ℃ for 5 minutes, then gradually cooled to room temperature at a ramp rate of -0.1 ℃/s. 10 µl of SaCas9 (5mg/ml) was mixed with sgRNA and incubated at room

temperature for 15 minutes to allow formation of sgRNA/SaCas9 ribonucleoprotein. 7.5 µl of SaCas9/sgRNA ribonucleoprotein and 7.5 µl single stranded DNA repair template (5mg/ml) was used in each transfection.

4.5.7 METACYCLOGENESIS

10^7 log-phase *T. cruzi* epimastigotes were centrifuged at 3000xg for 10 min, and gently resuspended in 10ml TAU media (Contreras, Salles et al. 1985) in a T75 flask. Parasites were incubated at 27℃ for 2 h in a horizontal position to allow attachment to flask. Supernatant were carefully discarded. 10 ml of TAU3AAG (Goldenberg, Contreras et al. 1987) media was added to the flask. Parasites were incubated at 27℃ for 6 days in a horizontal position. Collect the supernatant of the flask can carefully wash the wall of the flask with 5 ml. Centrifuge the supernatant at 3000g for 10 min to collect *T. cruzi* parasites in a pellet.

4.5.8 FETUIN LABELING AND STAINING

Fetuin (NEB Cat. # P6042S) was labeled with AF633 using the Alexa Fluor™ 633 Protein Labeling Kit (ThermoFisher) following manufacturer's instructions. Briefly, 50 µl of 1M bicarbonate was added to 500 µl of 1mg/ml Fetuin protein. 1 vial of reactive dye was added to the protein solution and incubated for 1 hour at room temperature with continuous stirring. Labeled fetuin was purified using a column packed with Bio-Rad BioGel P-30 Fine size exclusion purification resin (provided by protein labeling kit). Degree of labeling was determined to be 0.52 moles dye per mole protein following manufacturer's instructions.

To perform Fetuin-AF633 staining, *T. cruzi* epimastigotes (2×10^7) were washed twice in PBS, then incubated at room temperature for 15 min with 8 µl Fetuin in 1000 µl PBS. Parasites were analyzed by a CyAn flow cytometer (Beckman Coulter).

4.5.9 VERO CELL CULTURE AND *T. CRUZI* INFECTION

Vero cells were seeded onto T75 flasks in complete RPMI medium with 10% fetal bovine serum (FBS) and grown to 50% confluence at 37 ℃ in an atmosphere of 5% CO_2. Vero cells were incubated with 10^7 *T. cruzi* metacyclic-trypomastigotes or trypomastigotes for 12 h before wash twice with PBS and incubated in RPMI medium with 10% FBS at 37 ℃ and 5% CO_2. At 7-12 days post infection, *T. cruzi* trypomastigotes released into the supernatant of infected vero cells were collected by centrifugation at 200xg for 1 min to remove cellular debris and a second centrifugation at 3000xg for 10min to collect *T. cruzi* parasites in a pellet.

4.5.10 PLASMID CONSTRUCTION

The *T. cruzi* pTrex-n-tdTomato and pTrex-b-ts611/2 plasmids were constructed by subcloning the coding sequence of tdTomato [Genbank accession: AY678269], or ts611/2 (Buschiazzo, Campetella et al. 1997) respectively into multiple cloning site of the pTrex plasmid (Vazquez and Levin 1999) containing neomycin phosphotransferase gene (pTrex-n) or blasticidin-S deaminase gene (pTrex-b).

4.5.11 FLOW CYTOMETRY AND FACS

Flow cytometry and FACS were performed as described in Chapter 2

4.5.12. NONSPECIFIC PRIMER ANCHORED PCR

Genomic DNA was extracted and purified using the QIAamp DNA Blood Mini Kit (250) (QIAGEN) following the manufacturer's manual. The nonspecific primer anchored PCR reactions were carried out in 4 steps: (1) Assembly a 30 µl PCR mix consisting of 6.2 µl 5x GC buffer, 0.6 µl of 10 mM dNTP, 50 ng genomic DNA from *T. cruzi,* 0.3 µl of 10 µM random

primer (SAP1), and 1µl Phusion polymerase and water. The PCR mix was subjected to denaturation at 94 °C for 5 min followed by incubation at 30 °C for 3min, ramp to 65 °C at 0.1 °C/s, and incubate at 65 °C for 5 min. (2) Add 1.5 µl of 10mM GFP11 primer (F4) to the PCR mix and perform 10 cycles of 98 °C for 30s; 65 °C for 3min followed a 72 °C for 5 min. (3) To the PCR mix, add 2.5 µl of 10mM GFP11 primer (F3), 4 µl 5x GC buffer and 13.5 µl water. The PCR mix was subjected to an initial denaturation at 98 °C for 2 min followed by 20 cycles of 98 °C for 30s; 65 °C for 3min and 1 instance of a 72 °C for 5 min. (4) Assembly a 25 µl PCR mix consisting of 5 µl 5x GC buffer, 0.5 µl of 10 mM dNTP, 1 µl of 50x dilution of PCR product from (3), 1.25µl of 10 µM random primer (SAP2), and 0.5 µl Phusion polymerase and 15.5 µl water. The PCR mix was subjected to an initial denaturation at 98 °C for 2 min followed by 30 cycles of 98 °C for 2 min; 65 °C for 3 min and 1 instance of a 72 °C for 5 min. All PCR reagents (except water and genomic DNA) were from Phusion High-Fidelity PCR Kit (NEB). All thermocycling steps were performed using a Peltier thermocycler, model PTC-200. PCR Products were collected and purified with E.Z.N.A. Cycle Pure Kit (OMEGA bio-tek).

Primer sequences are listed as follows:

SAP1 (“NNNNCT”): 5'-
CTGAGACCTATTACCGCCCAGAGGATCTCCGTGTGCAATGCATNNNNCT

SAP1 (“NNNNNA”): 5'-
CCTGAGACCTATTACCGCCCAGAGGATCTCCGTGTGCAATGCATNNNNNA

SAP2: 5'-CCTGAGACCTATTACCGCCCAGAGGA

F3: 5'-ATGGTGCTTCATGAGTACGTG

F4: 5'-TCCGTGACCACATGGTGCTT

4.6 REFERENCES

Buschiazzo, A., O. Campetella and A. C. Frasch (1997). "*Trypanosoma rangeli* sialidase: cloning, expression and similarity to *T. cruzi trans*-sialidase." Glycobiology **7**(8): 1167-1173.

Cabantous, S., T. C. Terwilliger and G. S. Waldo (2005). "Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein." Nat Biotechnol **23**(1): 102-107.

Carvalho, S. T., M. Sola-Penna, I. A. Oliveira, S. Pita, A. S. Goncalves, B. C. Neves, F. R. Sousa, L. Freire-de-Lima, M. Kurogochi, H. Hinou, S. Nishimura, L. Mendonca-Previato, J. O. Previato and A. R. Todeschini (2010). "A new class of mechanism-based inhibitors for *Trypanosoma cruzi trans*-sialidase and their influence on parasite virulence." Glycobiology **20**(8): 1034-1045.

Contreras, V. T., J. M. Salles, N. Thomas, C. M. Morel and S. Goldenberg (1985). "In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions." Mol Biochem Parasitol **16**(3): 315-327.

Goldenberg, S., V. T. Contreras, M. C. Bonaldo, J. M. Salles, M. P. A. L. Franco, J. Lafaille, M. Gonzales-Perdomo, J. Linss and C. M. Morel (1987). "In vitro differentiating systems for the study of differential gene expression during *Trypanosoma cruzi* development." **v. 42**.

Horn, D. (2014). "Antigenic variation in African trypanosomes." Mol Biochem Parasitol **195**(2): 123-129.

Jackson, A. P., A. Berry, M. Aslett, H. C. Allison, P. Burton, J. Vavrova-Anderson, R. Brown, H. Browne, N. Corton, H. Hauser, J. Gamble, R. Gilderthorp, L. Marcello, J. McQuillan, T. D. Otto, M. A. Quail, M. J. Sanders, A. van Tonder, M. L. Ginger, M. C. Field, J. D. Barry, C. Hertz-Fowler and M. Berriman (2012). "Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species." Proc Natl Acad Sci U S A **109**(9): 3416-3421.

Martin, D. L., D. B. Weatherly, S. A. Laucella, M. A. Cabinian, M. T. Crim, S. Sullivan, M. Heiges, S. H. Craven, C. S. Rosenberg, M. H. Collins, A. Sette, M. Postan and R. L. Tarleton (2006). "CD8+ T-Cell responses to *Trypanosoma cruzi* are highly focused on strain-variant *trans*-sialidase epitopes." PLoS Pathog **2**(8): e77.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953). "Equation of State Calculations by Fast Computing Machines." Journal of Chemical Physics **21**(6): 1087-1092.

Mugnier, M. R., G. A. Cross and F. N. Papavasiliou (2015). "The in vivo dynamics of antigenic variation in Trypanosoma brucei." Science **347**(6229): 1470-1473.

Peng, D., S. P. Kurup, P. Y. Yao, T. A. Minning and R. L. Tarleton (2015). "CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*." MBio **6**(1): e02097-02014.

Piras, M. M., D. Henriquez and R. Piras (1987). "The effect of fetuin and other sialoglycoproteins on the in vitro penetration of *Trypanosoma cruzi* trypomastigotes into fibroblastic cells." Mol Biochem Parasitol **22**(2-3): 135-143.

Rubin-de-Celis, S. S., H. Uemura, N. Yoshida and S. Schenkman (2006). "Expression of trypomastigote *trans*-sialidase in metacyclic forms of *Trypanosoma cruzi* increases parasite escape from its parasitophorous vacuole." Cell Microbiol **8**(12): 1888-1898.

Schenkman, R. P., F. Vandekerckhove and S. Schenkman (1993). "Mammalian cell sialic acid enhances invasion by *Trypanosoma cruzi*." Infect Immun **61**(3): 898-902.

Soares Medeiros, L. C., L. South, D. Peng, J. M. Bustamante, W. Wang, M. Bunkofske, N. Perumal, F. Sanchez-Valdez and R. L. Tarleton (2017). "Rapid, Selection-Free, High-Efficiency Genome Editing in Protozoan Parasites Using CRISPR-Cas9 Ribonucleoproteins." mBio **8**(6).

Vazquez, M. P. and M. J. Levin (1999). "Functional analysis of the intergenic regions of TcP2beta gene loci allowed the construction of an improved *Trypanosoma cruzi* expression vector." Gene **239**(2): 217-225.

Weatherly, D. B., C. Boehlke and R. L. Tarleton (2009). "Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome." BMC Genomics **10**: 255.

Weatherly, D. B., D. Peng and R. L. Tarleton (2016). "Recombination-driven generation of the largest pathogen repository of antigen variants in the protozoan *Trypanosoma cruzi*." BMC Genomics **17**(1): 729.

Zhang, Y., D. Q. Liu, W. X. Yang, Y. N. Li and H. F. Yan (2011). "A Nonspecific Primer Anchored PCR Technique for Chromosome Walking." Brazilian Archives of Biology and Technology **54**(1): 99-106.

4.7 FIGURE AND TABLES

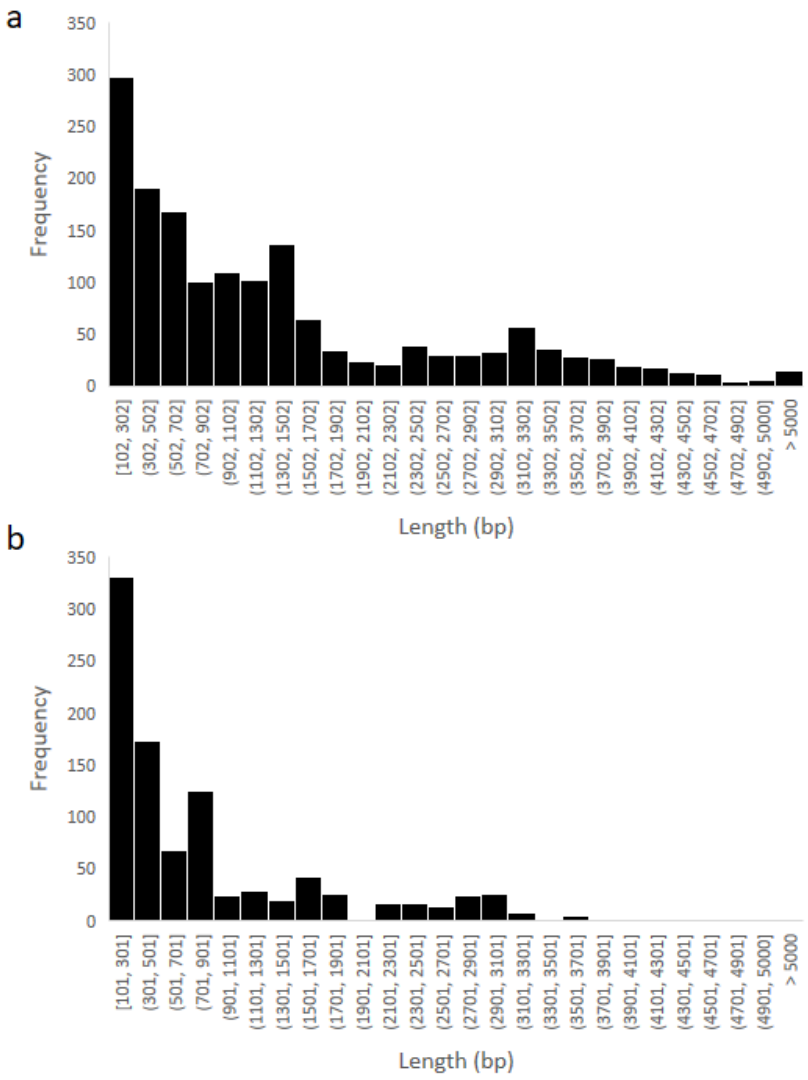


Figure 4.1. Length distribution of TcTS genes (a) being targeted and (b) not targeted by the gRNA set.

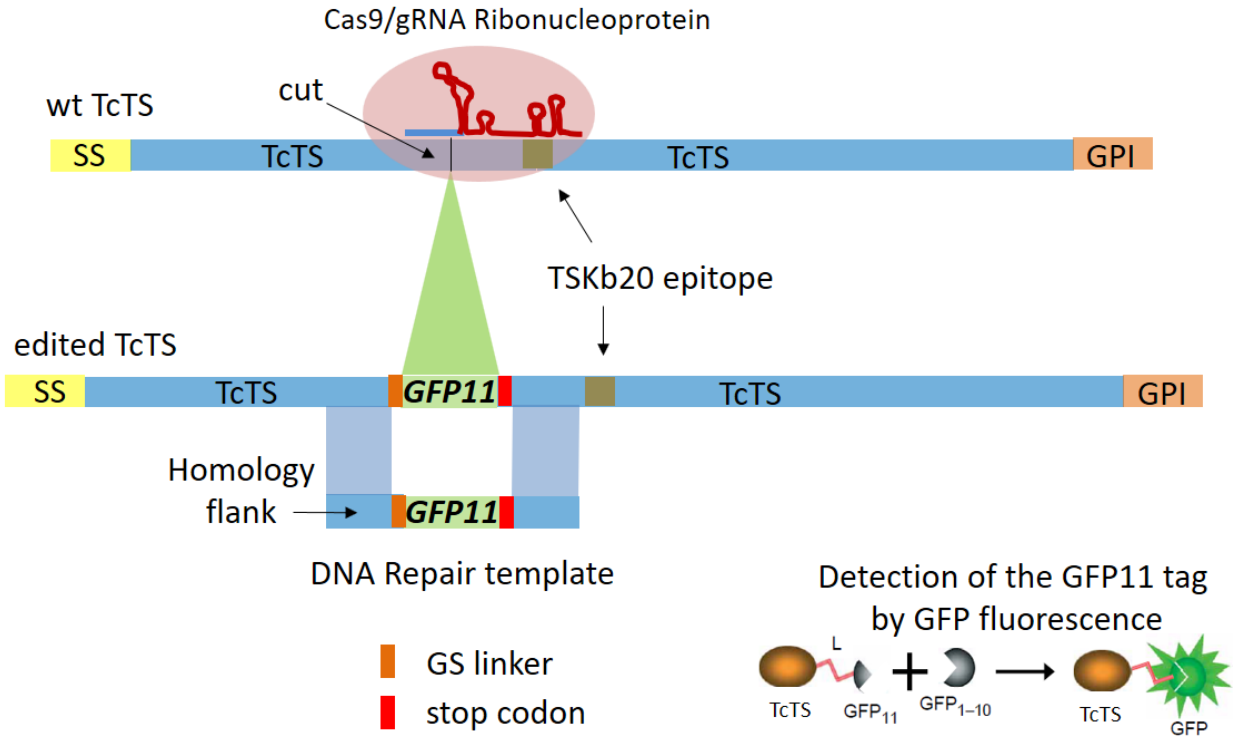


Figure 4.2. Schematic diagram of a full length TcTS gene (with signal sequence, SS, and GPI anchor site, GPI) being edited by the Cas9/gRNA ribonucleoprotein complex, and the expected editing outcome with insertion of GS linker, GFP11 and a stop codon.

Schematic diagram of detection of TcTS-linker-GFP11 fusion protein by complementation with GFP1-10

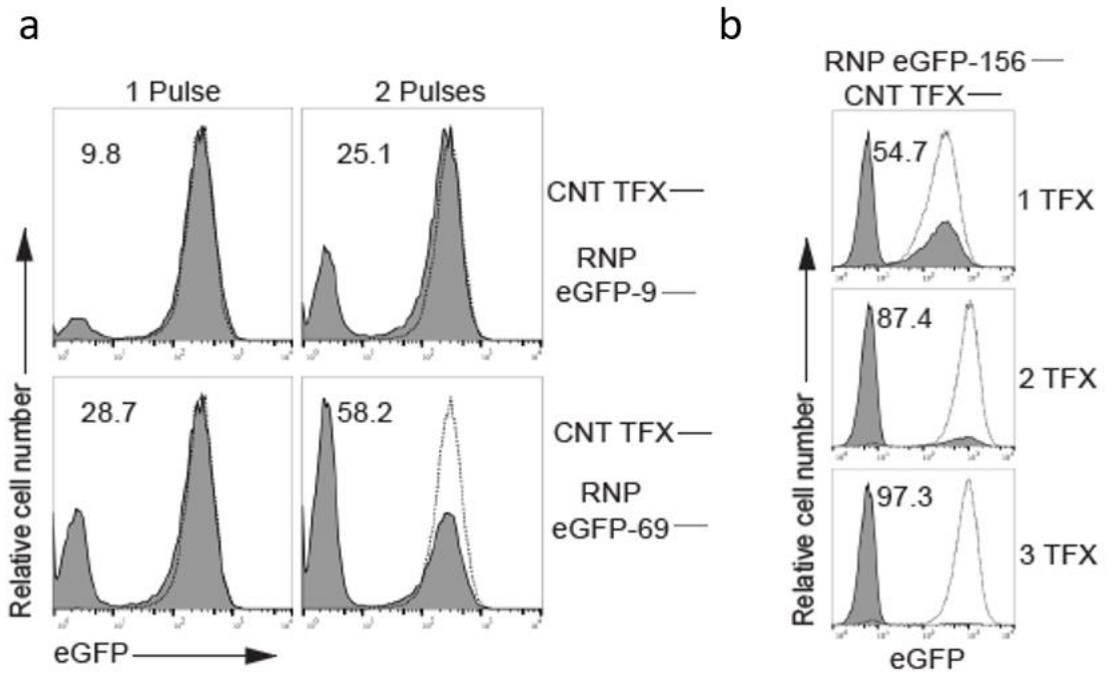


Figure 4.3. (a) Epimastigotes from *T. cruzi* expressing eGFP were electroporated with water (dashed line) or RNP complexes containing SaCas9 and a sgRNA GFP-9 or GFP-69 targeting eGFP (gray filled). Parasites received 1 or 2 pulses of electroporation (the 2 pulses were applied consecutively). (b) Epimastigotes from *T. cruzi* expressing eGFP were electroporated with water (dashed line) or RNP complexes containing SaCas9 and a sgRNA targeting eGFP (eGFP-156; gray filled). Parasites received 1, 2 or 3 electroporations with 2-day intervals between each. Loss of eGFP or tdTomato fluorescence was determined seven days post-first transfection. This figure is taken from (Soares Medeiros, South et al. 2017).

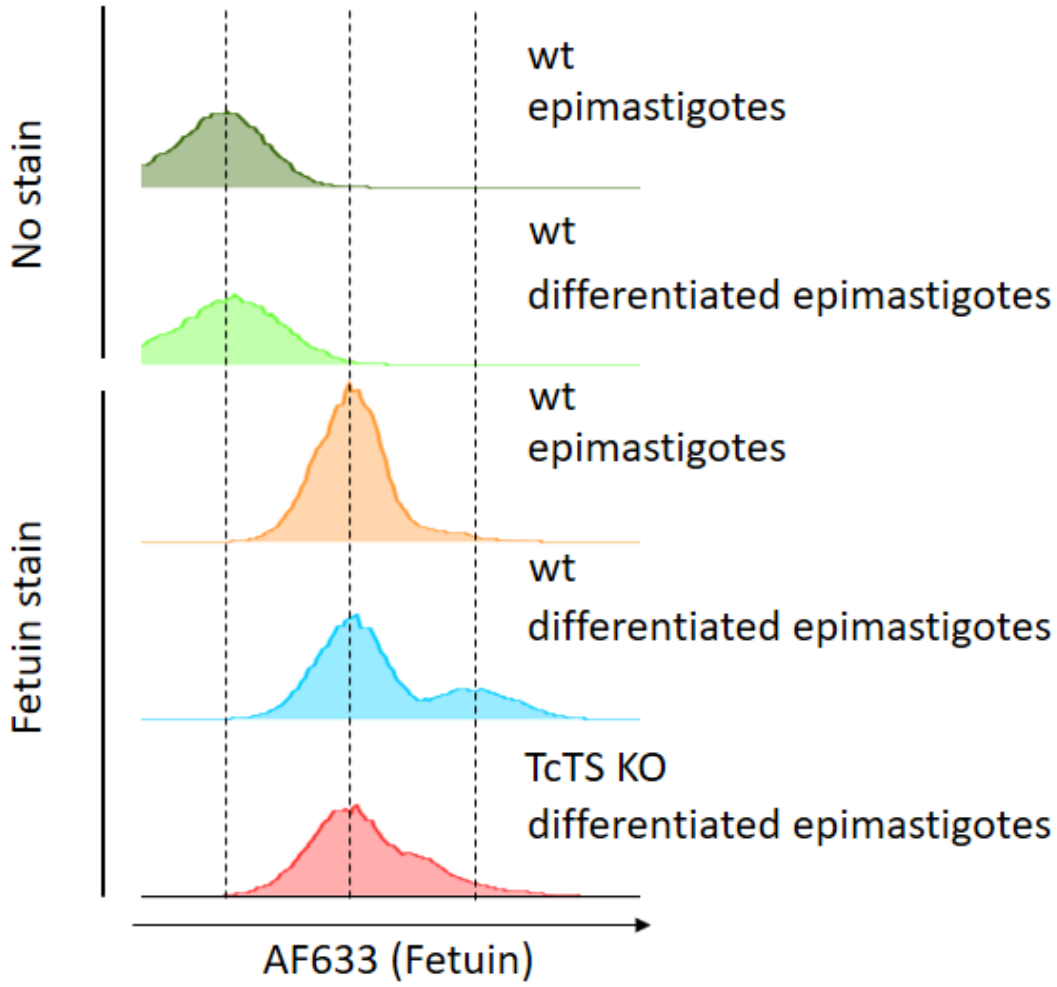


Figure 4.4. Flow cytometry analysis of surface transialidase protein level using fetuin staining. Differentiated *T. cruzi* epimastigotes and undifferentiated epimastigotes were stained with Fetuin-AF633 conjugate.

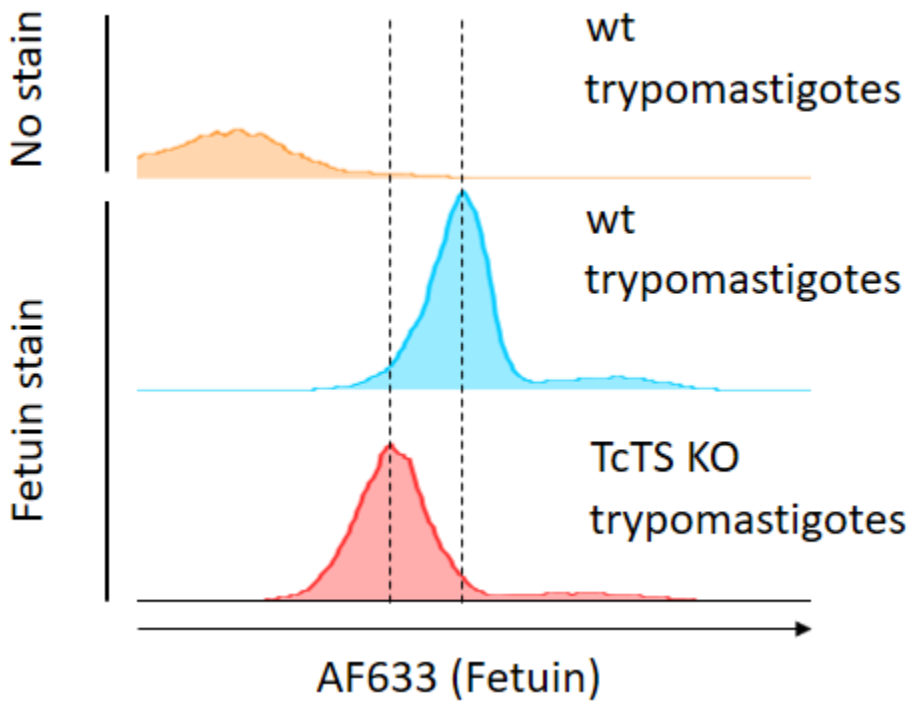


Figure 4.5. Flow cytometry analysis of *T. cruzi* trypomastigotes parasites stained with Fetuin-AF633 conjugate

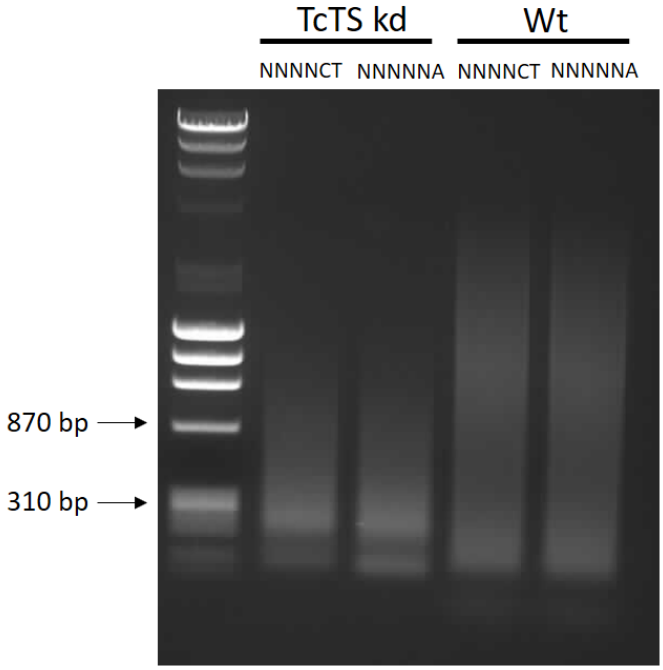


Figure 4.6. PCR amplification with GFP11 and non-specific anchoring primers using genome DNA from TcTS knockdown and wildtype parasites. “NNNNCT” and “NNNNNA” are non-specific anchoring sequences.

Table 4.1. gRNAs used in the knockdown of TcTS.

Epitope group	gRNA target sequence (including PAM)	Name	Target number (TcTS with epitope)	Target number (All TcTS)
Tskb20	GTATTGTCCGTCAGCCAAAGATGGAGT	14_2	286	299
	GTCGATGGACACCGTCGCCACAAGAGT	14_8	217	466
	ATCTCCCACTTCTACATTGGAGGGGAT	14_9	196	73
	GCCGTCGTCATGAGCGTGTCGGTGGGT	14_10	181	776
	CACACAACCGTCGCCGCGTGACCGGAT	14_11	162	819
	TGTTTGTTGCCCCACACGCGCGAGAGT	14_7	119	350
	GTGGATGGACACCGTCGCCACAAGAGT	14_4	105	282
	TGTTGCTGCACTCTTCCTGTGAGGGGT	14_12	85	717
	CTTCCTTCCCTCCTCCTTCCGCTGGAT	14_5	79	769
	ATCTCACACTTCTACATTGGAGGGGAT	14_6	73	78
	GGAAGGAGGAGGGAAGGAGGAGAGAGT	14_1	58	799
	GTCAGAAGCACAGAGACCATACCGGGT	14_14	37	125
	CACACAACCGTCGCCGCGTGACTGGAT	14_13	9	742
	CTTGTGGCTGACGGACATGCAGCGAAT	14_3	6	21
Tskb18	GTATTGTCCGTCAGCCAAAGATGGAGT	7_1	12	108
	CACATGGAGGGACGAGTACCTCGGGGT	7_2	10	322
	GTCTTTCTGTACAACCGCCCACTGAGT	7_7	9	45
	GCCGTCGTCATGAGCGTGTCGGTGGGT	7_3	9	784
	AGTGCTCCCCTTCGCAACCTGGTGGAT	7_6	8	30
	agcccaatgaagagcgtgctgatgggt	7_5	8	12
	GGAAGAGCTCGGCGCAGCGAAGGGGAT	7_4	5	19

Table 4.2. Deep sequencing of disruption sites in TcTS knockdown clones.

Clone #	Number of TcTS gene with evidence of intended disruption	% of targetable TcTS
53	283	21.23
109	301	22.58
10	303	22.73
61	314	23.56
107	318	23.86
96	340	25.51
40	342	25.66
52	405	30.38
3	427	32.03
32	455	34.13
1	524	39.31
6	572	42.91
42	613	45.99
8	644	48.31
30	691	51.84
15	742	55.66
73	790	59.26

CHAPTER 5

CONCLUSION AND FUTURE WORK

Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, affects 6-7 million people worldwide. If not treated promptly in the acute infection phase, most infections enter a chronic phase in which progressive and debilitating conditions develop over decades. The anti-*T. cruzi* CD8+ T response produced by the host in response to this intracellular parasite, is critical to controlling the infection but is often insufficient to completely clear it. The substantial delay in the onset and development of the anti-*T. cruzi* CD8+ T cell response suggests an act of immune evasion. *T. cruzi* expresses a surface-anchored enzyme, *trans*-sialidase (TcTS), encoded by a gene family consisting thousands of members. It is hypothesized that the TcTS gene family repertoire plays a role in the immune evasion process. One way to address this hypothesis is to assess the impact of knockout of a substantial number of these TcTS genes.

In this thesis work, I aimed to build the foundation to address this problem by first developing the tools that would allow me to knock-down of the expression of TcTS genes in *T. cruzi*. In the absence of a functional RNAi system or other genetic tools to knockout thousands of genes in *T. cruzi*, I adapted the CRISPR/Cas9 system to conduct large-scale genome editing in *T. cruzi*.

In the first part of this work, I achieved, for single genes, >60% gene knockout efficiency by electroporating gRNAs into a *T. cruzi* line stably expressing SpCas9, or >90% efficiency when SaCas9/gRNA ribonucleoprotein complexes are introduced into wild-type lines. Next, I demonstrated precise gene-editing using DNA donor templates with the CRISPR/Cas9 system.

Leveraging the multiplexity of the CRISPR/Cas9 system, I knocked-down a gene family of 65 members using 3 gRNAs, as a proof of principle of gene family knockdown in *T. cruzi*.

In the second part of this work, I developed a computation tool, EuPaGDT: Eukaryotic pathogen gRNA design tool, to aid in the design of gRNA(s) targeting over 2000 TcTS genes. EuPaGDT can enumerate all gRNAs found in a collection of genes and characterize each gRNA. A key step in the characterization step is the search for genome-wide on-targets, which allows identification of gRNA(s) that can target a number of genes – 100's in the case of the TcTS gene family. Using EuPaGDT, I designed a set of 21 gRNAs, when used in combination, can target 63% of all TcTS genes and 92% of the TcTS genes capable of generating protein products.

In the third part of the work, I sequentially delivered each of the 21 gRNAs in complex with Cas9 along with a corresponding DNA repair template to make precise edits to disrupt the coding capability of TcTS genes. Metacyclic and culture-derived trypomastigotes from the knockdown line showed a decreased surface *trans*-sialidase protein level compared to the wild-type controls and the intended sequence editing could be detected in the genomic DNA of the TcTS knockdown parasites.

Future work will analyze TcTS expression level in clonal lines derived from the TcTS knockdown population. Parasite surface TcTS protein level will be assayed by fetuin staining and global TcTS transcripts level will be analyzed by RNAseq. Deep sequencing of the genomic TcTS disruption sites is expected to provide additional evidence of TcTS knockdown. Assessing the ability of clonal lines to elicit TcTS epitope-specific CD8+ T cell responses in mice infections will be a third method to assess TcTS knockdown efficiency in these lines. Results from all 3 methods will be cross-referenced to select a clonal *T. cruzi* line with the lowest expression of TcTS genes. The ability to establish chronic infections in mice will be analyzed for

this clonal line, as a first step to test the hypothesis that the repertoire of TcTS gene family plays a role in the immune evasion process.

This thesis work has a few innovations. I pioneered in adapting the CRISPR/Cas9 system for efficient genome editing in *T. cruzi*. Instead of using plasmids to deliver all components of the CRISPR/Cas9 system and then drug-selecting a target population of genome-edited parasites, I transfected sgRNAs or Cas9/sgRNA ribonucleoprotein complexes into *T. cruzi*, generating a parasite population with 60% knockouts in as little as 4 days. To aid the design of gRNAs targeting thousands of TcTS genes, I created a versatile gRNA design tool for eukaryotic pathogens. This tool can design gRNAs for large gene families, whole genomes, and can design gRNAs for gene-tagging experiments. To date, this tool has processed over 10,000 gRNA-design requests from the eukaryotic pathogen research community.

This work demonstrated the feasibility of knocking-down gene families in *T. cruzi*. I created TcTS gene family knockdown parasites to test the longstanding hypothesis that the TcTS gene family repertoire plays a role in the immune evasion process.