Imperial College London

Doctoral Thesis

Cellular Barcoding of Protozoan Pathogens for Within-Host Population Dynamics and *in vivo* Drug Discovery

---

Ceire Joanne Wincott

A thesis submitted for the degree of Doctor of Philosophy

Department of Life Sciences

Imperial College London

June 2023

Statement of Originality

I, Ceire Wincott, declare that the work presented in this thesis entitled "Cellular barcoding of protozoan pathogens for within-host population dynamics and *in vivo* drug discovery" is my own. I confirm that information taken from other sources has been acknowledged as appropriate, and any work carried out by others has been clarified in the text, corresponding figure legends, and or below. Figures were made using Adobe Illustrator® and BioRender®.

Section 4.2.1 – Experiments were conducted in the Child Lab before my joining by Gaya Sritharan and Matthew Child.

Section 4.3 – Transfections in *T. brucei* were carried out by the Tiengwe Lab and in *C. parvum* were conducted in Sateriale Lab.

Section 5 – *In vivo* experiments were conducted in the Ewald Lab at the University of Virginia. This work has been reported in Cell Reports Methods.

Section 4 & 5 – Some work has been adapted from that published in Wincott et al.[1] appended to this thesis.

Section 5 – Parasite replication assay was conducted by Rebecca Oxtoby, masters student, under my direct supervision.

Copyright Declaration

IMPERIAL COLLEGE LONDON

Abstract

Faculty of Natural Sciences

Department of Life Sciences

Doctor of Philosophy

Cellular Barcoding of Protozoan Pathogens for Within-Host Population Dynamics and *in vivo* Drug Discovery

By Ceire WINCOTT

The obligate intracellular apicomplexan parasite *Toxoplasma gondii* has broad infectious ability causing disease in humans and animals, some of which can be fatal. Existing treatments for *T. gondii* infections have notable side effects, and the emergence of resistance to first-line therapies is a growing concern. Understanding the fundamental aspects of *T. gondii* biology necessitates studying *in vivo* host-pathogen interactions. However, tracking parasite populations without artificially influencing infection dynamics has posed significant challenges. To address this, we propose a cellular barcoding technique combined with Next Generation Sequencing (NGS) technology to genetically identify and assess the representation of parasite populations. This approach can be applied not only to *T. gondii* but also to *T. brucei* and holds potential for future application to other pathogens.

Using our cellular barcoding methodology, we conducted population dynamics studies to investigate *T. gondii* colonisation of the brain parenchyma. Surprisingly, we discovered that the blood-brain barrier (BBB) allows relatively unrestricted traversal by *T. gondii*, imposing a less stringent bottleneck than anticipated. Moreover, we observed the dynamic nature of chronic *T. gondii* infection, as brain cyst numbers continued to decrease over several months.

Furthermore, we employed the cellular barcoding methodology to facilitate multiplexed *in vivo* drug screening. Through this approach, we successfully identified small molecule fragments with anti-parasitic effects. Our proof-of-concept data supports the use of this screening platform for iterative drug molecule development. Additionally, in concurrent studies, one of the identified hit fragments exhibited selective inhibition of translation in *T. gondii* compared to HEK293 cells, prompting further characterisation efforts.

Acknowledgements

Abbreviations

BBB – Blood brain barrier

GPI – glycosylphosphatidylinositol

MIC - microneme surface-exposed transmembrane adhesins

PV – Parasitophorous vacuole

MJ – Moving junction

RON – Rhoptry neck

ROP – Rhoptry bulb proteins

PVM – Parasitophorous vacuole membrane

HIV - Human immunodeficiency virus

AIDS – Acquired immune deficiency syndrome

DHFR – Dihydrofolate reductase

DNA - Deoxyribonucleic acid

gDNA – Genomic DNA

R&D – Research & development

HTS – High-throughput screens

AI – Artificial intelligence

DOS – Diversity orientated synthesis

MoA – Mechanism of Action

CRISPR – Clustered regularly interspaced short palindromic repeats

RNA – Ribonucleic acid

sgRNA – Single-guide RNA

DSB – Double-stranded break

PAM – Protospacer adjacent motif

NHEJ – Non-homologous end-joining

HDR – Homology directed repair

PCR – Polymerase chain reaction

NEB – New England Biolabs®

NGS – Next generation sequencing

UPRT – Uracil phosphoribosyl-transferase

PE – Paired ends

DMEM – Dulbecco's Modified Eagle Medium

DMSO – Dimethyl sulfoxide

FUDR – 5'-fluro-2'-deoxyuridine

ORF – Open reading frame

HFF – Human foreskin fibroblasts

SD – Standard deviation

PBS – Phosphate-buffered saline

AAALAC – Association for Assessment and Accreditation of Laboratory Animal Care

IACUC – Institutional animal care and use committee

IP – Intraperitoneal

FBS – Foetal bovine serum

FITC – Fluorescein isothiocyanate

PEC – Peritoneal exudate cells

WITS – Wild-type isogenic tagged strains

MRC – Medical research council

STAMP – Sequence tag–based analysis of microbial populations

STAMPR – Refined Sequence tag–based analysis of microbial populations

CNS – Central nervous system

PDD – Phenotypic drug discovery

PFA – Paraformaldehyde

IAA – Iodoacetamide

MTT - 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

EGFR - Epidermal growth factor receptor

BTK – Bruton's tyrosine kinase

PDAC – Pancreatic ductal adenocarcinoma

CHX – Cycloheximide

HEK 293 – Human embryonic kidney 293

Figures

Publications

Benns, Henry J. Wincott, Ceire J. et al. (2021). ' Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery'. In: *Curr Opin Chem Biol*, 60, p. 20-29.

Alves, Eduardo. et al. (2021). 'An Extracellular Redox Signal Triggers Calcium Release and Impacts the Asexual Development of *Toxoplasma gondii*'. In: *Front Cell Infect Microbiol*, 11.

Wincott, Ceire J. et al. (2022). 'Cellular barcoding of protozoan pathogens reveals the within-host population dynamics of *Toxoplasma gondii* host colonization'. In: *Cell Rep Meth*, 2.

Benns, Henry J. et al. (2022). 'CRISPR-based oligo recombineering prioritizes apicomplexan cysteines for drug discovery'. In: *Nat Micro*, 7, p. 1891-1905.

Presentations

ToxoUK Online Symposium, December 2020: 3-minute Flash Talk – Cellular Barcoding of *Toxoplasma* gondii Reveals Unexpectedly Permissive Host Brain Colonisation

Imperial College London Department of Life Sciences Postgraduate Symposium, August 2021: Poster Presentation - Molecular Barcoding for Investigation of Within-host Population Genetics & Drug Discovery (First Place Prize)

Molecular Parasitology Meeting, October 2021: Flash Talk - Cellular Barcoding of *Toxoplasma gondii* Reveals Unexpectedly Permissive Host Brain Colonisation and a Stringent Bottleneck in the Gut

ToxoUK University of Cambridge, December 2021: Flash Talk - Molecular Barcoding of *Toxoplasma gondii* for Multiplexed *in vivo* Drug Screening

RSC Chemical Biology Symposium, May 2022: Poster Presentation - Molecular Barcoding to Enable Multiplexed *in vivo* Antiparasitic Drug Discovery

EMBO Workshop New frontiers in host-parasite interactions, from cell to organism, October 2022: Poster Presentation - Molecular Barcoding of *Toxoplasma gondii* for Multiplexed *in vivo* Drug Screening

Contents

1. Introduction

1.1. *Toxoplasma gondii*, an Apicomplexan parasite

1.1.1. Apicomplexa, a phylum

Apicomplexa are a group of parasitic, single-cellular eukaryotes which infect a vast range of host organisms. This phylum includes several parasites important for both human and veterinary health including *Plasmodium* species responsible for malaria, and those of the *Cryptosporidium* genus responsible for severe diarrheal disease[2]. Malaria is a mosquito borne disease with an estimated 200 million cases and 600,000 deaths annually. Approximately half of the world's population is considered at risk of malaria[3]. *Plasmodium falciparum* is considered the most important of the six plasmodial species which present a threat to human health[4]. *Cryptosporidium* parasites, particularly *Cryptosporidium parvum* (*C. parvum*) and *Cryptosporidium hominis* (*C. hominis*), pose significant global health challenges. They are the causative agents of cryptosporidiosis which presents as severe diarrheal disease and is particularly problematic in areas with limited waste treatment and water sanitising facilities. In immunocompromised individuals and children this disease can often prove fatal[5].

Apicomplexan parasites are also of veterinary importance, for example cryptosporidiosis in livestock such as cattle can be fatal and rapid spread through herds can have devastating financial consequences for farmers[6]. Additionally, the *Neospora* genus of Apicomplexa causes spontaneous abortion in cattle and can lead to paralysis in dogs[7]. Parasites of the *Eimeria* genus are responsible for coccidiosis in livestock such as chickens and other poultry. It is often fatal in younger or immunocompromised animals and can also carry a significant financial burden due to the speed of spread through a population and the cost of treating outbreaks[8].

Toxoplasma gondii (*T. gondii*), is an obligate intracellular parasite of the Apicomplexa phylum. In 1908 Nicolle and Manceaux observed a protozoan in the tissues of a gundi rodent (*Ctenodactylus gundi*) whilst studying leishmaniasis. On investigation it was confirmed as a new organism and named *Toxoplasma gondii*, *toxo* meaning arc or bow, *plasma* meaning life. Splendore also observed this organism in 1908 in a rabbit however did not name it[9]. This protozoan was observed in a range of hosts over the next 30 years however a viable isolate was not obtained until 1937 by Sabin and Olitsky in New York[10]. *T. gondii* was not realised as a human pathogen until 1939 when a child became ill and died at 31 days old with wide-spread encephalomyelitis, later diagnosed as congenital toxoplasmosis. Tissue from the child was used to inoculate mice, rabbits, and rats all of which developed symptomatic toxoplasmosis[11]. Acquired toxoplasmosis was first recorded in 1941 when a six-year-old boy developed

headaches and convulsions, dying after 30 days of illness. On inoculation of mice with cerebral cortex homogenate, *T. gondii* was identified[12].

T. gondii is perhaps the most successful parasite of the apicomplexan phylum due to the ability to infect most nucleated mammalian cells. It is also widely distributed amongst humans with estimates that 30-50% of the global population are infected. Ingestion of contaminated meat is the most common cause of infection. Other transmission routes include faecal-oral, through litter trays of domestic animals or contamination of water supplies, and the previously discussed vertical transmission in pregnant women, from mother to foetus[13].

1.1.2. The life cycle of *T. gondii*

T. gondii has two distinct life cycle phases, which require a transition between the sexually replicating and asexually replicating form of the parasite.

Despite *T. gondii's* broad infectious ability, members of the Felidae family are the only known definitive host[14]. High concentrations of linoleic acid in the feline intestine are conducive to parasite sexual development[15]. Upon ingestion of a bradyzoite cyst, a combination of proteolytic enzymes and stomach acid rupture the cyst wall releasing bradyzoites into the digestive tract of the host. These bradyzoites invade enterocytes, and following an indeterminate number of asexual replication cycles they differentiate into either multiple micro- or a single macrogametocytes. It is unknown what determines whether a cell will become a
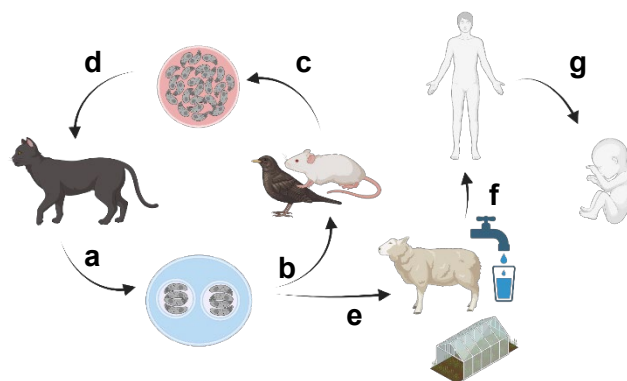
Figure 1 | The life cycle of *Toxoplasma gondii*. (a) *T. gondii* undergoes sexual reproduction, generating oocysts, in the digestive tract of felids, the definitive host, and these oocysts are shed in faeces. (b) After being shed in faeces, *T. gondii* oocysts can infect animals like mice and birds in the environment. During the acute stage of infection, the parasite resides in a variety of tissues as the rapidly replicating tachyzoite form. (c) *T. gondii* converts to the encysted bradyzoite form which is found primarily in brain, skeletal muscle, cardiac, and ocular tissues. (d) When felids ingest prey animals chronically infected with *T. gondii*, the bradyzoite cysts rupture, allowing the cycle to begin again. (e) Oocysts in the environment can also contaminate livestock, crops, and water supplies. (f) *T. gondii* can infect human hosts through these routes, and although toxoplasmosis is usually asymptomatic and controlled by the host's immune response, it can be serious and sometimes fatal in immunocompromised individuals. (g) A *T. gondii* infection in immune naive pregnant individuals can traverse the placenta. This congenital transmission can result in miscarriage or foetal deformity.

micro- or macrogametocyte or what the specific trigger for differentiation is[16]. Fertilised by a macrogamete forms an oocyst which is released into the intestine and shed in faeces. On ejection from the protected internal environment and exposure to the highly oxygenated exterior the oocysts sporulate (Figure 1a). These sporulated oocysts consist of 4 sporozoites protected by a dense cyst wall which persist and remain infectious for up to 18 months[17,18].

Mature oocysts infect intermediate hosts through ingestion of infected tissue (carnivorism) and contaminated water or food sources (faecal-oral transmission) (Figure 1b). On ingestion by an intermediate host, sporozoites excyst in the digestive tract and differentiate to tachyzoites. Tachyzoites are a rapidly replicating asexual form of the parasite, which multiply by endodyogeny in parasitophorous vacuoles within-host cells. Tachyzoites distribute themselves throughout the host via the vasculature and are responsible for the acute infection in humans and other animals. The acute infection is responsible for the clinical disease, toxoplasmosis. The majority of tachyzoites are thought to be cleared by the host immune response within two to three weeks of the initial infection[16]. Some will convert to bradyzoites, a sedentary immune-evasive encysted parasite form which typically resides in the striated muscle tissue, including the heart, and the central nervous system (Figure 1c). This chronic infection persists for the natural life of the host. Ingestion of these tissue cysts through predation by another intermediate or definitive host leads to the recommencing of the asexual or sexual life cycle, respectively (Figure 1d). While spontaneous reconversion of bradyzoites to tachyzoites has been documented, in immunocompetent hosts these parasites are thought to be rapidly cleared by the host immune system. However, if a host is immunocompromised this can lead to recurrence of the acute infection, toxoplasmosis, which may be fatal[19].

1.1.3. *T. gondii* Traversal of the Blood-brain Barrier and Brain Colonisation

The brain is a frequently colonised niche in the *T. gondii* chronic infection. The blood-brain barrier (BBB) was, until recently, thought to be highly selective and impermeable to most cells and organisms. Three mechanisms have been postulated for the traversal of *T. gondii* across the BBB[20]. The first is paracellular entry, the direct migration of *T. gondii* through junctions in the endothelial cell layer (Figure 2a). This proposition is supported by *T. gondii*'s ability to actively move by 'gliding motility', propelled by an actin-myosin motor and the glideosome macromolecular complex[21,22]. *T. gondii* tachyzoites have been shown to cross polarised cell-monolayers *ex-vivo* leading to speculation that this gliding motility may be the basis for BBB traversal[23].

The second proposed method is transcellular migration; extracellular parasites could infect the endothelial cells of the BBB and egress on the basolateral side (Figure 2b). This method of parasite transport has been documented for parasites crossing the gut epithelium,
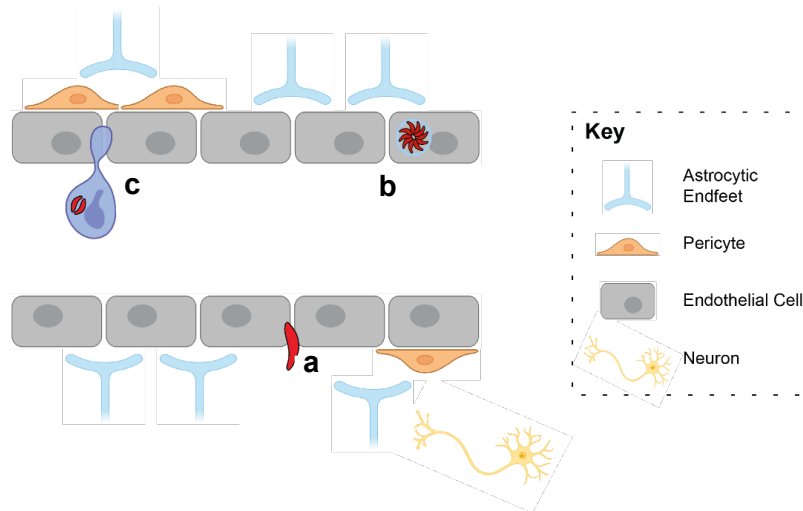
Figure 2 | Three proposed mechanisms for how *Toxoplasma gondii* crosses the blood-brain barrier. (a) The first proposed route of *T. gondii* traversal of the BBB is direct migration. This suggests that *T. gondii* can move directly through junctions in the endothelial cell layer. This theory is supported by *T. gondii*'s recognised ability for 'gliding motility'. (b) Second proposed route is transcellular migration. *T. gondii* may infect BBB endothelial cells and then egress on the basolateral side. Parasites are known to traverse the gut epithelium using a similar method. (c) The final proposed route for traversal of the BBB is the 'Trojan Horse' model. Parasites invade immune cells which are capable of passing the BBB.

and recent studies have shown that *T. gondii* can infect many types of endothelial cells including those of the BBB. Subsequent *in vitro* studies demonstrated that parasites could adhere to, invade, and egress from BBB endothelial cells, from which it would be possible to infect the brain parenchyma[24].

The final possible mechanism for *T. gondii* brain colonisation is the 'Trojan Horse' model (Figure 2c). In this model, parasites invade immune cells, which can traverse the BBB, and then egress once the transition is complete. *In vitro* studies have confirmed that infected immune cells are capable of traversing endothelial barriers and have increased motility[25]. Despite confirmation that *T. gondii* can effectively traverse the BBB it is still expected to impose a significant bottleneck on colonisation of the brain parenchyma.

1.1.4. The lytic cycle of T. gondii

T. gondii is an obligate intracellular parasite, and so mechanisms for attachment and invasion of host cells are fundamental to survival in the host environment. In intermediate hosts the lytic cycle of the tachyzoite parasite form is central to the parasite's pathogenicity. This cycle can be broken down in to five phases: (1) host cell attachment, (2) invasion, (3) parasitophorous vacuole (PV) formation, (4) replication and (5) egress[26].

The plasma membrane of *T. gondii* is covered in glycosylphosphatidylinositol (GPI)-anchored surface antigens known as SAGs. These reversibly interact with host cell surface glycans to facilitate initial contact[27]. Once this contact is instigated the parasite orientates so
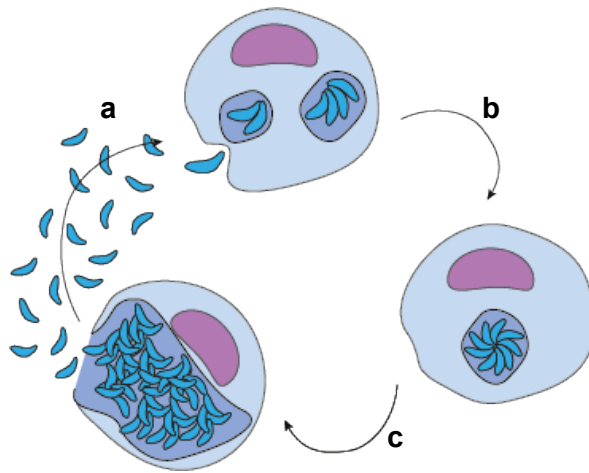
Figure 3 | The lytic cycle of *Toxoplasma gondii*. (a) Free tachyzoite invade nucleated host cells. Parasites attach to the host cell membrane, invade, and form the parasitophorous vacuole during the process of invasion. Within this vacuole parasites are protected from the host cell environment. The vacuole membrane has been proposed to act a molecular sieve, ensuring access to the necessary nutrients. (b) Within the parasitophorous vacuole *T. gondii* replicates asexually via endodyogeny. One replication cycle takes approximately 6-8 hours. (c) Once the host cell cannot support further replication the parasites exit in a process known as egress. Egress results in the lytic destruction of the host cell.

the apical end is in contact with the host cell membrane to engage in a higher-affinity interaction. This is mediated by microneme surface-exposed transmembrane adhesins (MIC) protein secretion followed by moving junction (MJ) formation, a joint structure between microneme protein apical membrane antigens and proteins residing in the rhoptry neck (RON proteins). The MJ acts as a region of close apposition between host and parasite. Subsequently rhoptry bulb proteins (ROPs) are secreted into host cell cytoplasm, detection of ROPs acts as further confirmation that the host cell membrane has been breached[28]. Once the MJ has been established, the parasites press forwards, invaginating the plasma membrane and forming the PV (Figure 3a)[27,29]. The PV provides a replicatory niche for the parasite with the PV membrane (PVM) acting as a molecular sieve ensuring *T. gondii* has access to the necessary nutrients[30]. Parasites replicate asexually via endodyogeny within the PV with one cycle taking approximately 6-8 hours (Figure 3b). Once the host cell can no longer support replication, the parasites rapidly exit the cell in a process known as egress (Figure 3c). Egress is the total lysis of the infected cell releasing all tachyzoites within[31]. It is this destruction of host tissue which characterises the pathology of toxoplasmosis.

1.1.5. Toxoplasmosis

Infection by *T. gondii* is prevalent worldwide in both humans and other animals. Serological surveillance, testing for anti-*Toxoplasma* IgG antibodies, suggests that one third of the global human population have been exposed to *T. gondii* with a proportion of these expected to remain chronically infected[32].

In healthy individuals *T. gondii* infection is typically asymptomatic with a small number of individuals experiencing mild, flu-like, symptoms[33]. In these instances, the immune system develops a protective response against *T. gondii*. An individual may then develop a latent, chronic infection, protected from any reactivation of acute infection by the developed immunity. However, in immunocompromised individuals such as those suffering from HIV/AIDS or undergoing chemotherapy treatment, recrudescence of the acute infection can be life-threatening[33–35]. Severe toxoplasmosis manifests as inflammation of infected tissues, primarily striated muscle (skeletal and cardiac), brain, lung and the eye. These symptoms are debilitating and often fatal if left untreated. *T. gondii* infection was the leading cause of brain lesions during the 1980's AIDs pandemic and is still considered on the most prolific opportunistic pathogens in those with HIV[35]. Another significant health risk associated with toxoplasmosis is congenital transmission. If a *T. gondii* infection naive individual is exposed to the parasite whilst pregnant, then tachyzoites can traverse the placenta and lead to spontaneous abortion or severe toxoplasmosis in the infant. Congenital infection-associated pathologies include ocular toxoplasmosis and severe neurological impairment[36,37].

Further to *T. gondii's* clinical relevance in humans it is associated with various pathologies in other animals, including livestock. This is problematic on two levels; (1) livestock that die from infection place an economic burden on farmers and a strain on the food supply chains, and (2) infected livestock that survive go on to pose an infection risk to humans through consumption of meat[38–40].

1.1.6. Current Treatments

Despite the significant burden posed to human and animal health by *T. gondii* worldwide, treatment options are limited. In immunocompetent individuals where disease is likely to be sub-clinical or mild, treatment may not be indicated. For those requiring treatment, the gold-standard is a co-prescription of pyrimethamine and sulfadiazine (pyr-sulf) (Figure 4)[41,42]. Pyrimethamine and sulfadiazine inhibit dihydrofolate reductase (DHFR) and sulfadiazine dihydropteroate synthetase respectively, preventing synthesis of folic acid[43]. Inhibition of these enzymes impacts DNA synthesis and cell proliferation. Pyr-Sulf treatment is only effective against the rapidly replicating tachyzoite parasite form so must be given in the acute stage of infection to achieve benefit. Pyr-Sulf treatment is associated with high incidence of side-effects and has a significant rate of failure. There is no treatment for the latent chronic infection rendering individuals at risk of acute infection recurrence for the remainder of their lives. As a
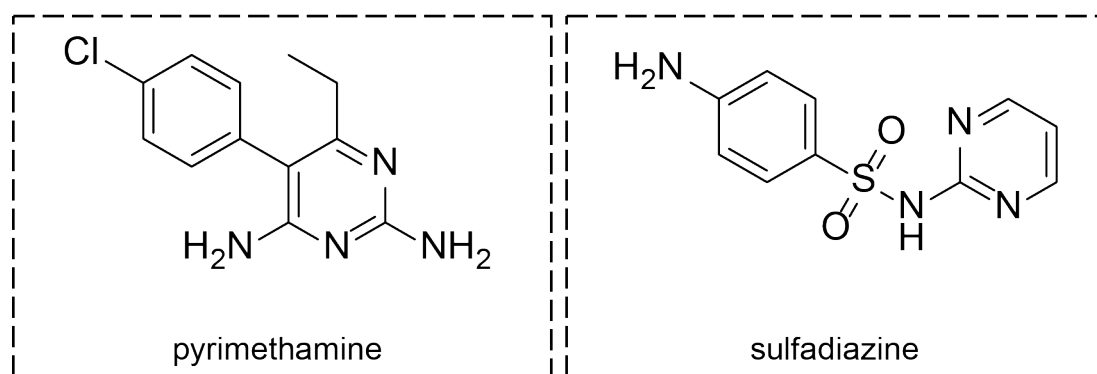
Figure 4 | The current gold-standard treatment for acute toxoplasmosis, co-prescription of pyrimethamine/sulfadiazine. This treatment is associated with high incidence of unpleasant side-effects and is only effective against the acute infection. Increasing resistance is being documented highlighting the need for new treatments.

result, treatment of immunocompromised patients may need to be ongoing for the duration of their compromised immune system[43].

1.2. Drug Discovery

1.2.1. Target-based Drug Discovery

The decrease in output from the research & development (R&D) sector of the pharmaceutical industry over recent years is well documented[44] and attributed to a range of factors for example, the discovery of all so-called 'low-hanging fruit'[45]. A recent analysis indicates that without substantial improvement, the pharmaceutical industry is unable to sustain sufficient output to offset the cost of R&D and patent expirations[46]. One possible contributory factor to this decrease in output is the shift to target-based drug-discovery approaches. Since the genomics era, drug discovery has focussed on targets with a clear and defined role in disease pathogenesis (Figure 5).

Target-based approaches incorporate the formation of a clear molecular hypothesis, which can then be investigated using rational medicinal chemistry techniques to create an inhibitor specific to the target. This approach requires specialised chemical and, ideally, structural knowledge. With the advancement of crystallography techniques and *in silico* platforms to aid in this rational design, it is of little surprise that this became the focus of much drug discovery research. However, addressing the molecular hypothesis does not reliably lead to inhibiting disease pathogenesis as the wider biological context of the target is often neglected, contributing to the high attrition rate associated with target-based drug discovery[47,48]. Despite this, target-based drug discovery has been favoured in recent years in part due to the confirmation of mechanism of action (MoA) in the preliminary stages of discovery. Inhibitors are designed specifically for these well-defined targets, minimising the risk of non-specific interactions and therefore unwanted side effects. This premise was in part to circumvent the risk of non-specific interactions posed by the incomplete screening
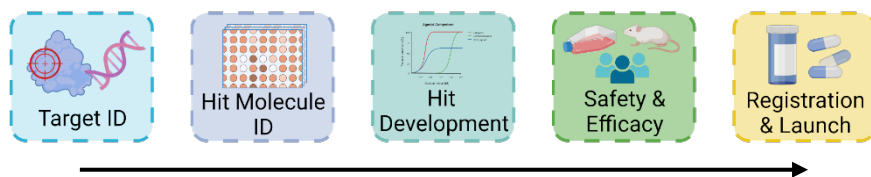
Figure 5 | Drug Discovery, the conventional pipeline. Most drug-discovery projects will begin with the process of target identification. A protein with disease connotations, e.g up-, down-regulation or mutation in a disease state would be regarded as a potential target of interest. Whilst this targeted approach means a greater understanding of potential drug mechanisms of action much of the early stage validation will be carried out against the target in isolation which may give an incomplete picture. Once a target has been identified high-throughput screens will be conducted using fragment libraries to detect molecules which may have an inhibitory effect. Once a hit molecule has been identified, further studies develop it to improve the inhibitory effect, specificity, efficiency of synthesis etc. Once a suitable hit drug has been developed this will be tested for safety and efficacy. First in an *in vitro* lab environment, subsequently in pre-clinical *in vivo* trials and finally in human trials. Initially these will be done in healthy volunteers to assess safety before progressing on to patient studies to determine efficacy. If the drug progresses successfully though trials it will then be registered and launched by pharmaceutical companies. This process takes approximately 10 years from start to finish with high rates of failure.

environment available in *in vitro* studies[49]. However, as previously alluded to, this wide-spread use of target-driven drug discovery has been associated with exceedingly high attrition rates in later stage trials overwhelmingly due to lack of efficacy[50].

1.2.2. Phenotypic Drug Discovery

Phenotypic drug discovery (PDD) is the identification of potential hit molecules by their effect on a disease state either *in vitro* or *in vivo*. Potential drug compounds are considered in a wider cellular context and if significant improvement in condition is observed, be that reduction in parasitaemia, prevention of metastasis or other indicators, that molecule may be considered a hit. Whilst this does not facilitate the immediate identification of the MoA it does allow for serendipitous events such as a drug having multiple targets with a combined effect or a protein being a viable target only in certain confirmations[51]. Prior to the shift towards target-focussed discovery, PDD was the commonly used methodology[52]. In comparison to target-based approaches this method interrogates the entire disease model, rather than one component in isolation, often simplifying the translation of results into a drug candidate with therapeutic effect[53]. Phenotypic methods are frequently incompatible with modern high-throughput screens (HTS) and are often unsuited to multiplexing as they rely on intricate interactions within a biological system, contributing to their fall from favour in recent years[52]. PDD has had great success with regards to first-in-class drug discovery largely due to the target agnostic approach removing the reliance on knowledge of a relationship between a specific target and disease state[54].

1.2.3. Medicinal Chemistry; Traditional and Diversity Orientated Synthesis

Target-based medicinal chemistry has been structured to adhere to rules and guidelines, for example Lipinski's rule-of-five, by which drug-like molecules are supposed to abide. This framework places limits on properties including lipophilicity and molecular weight of a compound, based on analyses of existing drugs[55]. While these are useful guidelines for rational design and in the categorisation of drugs, they can also function as arbitrary cut-offs that may limit the discovery process. Notably, Lipinski's rules were formed through retrospective analysis of available drug compounds and were not intended to act as definitive features for future discovery. Familiarity with procedures and reagents will inevitably play a role in any synthesis. Moieties which have been shown to work against structurally or biochemically similar targets are, rationally, the first place many medicinal chemists will begin. This perpetuates the ongoing use of familiar molecular structures, slowly narrowing the diversity of drug candidates published. Although this is not inherently problematic, through unconscious bias it is possible that highly efficacious candidates are missed, and novel discovery potential is limited.
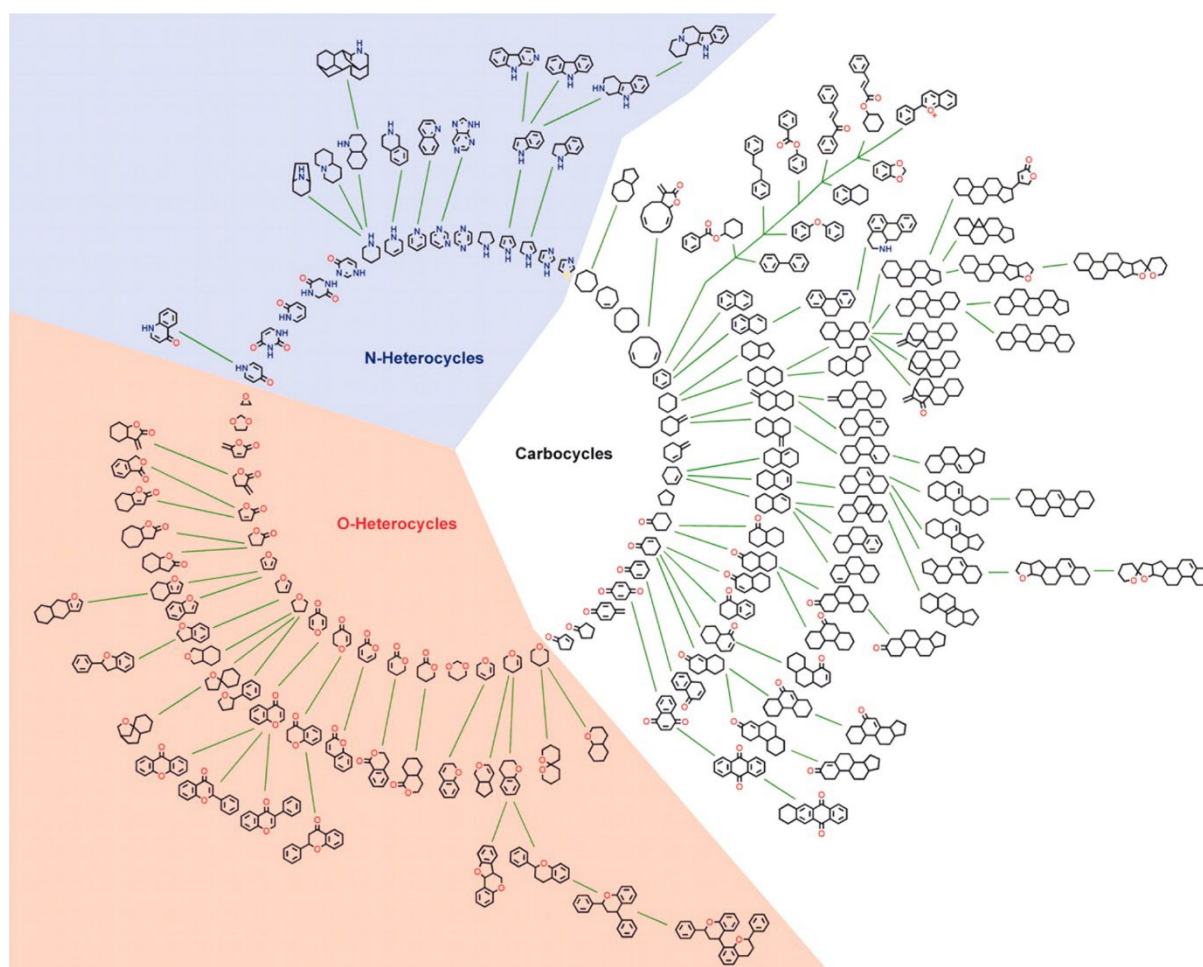
Figure 6 | This figure is taken from van Hattum & Waldmann's Biology-Oriented Synthesis: Harnessing the Power of Evolution, 2014[63]. Unbiased synthesis methods can be used to discover many drug compounds by utilising key scaffolds identified from natural products.

A recent analysis of reported crystal structures from the hydrothermal synthesis of amine templated metal oxides showed how far this bias truly goes. The authors of the study found that 79% of reported structures incorporated just 17% of available amine reactants. On substituting so-called 'common' amine reactants for 'uncommon' reactants in the reported reactions it was noted that there was an almost identical reaction success rate, with similar yields obtained. The 'uncommon' and 'common' reagents used in this study were all available from major suppliers in equal quantities and were of similar prices. These factors considered, inherent bias towards familiarity is one of the only remaining explanations for the un-representative distribution of amines typically used in chemical syntheses[56]. Another study used machine-learning artificial intelligence (AI) to predict the structure of novel antimicrobial agents with the capacity to screen 107,349,233 small molecules in four days. While the ability to explore such vast amounts of chemical space is an invaluable tool for drug discovery, this approach shares fallibilities with previous methods. The model was trained on established drug libraries and, rather than interrogating the effect of small molecules on microbes, it sought to identify molecules similar to previously identified antibiotics thereby inheriting human bias[57]. To acknowledge and address this bias in medicinal chemistry, the principle of diversity orientated synthesis (DOS) has been established (Figure 6)[58].

This strategy focuses on sampling a wide and varied portion of chemical space unrestricted by guidelines and arbitrary cut-offs to identify previously ignored bio-active small molecules, particularly those with the potential to target the so-called undruggable proteome[59]. Following investigations into how best to achieve this diversity with regards to molecular structure and range of targets engaged, variation in core scaffold was identified as having the most significant impact on subsequent interactions[59–61]. A DOS approach conceived by Waldmann et al. incorporated the bioinformatic analysis of scaffolds frequently observed in natural products (Figure 6)[62], the principle being that this library of diverse natural product derived scaffolds could then be used as a template for further medicinal chemistry studies. The rationale is that while it is important to sample a wide range of chemical space, fundamental constraints on time and synthesis capacity mean that some selection process must be implemented. By using scaffolds which are frequently identified in natural products, and are therefore pre-validated as biologically active, this selection can be done without enforcing human bias[58,63]. Weber et al. developed a strategy wherein initial reactions were selected from a sample set at random, with products of these reactions then tested in biological assays and the most efficacious molecules combined and taken forward into a second stage of synthesis. The Nelson Group at Leeds University has developed a strategy for activity-based development, which revolves around the concept of pluripotency. This methodology employs a plate-based format, where reactive intermediates are combined with

various substrates and solvents in reactions. After an incubation, the compounds generated are analysed to identify the chemical products. These products are then screened in biochemical assays, and those demonstrating activity are further pursued.[64]. These strategies begin to address the bias in medicinal chemistry, as the focus is on testing a diverse set of molecules and taking forward those that successfully affect the targeted phenotype. However, by focussing on reactions which synthesise drug-like molecules, bias is not eliminated[65–67].

1.2.4. Covalence in the Clinic

Interactions between small molecule drug compounds and their biological targets can occur through a variety of mechanisms. One such mechanism, covalent binding, is characterised by the formation of an irreversible (covalent) bond between the compound and its target. Drugs functioning through a covalent mechanism of action (MoA) have experienced a resurgence in the last three decades. For many years, covalent drugs were considered unsafe with an increased propensity for off target effects. Combined with extended duration of action due to irreversible binding, concerns regarding toxicity and unwanted side effects prohibited development[68,69]. Drugs which are now known to act through a covalent MoA, for example aspirin, were often in wide-spread use before this was discovered[70,71]. Antibiotics such as penicillin and Fosfomycin, which have covalent MoAs, were derived, or inspired, by natural products with observed phenotypic impact[72,73]. Covalent drugs have also played a role in cancer therapy, the pro-drugs 5-fluorouracil[74] and gemcitabine[75] are chemotherapies used to treat colorectal, breast, pancreatic and bladder cancers amongst others. Covalent drugs are commonly and effectively used to treat an array of disease states however with enhanced focus on this modality in drug discovery studies their utility may continue to increase.

Key milestones in the recent resurgence of covalent drugs include approval of the first inhibitor of the receptor tyrosine kinase EGFR, afatinib, in 2013 (Figure 7)[76]. EGFR drives progression of non-small cell lung cancer and so is considered a target of importance in

oncology. By modifying reversible inhibitors approved in the early 2000s[77,78] with moderate electrophiles, such as acrylamides, efficacy of these treatments was improved. This practice is referred to as ligand first, covalent drug discovery, as an existing ligand is modified for use as a covalent inhibitor. However, afatinib did not display sufficient selectivity for disease specific mutants of EGFR relative to wild-type, leading to high toxicity[79]. Improvements were made leading to second-generation drugs including, osimertinib[80] and rociletinib[81] which exhibited enhanced safety profiles facilitating higher treatment doses and better therapeutic effects. The targeting of a single mutated cysteine is a limitation of EGFR covalent inhibitors as this makes them susceptible to resistance[82].

Bruton's tyrosine kinase (BTK) inhibitors are another example of ligand first covalent drug discovery. BTK was identified as a target of interest in chronic lymphocytic leukaemia[83] and rheumatoid arthritis[84]. An acrylamide containing compound, later named ibrutinib (Figure 7), was designed to facilitate in target validation through fluorescent labelling of BTK[85] however was found to have sufficient inhibitory effects to proceed with drug trials[86–88]. Ibrutinib continues to exhibit clinical and financial success demonstrating the potential for blockbuster covalent drugs[89,90].

With renewed confidence in covalent small molecules as viable drugs, research has shifted toward an approach known as "electrophile-first discovery." This approach involves the initial design and development of covalent inhibitors, as opposed to modifying reversible inhibitors. The most prominent example of electrophile first discovery is inhibitor development for the GTPase KRAS(G12C). The G12C mutation is present in approximately 13% of lung adenocarcinoma cases and 3% of colorectal cancers[91]. KRAS(G12C) was identified as an oncogenic driver ~30 years ago and long eluded drug-discovery teams, with GTPases considered 'undruggable'[92,93]. Their conversion between conformations resulted in no clear
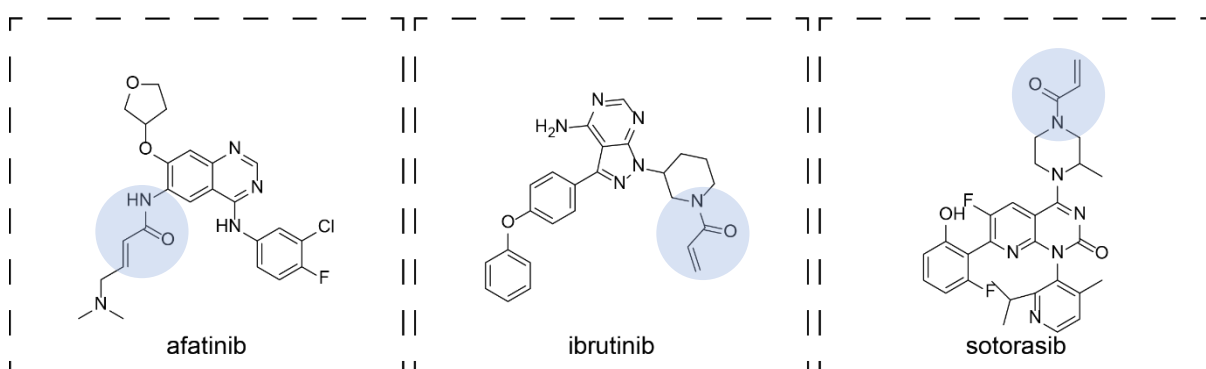
Figure 7 | Covalent drug compounds in the clinic. Acrylamide electrophilic warheads highlighted in blue. Afatanib, an EGFR inhibitor, is utilized for non-small cell lung cancer treatment. Ibrutinib, a BTK inhibitor, has achieved consistent financial success since its registration in 2013[88]. Sotorasib, developed using an electrophile-first approach, is employed for treating cancers with the KRAS G12C mutation.

active site to target and reversible inhibitors were unable to compete with the high affinity for and abundance of GTP and GDP, their natural substrates[94]. An inhibitor with a covalent modality offers advantages in this case. By targeting a mutated cysteine present only in the disease state, the risk of toxicity is significantly reduced. Additionally, due to irreversibly binding the target, covalent inhibitors are not in constant competition with natural substrate, likely requiring a lower effective concentration. Using an electrophile-first design approach several covalent molecules reached trials for treatment of KRAS(G12C) mutated cancers[95–97], and in 2021 sotorasib was approved for use in treating non-small cell lung cancer (Figure 7)[98]. The success in this case further validates the use of covalent inhibitors as precision therapies for modified disease states[99].

1.2.5. CRISPR-Cas9 Gene Editing

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 technology is a modified version of the bacterial CRISPR-Cas mechanism, a protective adaptive immune response. Vectors, DNA constructs that carry the components necessary for utilising the CRISPR-Cas9 gene editing system, are constructed to target the desired locus. Vectors typically consist of several elements; (1) A guide RNA (gRNA) sequence which has a
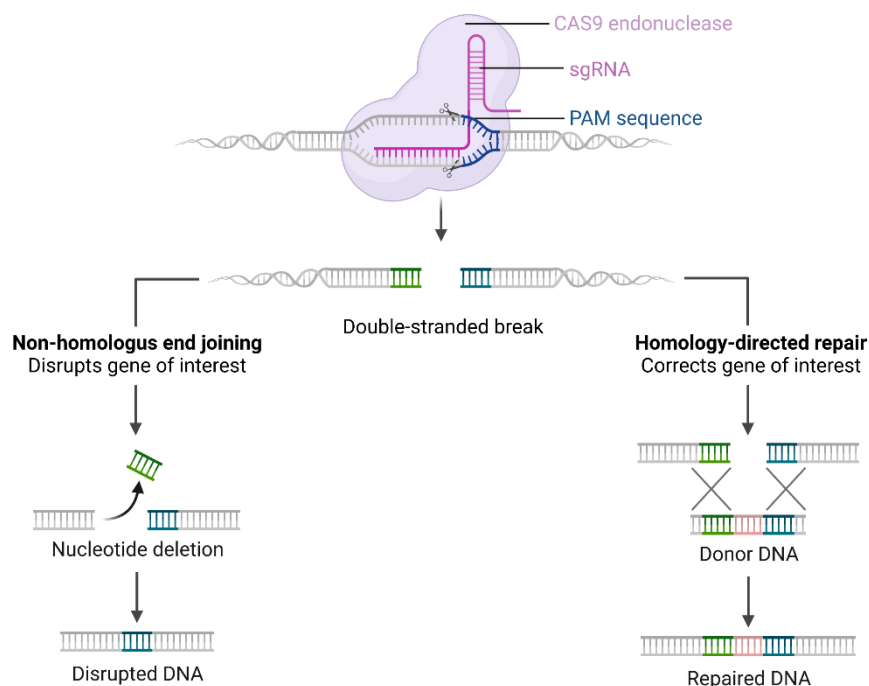
Figure 1 | CRISPR-Cas9 gene editing results in a double-stranded DNA break. (a) NHEJ is a mechanism used by cells to repair DSBs, in the absence of a homologous template. This repair pathway is considered error-prone as it can lead to the introduction of small insertions or deletions at the site of repair, potentially causing mutations. (b) HDR repairs DSBs and ensures precise restoration of the original DNA sequence by using homologous templates.

complimentary sequence in the targeted genomic location. (2) A transactivating CRISPR RNA (tracr) that, in combination with the gRNA, forms a scaffold that guides Cas9 endonuclease to induce a double stranded break (DSB) at specific genetic locus [100]. (3) The Cas9 gene which encodes the Cas9 endonuclease. (4) The promotor, which facilitates expression of the gRNA coding sequence. The relative binding efficiency between the gRNA and target DNA sequence will determine the specificity of a given CRISPR-Cas9 system. Preceding the gRNA recognition sequence is a promoter, which facilitates expression of the Cas endonuclease and the guide RNA.

In the target genetic sequence, the protospacer adjacent motif (PAM) follows the gRNA recognition sequence. The PAM is required for successful binding of the gRNA. The targeted endonuclease activity of Cas9 to a position three nucleotides 5´ of the PAM results in the formation of double-strand breaks (DSBs), which are repaired through either non-homologous end joining (NHEJ), or homology-directed repair (HDR) machinery. Taking advantage of HDR, it is possible to use the targeted formation of DSBs to facilitate precise incorporation of DNA such as oligonucleotide donor repair templates into the disrupted locus (Figure 8a)[101]. These repair templates are designed to have regions of homology to the targeted locus which ensure specificity to the target site, with Cas9 activity triggering high-efficiency HDR in a 30 nt region 5´ and 3´ of the DSB. NHEJ creates unpredictable recombination, introducing insertions and deletions which disrupts gene expression by shifting the reading frame of a sequence resulting in incorrect translation leading to production of non-functional proteins (Figure 8b)[102]. This methodology has been utilised to evaluate the essentiality of genes. By employing a CRISPR-Cas9 system to target specific genes and subsequently monitoring cell survival and proliferation, it becomes possible to distinguish between essential and non-essential genes.

1.2.6. Next-Generation Sequencing

Next-generation sequencing (NGS) facilitates the parallel sequencing of many DNA samples in a comparatively rapid and low-cost manner. In the case of amplicon deep sequencing, the DNA fragment of interest is amplified by PCR and unique combinations of 'indexes' are appended in subsequent PCR. During sequencing, these DNA fragments are denatured and passed over a flow-cell. Index complimentary nucleotides are present on the flow-cell surface and bind the 5′ of the single strand DNA. Primed from the 3′ end, a new, complimentary, strand is synthesised, and the original strand is denatured. The 3′ end then anneals to a complimentary nucleotide of the flow-cell surface forming a bridge. The 5' end is then released from the flow-cell with the 3' bound to a new second strand synthesis location. Multiple cycles result in the formation of DNA fragment 'clusters' on the flow-cell surface. The DNA is cleaved from the flow-cell surface and denatured, generating templates for sequencing by synthesis. Sequencing by synthesis is the imaging of a fluorescently labelled reversible terminator as

a Library preparation

gDNA

↓ Amplicon generation by PCR

Indexing adapter added

DNA library

b DNA library bridge amplification

Library hybridization

↓

Amplified clusters

Bridge amplification cycles

c DNA library sequencing

Fluorescently labeled nucleotides

Sequencing cycles

↓

Data collection
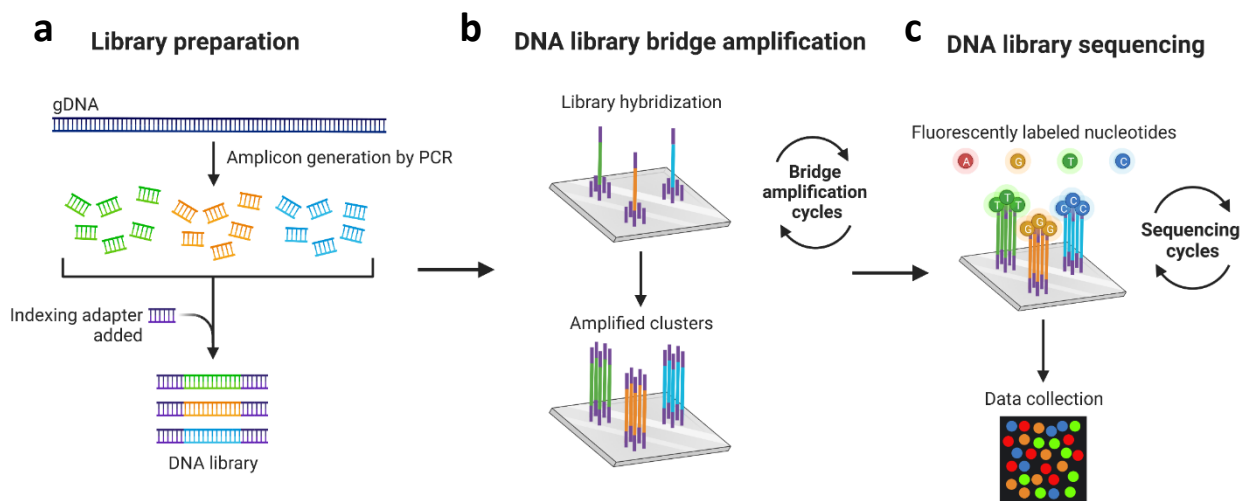
Figure | 9 Next-generation sequencing is high-throughput, massively parallel DNA sequencing technology that allows for rapid and efficient sequencing of large amounts of DNA. (a) PCR is used to generate DNA amplicons, which are then indexed using unique adapter primer combinations. (b) To generate enough material for sequencing, samples are first clustered on the flow-cell surface and then subjected to multiple rounds of amplification. (c) The sequence is generated by imaging fluorescently labelled reversible terminators after the addition of each dNTP.

each dNTP is added in DNA synthesis. The terminator is cleaved after each incorporation to facilitate integration of the next base[103,104].

1.2.7. DNA & Short Oligonucleotide Barcodes

The increase in high throughput drug screening has precipitated research and development of auxiliary techniques to aid their implementation. Cellular barcoding is an innovative molecular technique which has revolutionised various fields of biological research by enabling the accurate identification and differentiation of individual entities within complex mixtures. At its core, cellular barcoding involves the incorporation of short DNA or RNA sequences, known as oligonucleotides, into target molecules, such as DNA fragments or proteins. These unique molecular tags act as distinctive signatures, allowing for the precise tracking and analysis of multiple entities in parallel. By assigning specific barcodes to distinct components, researchers gain the ability to dissect complex samples and unravel intricate interactions with unprecedented precision [1,105,106]. This method finds applications across diverse disciplines, ranging from genomics and proteomics to single-cell analysis and high-throughput screening, fundamentally enhancing our understanding of intricate biological systems[107,108].

Barcoding has enhanced the capabilities of Next-Generation Sequencing (NGS). By incorporating unique barcodes into individual samples prior to sequencing, researchers have harnessed the power of multiplexing, allowing multiple samples to be analysed in a single sequencing run. Multiplexing has dramatically increased the efficiency and throughput of NGS,

enabling researchers to simultaneously examine numerous samples while minimising costs and time. Additionally, barcoding has facilitated the accurate demultiplexing of sequencing reads, ensuring that data from different samples can be sorted and analysed individually post-sequencing. This advancement has proven especially beneficial in large-scale genomics projects, metagenomics studies, and single-cell analyses, where the ability to distinguish and trace individual components is paramount[105,106].

Cellular barcoding has also played a pivotal role in advancing lineage tracing experiments by offering a sophisticated means to track and unravel the developmental or lineage history of individual cells within complex biological systems. By introducing barcodes into cells at specific time points or stages, researchers create distinct genetic identifiers that persist as the cells divide and differentiate. As cells undergo changes and progress along their developmental paths, the barcodes serve as molecular tags that provide a historical record of their lineage. Through the application of modern sequencing technologies, these barcodes can be read and analysed, enabling researchers to reconstruct the lineage relationships, trajectories, and diversification of cell populations over time. This approach not only sheds light on the dynamic processes that underlie tissue development, regeneration, and disease progression but also offers insights into the heterogeneity and functional properties of different cell subsets[109,110].

Examples of barcode aided lineage tracing experiments are those featuring the Restriction Site Tagged Poliovirus (RSTPV) technique. This is a methodology used in molecular virology to label and track the poliovirus genome. This method involves the deliberate introduction of restriction sites, into the genetic material of the poliovirus. These restriction sites act as markers that can be recognised and cut by restriction endonucleases. By strategically placing these restriction sites within the viral genome, researchers can create unique DNA fragments when the virus is digested with the corresponding restriction enzymes. The resulting DNA fragments can then be subjected to various molecular analyses, such as PCR or DNA sequencing. This approach allows researchers to gain insights into the genetic characteristics, mutations, and variations of the poliovirus within a specific outbreak or context. Additionally, the restriction sites serve as specific landmarks, aiding in the accurate mapping of the viral genome. Using this technique in murine poliovirus infections, researchers successfully identified a bottleneck occurring between inoculation and brain colonisation, which resulted in limited genetic diversity of the persisting virus in the murine brain. The bottleneck observed between inoculation and brain colonisation in the case of restriction site-tagged poliovirus is hypothesised to be an anti-viral state rather than a physical bottleneck. This hypothesis is supported by the fact that the bottleneck is no longer observed with a 10^7 fold increase in viral inoculum. It is believed that this bottleneck represents an innate immune

response aimed at limiting the pathogenicity of the virus as colonisation of the CNS and fatal disease outcomes are closely linked[111,112].

Wild-type Isogenic Tagged Strains (WITS) are wild-type organisms tagged with specific genetic barcodes. These barcodes function as neutral alleles so complex populations can be monitored over time and analysed by quantitative NGS. This approach allows researchers to study the dynamics and changes in population structure during infection. Using WITS, scientists have been able to investigate the bottlenecks that occur during gastrointestinal colonisation by Salmonella. The analysis of WITS data has shed light on the genetic diversity and clonal expansion of Salmonella populations within the gastrointestinal tract. It has revealed the presence of genetic bottlenecks during colonisation, indicating that only certain clones or variants of Salmonella can establish a successful infection. The use of WITS in studying Salmonella infections has provided valuable information about population dynamics, genetic diversity, and the impact of bottlenecks during gastrointestinal colonisation, contributing to a better understanding of Salmonella pathogenesis[113–115].

There have been recent advances in methodologies for analysis with regards to tracking microbial population dynamics during infections. Traditional population dynamics studies have been hindered by small numbers of tags, highly specialised mathematical models, and lack of systematic analyses. These limitations become increasingly pronounced in complex populations. Using *in vivo* studies of *Vibrio cholerae* (*V. cholerae*) as a model, researchers developed Sequence tag–based analysis of microbial populations (STAMP). This methodology combines population analysis frameworks with high-throughput DNA sequencing technology and large libraries of tagged pathogens. STAMP allows the determination of relative abundances of individually tagged strains in the infection inoculum, offering insight into population bottlenecks and the founding population size. This technique provides a robust framework for characterising microbial dissemination and is a versatile tool for understanding the dynamics of pathogen populations in different contexts[110,116].

Cellular barcoding has been employed to enable multiplexed *in vivo* drug screening in studies involving Pancreatic ductal adenocarcinoma (PDAC) cells, known for their propensity for metastasis. In a plate-based format, 1152 distinct populations of PDAC cells were individually barcoded. Each population was subjected to a 6-hour treatment with a small-molecule irreversible inhibitor, followed by washing and pooling. The resulting samples were intravenously transplanted into mice, while the remaining cells served as the input sample. Over a 48-hour period, the cells underwent metastatic seeding, and subsequently, tumours were harvested from the lungs of the mice. The representation of barcodes in the tumours was compared to their representation in the input sample to assess if any cell populations were

absent. The absence of certain barcodes in the tumour indicates their likely treatment with a small-molecule exhibiting anti-metastatic properties[117].

1.2.8. Fluorescent Barcodes

Fluorescent barcoding, whilst having fewer possible permutations than DNA barcodes, may be more easily incorporated into drug screening workflows and readouts often utilise common equipment[118,119]. By incorporating further variables in addition to different colours, it is possible to increase the number of marker combinations. For example, encoding for different levels of signal intensity enables further differentiation. By simply incorporating no fluorescence, low intensity and high-intensity, the possible combinations are increased three-fold[120,121].

Multiplexing experiments facilitated by barcoding is cost-, resource- and time-saving. By carrying out multiple data-collections simultaneously, quality of comparisons between sets may be enhanced and truer evaluations of replicates is possible. Barcoding can be used to facilitate drug-screening or answering of fundamental biological questions from infection dynamics to lineage of tumour cells. By incorporating these techniques into laboratory workflows, throughput may be increased, and data collection streamlined, and new biology discovered.

1.3. Aims & Objectives

The primary aim of this thesis is to establish and apply cellular barcoding techniques in *T. gondii* and subsequently use these techniques to investigate in host population dynamics and enable multiplexed *in vivo* drug screening.

- To establish cellular barcoding techniques in *T. gondii*:
  - Develop a robust and reproducible protocol for generating barcoded *T. gondii* populations.
  - Validate the effectiveness and stability of barcodes in *T. gondii* cells over multiple generations.
  - Optimise the experimental conditions to minimise potential impact on parasite viability and growth.
- To assess transferability of the barcoding technique to other parasitic pathogens:
  - Investigate the applicability of the established barcoding technique to different parasitic pathogens.
  - Compare the efficiency and reliability of barcoding between *T. gondii* and other selected parasites.
  - Address potential challenges and limitations in adapting the technique to various pathogens.
- To investigate within-host population dynamics of murine *T. gondii* infection:
  - Apply cellular barcoding to track and analyse the dynamics of barcoded *T. gondii* populations within a murine host.
  - Monitor changes in relative abundance of distinct populations during various stages of infection.
- To establish multiplexed *in vivo* drug screening using distinct barcoded *T. gondii* populations:

- o Develop an experimental framework to simultaneously evaluate the efficacy of multiple drugs in a single host.
- o Identify potential drug candidates for further development based on their impact on specific parasite populations.

By addressing these aims and objectives, this thesis contributes to a deeper understanding of cellular barcoding techniques in *T. gondii*, provides insights into within-host parasite population dynamics, and demonstrates the potential of multiplexed *in vivo* drug screening for advancing anti-parasitic drug development.

2. Materials & Methods

Unless otherwise stated, all chemicals were purchased from Sigma Aldrich. All primers, oligonucleotides and plasmid constructs generated and/or used in this study are listed in the appendices (Chapter 7).

2.1. Molecular Biology

2.1.1. PCRs

Various oligonucleotide primers were used to amplify DNA fragments from genomic DNA (gDNA) by PCR (Section 7.1.4). PCRs were performed using different polymerases, following the associated manufacturer protocols. For parasite genotyping, the Q5® High-Fidelity DNA Polymerase (NEB) was used with 10 ng of template DNA. Products from these reactions were purified using Monarch® PCR & DNA Cleanup Kit (NEB). To generate amplicons for Illumina sequencing, the KAPA HiFi HotStart ReadyMix Kit (Roche) was used with 20 ng of template DNA. These NGS amplicons were purified using AMPure XP magnetic beads (Beckman Coulter) following the manufacturer's guidelines with a volumetric bead-to-sample ratio of 1:1.25. Correct amplification was routinely assessed by visualising bands following electrophoresis on a 1% (v/v) agarose gel stained with SYBERSafe (Thermo Fisher). To generate gDNA templates for PCR, total DNA was extracted from cell pellets using the DNeasy® Blood & Tissue Kit (QIAGEN).

2.1.2. Plasmid Preparation

To scale up plasmids for transfection of the Tg*UPRT* locus, 1600 ml of culture was used, and plasmids were isolated using the QIAGEN® Plasmid Gigaprep Kit (QIAGEN). Prior to transfection, all constructs were purified by extraction with 1 vol phenol:chloroform:isoamyl alcohol (25:24:1), followed by 1 vol chloroform. Purified constructs were then desalted by ethanol precipitation (0.1 vol 3M sodium acetate pH 5.2, 2 vol 100% ethanol), washed in 70% ethanol, air dried and resuspended in the appropriate transfection buffer.

2.1.3. Next Generation Sequencing Library Preparation

Frozen cell pellets of parasites were thawed to room temperature and gDNA extracted using the DNeasy Blood & Tissue Kit (Qiagen). Genomic DNA libraries were prepared following the 16S Metagenomic Sequencing Library Preparation guide (Illumina). In brief, an~300 base pair amplicon region encompassing the 6 nt barcode sequence was amplified (30 cycles) from the barcoded UPRT locus in *Toxoplasma gondii* using primer sequences (5′ to 3′)

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtggatgtgtcataccatggagtttcctg and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgttttagtgtaacaaagtggacagcagc.

The barcoded AAT6 locus in *Trypanosome brucei brucei* was amplified using primer

sequences (5′ to 3′)

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGatgcaatcagaagacgaggtttaagtag

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGagcgagttcaatactaacaataaccagg.

These primer sequences include the specified Illumina adapter overhang sequences (bold, uppercase). AMPure XP beads were used to purify the resulting PCR product. An indexing PCR (10 cycles) was carried out using the purified product as the template to add dual indices and sequencing adapters to the amplicon using the Nextera XT Index Kit (Illumina). Indexed libraries were then cleaned using AMPure XP beads and quantified on the Quantus Fluorometer using the QuantiFluor ONE dsDNA System (Promega). Amplicons were purity-checked and sized on a TapeStation using D1000 ScreenTape System (Agilent). For each NGS run, typically 8 to 25 uniquely indexed libraries were pooled at equimolar concentrations for multiplexed outputs on either an Illumina MiSeq or NextSeq sequencer using the MiSeqV3 PE 75 bp kit or NextSeq 500/550 Mid Output v2.5 PE 75 bp kit respectively. PhiX DNA spike-in of 20% was used in all NGS runs. Testing the sensitivity of the NGS pipeline, a 96 well plate of 96 uniquely barcoded strains was set up (using transfected drug resistant parasites), with 10,000 parasites/well. For the plate (12 columns x 8 rows), a serial two-fold dilution was performed across the 12 columns, and all rows in the final column pooled after the final dilution. Genomic DNA was then prepared from the final pooled sample and processed for NGS as described.

2.1.4. Covalent fragment library curation

The library of 84 covalent fragments used in the first drug screen was curated from a collaborator's selection of 730 compounds. All fragments had a cysteine reactive acrylamide warhead. Compounds were filtered using the FragFp diversity algorithm in DataWarrior v5.5.0 with 84 fragments chosen for maximum diversity[122]. FragFp is a method used to measure the similarity between chemical compounds based on their fragment fingerprints.

The FragFp algorithm breaks down chemical compounds into smaller molecular fragments, such as functional groups or substructures. These fragments are then represented as a binary fingerprint, indicating their presence or absence in a compound. The fingerprints are generated by encoding the fragments into a fixed-length bitstring, where each bit represents the presence or absence of a specific fragment. To calculate the similarity between two compounds using FragFp, the algorithm compares their respective fingerprint bitstrings. The similarity score is determined by counting the number of shared fragments between the compounds and dividing it by the total number of fragments present in either compound. A higher similarity score indicates a greater resemblance between the compounds in terms of their fragment composition.

2.2. Tissue Culture

2.2.1. *T. gondii* Parasite Culture

T. gondii parasite strains were maintained by serial passage in confluent human foreskin fibroblasts (HFF-1 ATCC® SCRC-1041™). HFFs were cultured at 37°C with 5% CO2 in Dulbecco's Modified Eagle's medium supplemented with 10% fetal bovine serum and 2 mM L-glutamine. Tachyzoites were harvested via mechanical syringe lysis of heavily infected HFFs through a 25-gauge needle. RHΔ*ku80* parasites were used for *in vivo* and *in vitro* studies. PruΔ*ku80* parasites were used in *in vivo* experiments where chronic infections were established. For the cryopreservation of intracellular parasites, heavily infected monolayers in T-25 ($25cm^2$) flasks were trypsinised, gently resuspended in freezing solution (DMEM +25% FBS +10% DMSO) and transferred to cryovials for freezing at -80°C or in liquid nitrogen.

2.2.2. *T. brucei brucei* Parasite Culture

We used a bloodstream form Lister 427 strain *T. brucei brucei* for all experiments. The cells were cultured at 37°C, 5% CO2 in HMI-9 medium supplemented with 20% heat inactivated Foetal Bovine Serum, 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco).

2.2.3. Generation of barcoded *T. gondii* strains and libraries

60-nucleotide single-stranded oligos were designed to include a unique six nucleotide barcode sequence flanked by a stop codon and homology regions on either side. Barcodes were designed using the DNA barcode designer and decoder, nxcode (http://hannonlab.cshl.edu/nxCode/nxCode/main.html). The sequences of all oligos within the 96-member library can be found in key resources table. Barcoded libraries of tachyzoites were generated using two alternative strategies: For strategy A, 96 independent transfections were carried out in 16 well Nucleocuvette strips. 10 µg of the pSAG1::Cas9-U6::sgUPRT vector and 10 µg of the barcode oligo (equivalent to an~1:160molar ratio of plasmid to oligo) were co-transfected into approximately $1×10^6$ extracellular tachyzoites using the 4D-Nucleofector X Unit programme F1-115 (Lonza). 24 hours post-transfection, transgenic barcoded parasites were selected for using 5 µM 5'-fluro-2'-deoxyuridine (FUDR). Barcoded strains were independently maintained, and only pooled just prior to use. For strategy B, a single "one-pot" transfection was carried out. An oligo library pool containing roughly equal amounts of all barcode oligos was prepared. The ratio of the pSAG1::Cas9-U6::sgUPRT vector to the total oligo pool was the same as in strategy A, though here the final concentration of any single oligo within the pool was~100-fold less. Transfection and selection were performed as for A, with the complex barcoded strain library generated and maintained as a single population.

2.2.4. Generation of barcoded *T. brucei brucei*

60-nucleotide double-stranded oligos were designed to include a unique six nucleotide

barcode sequence (bold, upper case) flanked by a stop codon (lower case) and 24 bp homology regions on either side:

TGCAATCAGAAGACGAGGTTTAAGtagAAACACtgaCTCACACTAACCGTTTCGATTTAC. Amino acid transporter AAT6 (Tb927.8.5450) was selected as a suitable locus for barcode integration. DNA encoding the sgRNA sequence targeting the AAT6 locus (GTTTAAGTTCACATTGTCGC) was generated by PCR as described in (Rico et al., 2018), ethanol precipitated, and 10 µg was mixed with 10 ng pre-annealed oligonucleotides. The mixture (20 µl total volume) was added to $\sim$10^7 Cas9 expressing bloodstream form Lister 427 *T. b. brucei* cells in 100 µl Amaxa buffer and electroporated using Amaxa Nucleofector IIb (Lonza) program X-001. Transfected cells were immediately added to pre-warmed HMI-9 medium containing 270 µM eflornithine. Transfected cell cultures were passaged under selection every two days. Drug resistant parasites were harvested seven days after transfection for gDNA isolation. PCR amplicons encompassing the barcoded region of the AAT6 locus (from four independent transfections) were generated using ORF-specific PCR primer sequences (5′ to 3′) ATGAGAGAGCCGATACAAACTTCAAC and TCAGAGTTCAGCAATGACGCTG. Barcode integration was confirmed by Sanger sequencing of these amplicons. For the one-pot transfection strategy, complementary single-stranded barcoding oligos were annealed to produce 96 unique double stranded barcoding repair templates. Annealed barcoding oligos were then pooled, and used in a single transfection as described above.

2.2.5. Plaque Assays

Growth of *T. gondii* tachyzoites was assessed by means of plaque formation on confluent HFF monolayers. Monolayers grown in 6-well plates were seeded with 200-400 parasites and plaques were left to form undisturbed for 6-7 days. Following this, the monolayer is fixed using ice cold ethanol at 4°C for 30 minutes. Crystal violet, a broad-spectrum stain, is then applied to the monolayer staining everything but areas of clearance, caused by parasite lysis. Areas of clearance, referred to as plaques, are indicative of viable parasites. Plaques were enumerated manually, and statistical significance in plaque counts between cell lines and/or treated/untreated samples were tested using two-tailed unpaired Student's *t*-tests with unequal variance. The data are presented as mean (±SD) counts.

2.2.6. Drug Screening

Parasites barcoded in a plate-based format were transferred to a v-well plate diluted in media with drug fragments at the appropriate concentration. The treatment was for 1 hour in an incubator at 37°C. Following treatment, drug media was removed, and parasites were resuspended in sterile PBS. Parasites were counted and diluted to a concentration of 1 million/mL. 200 uL was used for each infection and the remaining parasites were taken as

the input sample. Parasites were sampled and passaged at five- and ten-days post infection.

2.2.7. MTT Assay

HFFs were cultured in 96-well plates in complete DMEM prior to incubation with the control and experimental treatment. The media was replaced with complete DMEM and 50 µg MTT (Roche). Cells were incubated for 4 hours at 37°C. Cells were treated with 10% SDS (Roche) in 0.01 M HCl (Roche) and incubated at 37°C overnight. Absorbance of the formazan was read at 570 nm on a microplate reader.

2.3. Animal work

2.3.1. Strains Used

Six-week-old female C57BL/6 or CBA/J mice were purchased from Jackson Laboratories. Mice were acclimated for seven days prior to infection. Six- to eight-week-old female Swiss Webster mice (originally form Jackson Laboratories) were obtained from the University of Virginia Centre for Comparative Medicine foster and sentinel colony. For studies using CBA/J and Swiss Webster mice, the animal protocols were approved by the University of Virginia Institutional Animal Care and Use Committee (protocol # 4107-12-18). All animals were housed and treated in accordance with AAALAC and IACUC guidelines at the University of Virginia Veterinary Centre for Comparative Medicine. The procedures involving C57BL/6 (multiplex experiment) mice were approved by the local ethical committee of the Francis Crick Institute Ltd, Mill Hill Laboratory and are part of a project license approved by the Home Office, UK, under the Animals (Scientific Procedures) Act 4001986.

2.3.2. Population Dynamics Experiments

For intraperitoneal (IP) infection, the pooled barcode parasite library was expanded on HFFs in a T175 flask. Once full parasite vacuoles were observed, parasites were scraped and syringe lysed, counted on a hemocytometer and diluted to an inoculum of 37,000 viable parasites or 12,000 viable parasites in 200 µL of PBS per mouse. The numbers of viable parasites in the IP infection inoculums were determined by plaque assay. At the time of inoculation 2×10^6 parasites were frozen as an initial population control. In addition, three inoculum control samples were expanded immediately on HFF T25 flasks. After 48 or 72 hours three to five mice were euthanised to isolate parasites in the peritoneal exudate. Specifically, 10 mL of PBS was injected by 25G needle into the peritoneal cavity, mice were rocked vigorously, and peritoneal fluid removed by syringe. Parasites and exudate cells were washed twice in 10 mL of media containing penicillin/streptomycin, pelleted at 1,500 rpm and plated on HFFs T25 flasks. Parasites were harvested when they approached full lysis of the monolayer pelleted and frozen for gDNA isolation. After 28 days or three months, the remaining mice were euthanised. Carcasses were incubated in 20% bleach for 10 minutes

and the brain was excised in the biosafety cabinet under sterile conditions. To isolate parasites the brains were mashed though a 70 µm filter using 25mL PBS with 5% FBS and penicillin/streptomycin. Brain mash was pelleted for 10 minutes at 1,500 rpm, washed twice with PBS and penicillin/streptomycin then plated on HFF monolayers in T75 flasks. After 36 hours, media was changed to remove debris. Parasites were harvested by syringe lysis when the HFF monolayer was nearly lysed out (approximately two weeks), pelleted and frozen for gDNA isolation. To confirm cyst formation in the brain at one month (28 days) or three months post infection,1/50th of the mash was reserved, fixed in 4% paraformaldehyde for 15 minutes then stained with a 1:500 dilution of Dolichos bifluorus agglutinin conjugated to FITC in PBS (Vector Labs). FITC-positive cysts were confirmed by fluorescence and morphology under 20x magnification and the total cyst burden per brain was back-calculated.

2.3.3. Drug Screening

Parasites barcoded in a plate-based format were transferred to a v-well plate diluted in media with drug fragments at the appropriate concentration. Treatment was for 1 hour in an incubator at 37°C. Following treatment, drug media was removed, and parasites were resuspended in sterile PBS. Parasites were counted and diluted to a concentration of 1 million/mL. 200 uL was used for each infection and the remaining parasites were taken as the input sample. Infections were initiated through i.p injection of 200,000 parasites. Five days post infection parasites were harvested through PEC isolation using ice cold PBS and 5% FBS. Parasites were expanded in sterile *in vitro* culture with penicillin/streptomycin.

2.4. Analysis & Illustration

Schematics were created using Illustrator v22.1 (Adobe). General data storage, curation and/or statistical analyses was carried out in Microsoft Excel and GraphPad Prism 8.0. Nucleic acid sequences were designed and aligned using Benchling (www.benchling.com). Protein or nucleic acid quantity and purity was visualised using NanoDrop 2000C (Thermo Scientific) or 2200 TapeStation Analysis (Agilent) software. Images of DNA gels were acquired using BioRad ImageLab v6.1.0. NGS data were acquired using NextSeq Control Software v4.0 (Illumina), demultiplexed using CASAVA v2.17 (Illumina), and analysed using the Galaxy web server (www.usegalaxy.org). Chemical structures were routinely drawn in ChemDraw Professional v18.0 (PerkinElmer). Diversity filtering and physicochemical property analyses of compounds were performed using DataWarrior v5.5.0 (OpenMolecules) (Sander et al., 2015). Genetic selection bottlenecks experienced within the murine host were estimated by calculating changes in the relative frequencies of barcodes within dynamic *T. gondii* populations in relation to the starting population in the inoculum. The following equations were used to calculate chord distance:

$$\hat{F} = \frac{1}{k} \sum_{i=1}^{k} \frac{(f_{i,s} - f_{i,0})^2}{f_{i,0}(100 - f_{i,0})}$$

k = number of distinct alleles

$f_{i,0}$ = frequency of allele i at time 0

$f_{i,s}$ = frequency of allele i at sampling

$$N_b \approx N_e = \frac{g}{\hat{F} - \frac{1}{S_0} - \frac{1}{S_s}}$$

g = number of generations during competitive growth

S_0 = sample size at time 0

S_s = sample size at sampling

$$D_{ch} = \frac{2\sqrt{2}}{\pi} \sqrt{1 - \cos\theta}$$

D_{ch} = chord distance

k = number of distinct alleles

$$\cos\theta = \sum_{i=1}^{k} \sqrt{f_{P1,i}f_{P2,i}}$$

$f_{P1,i}$ = frequency of allele i in population 1

$f_{P2,i}$ = frequency of allele i in population 2

3. Barcoding

3.1. Introduction

Cellular barcoding is a methodology used to genetically identify individuals within complex populations. Short, neutral sequences of DNA are inserted into a genome of interest. This allows for the tracking of tagged cells and enables assessment of how environmental conditions impact their survival and replication. It is also possible to monitor the population dynamics of a particular organism or cell type. This has allowed researchers to understand how populations expand and contract as they move through different environments, and how this affects the genetic diversity of a population at a given time point[113]. This technique has been applied in the functional profiling of organism genomes to evaluate the impact of specific genes on molecular functions[123,124]. Cellular barcoding has also been used to facilitate multiplexed drug screening[117]. Using independent, identifiable populations, many different treatments may be tested concurrently with their individual effects observable through genetic sequencing of the persisting barcoded strains. DNA barcoding strategies have also been used in cancer research to identify patterns in immunoediting, which is the transformation of 'normal' cells into detectable tumour cells within populations as metastases form. With knowledge of these patterns, supplementary therapies can be deployed to limit this effect and improve efficacy of immunotherapy treatments[125]. Given the numerous documented use cases for cellular barcoding, its application in eukaryotic pathogens holds great potential to address previously challenging questions about host-pathogen dynamics. This approach opens up new avenues for investigating and unravelling the complexities of host-pathogen relationships.

3.2. Results

3.2.1. Molecularly Barcoding *T. gondii*

I aimed to develop a versatile method for barcoding eukaryotic pathogens. To accomplish this, I sought to insert a single barcode at a consistent position within a non-essential genomic locus and implement a selection strategy to enrich for barcoded cells. By utilising this approach, it would be possible to generate a population of barcoded parasites while eliminating unmodified, wild-type, parasites. I hypothesised that a CRISPR-Cas9 strategy could satisfy these requirements, with the double strand break introduced by Cas9 stimulating efficient HDR to mediate barcode insertion at a specific target locus[126–128]. I targeted the *UPRT* locus as successful barcode integration would result in disruption of the coding sequence, and as a result, provide barcoded strains with resistance to the pro-drug Floxuridine (FUDR) (Figure 10a)[129]. A prodrug is a pharmacologically inactive or less active compound that undergoes biotransformation in the body to be converted into an active drug. Once administered, the prodrug undergoes a chemical or enzymatic conversion, either spontaneously or facilitated by specific enzymes, to release the active drug molecule[130]. *T.*
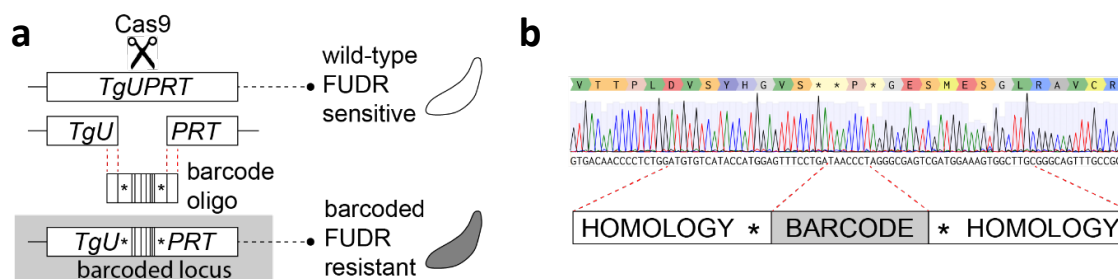
Figure 10 | *T. gondii* tachyzoites can be barcoded using a CRISPR-Cas9 system. (a) Schematic depicting the cellular barcoding strategy. Cas9 is targeted to the Tg*UPRT* locus using a specific guide RNA sequence. Cas9 endonuclease activity introduces double strand breaks, stimulating homology directed repair and insertion of a 'barcoding' oligo. Disruption of the Tg*UPRT* locus confers resistance to the drug FUDR enabling successfully transfected parasites to be selected for in cell culture. (b) Sanger sequencing chromatogram of a PCR amplicon derived from the Tg*UPRT* locus following barcode integration by HDR. The 'barcode' is six nucleotides long and flanked by two stop codons to prevent translation of barcoded region. There are homology regions of 24 nucleotides to e integration of the oligomer into the targeted locus by HDR.

gondii parasites deficient for NHEJ (RH*Δku80*) were used to increase recombination efficiency[131]. I designed barcoding oligonucleotides (oligos) that would destroy both the protospacer DNA sequence recognised by the CRISPR sgRNA and protospacer adjacent motif (PAM). This ensures single barcode integration by preventing cycles of gRNA recognition and Cas9 cleavage following the first integration event. The oligos consist of a six-nucleotide 'barcode' that serves as a genetic identifier, delimited by two stop codons. On either side of this barcode, there are 24 nucleotides that are homologous to the parent sequence. Homologous regions in a donor sequence are known to enhance recombination efficiency[132]. Barcode integration and the expected genomic modification were confirmed by Sanger sequencing of the *UPRT* locus in the drug-resistant parasite population (Figure 10b). These data indicate our barcoding strategy can be used to efficiently insert unique barcode sequences at a specific locus in *T. gondii*.

3.2.2. Implementation in a multiplexed, plate-based, format

Following confirmation (Figure 10b) that our strategy could be used to successfully integrate a single barcode, I sought to establish the amplicon derived BarSeq pipeline that would allow us to quantify the relative frequencies of many different barcodes within complex 'populations' of barcoded strains. I began by generating a library of 96 uniquely barcoded parasite lines using microcuvette-based transfection apparatus that allowed for 16 transfections to be performed in parallel. I envisaged that this multiplexed strategy would be useful for future chemical screening applications applied to each of the 96 uniquely barcoded strains prior to pooling (Figure 11a). I developed an NGS pipeline to quantify the frequency of individual barcodes after pooling the barcoded strains. Genomic DNA was extracted from the pooled
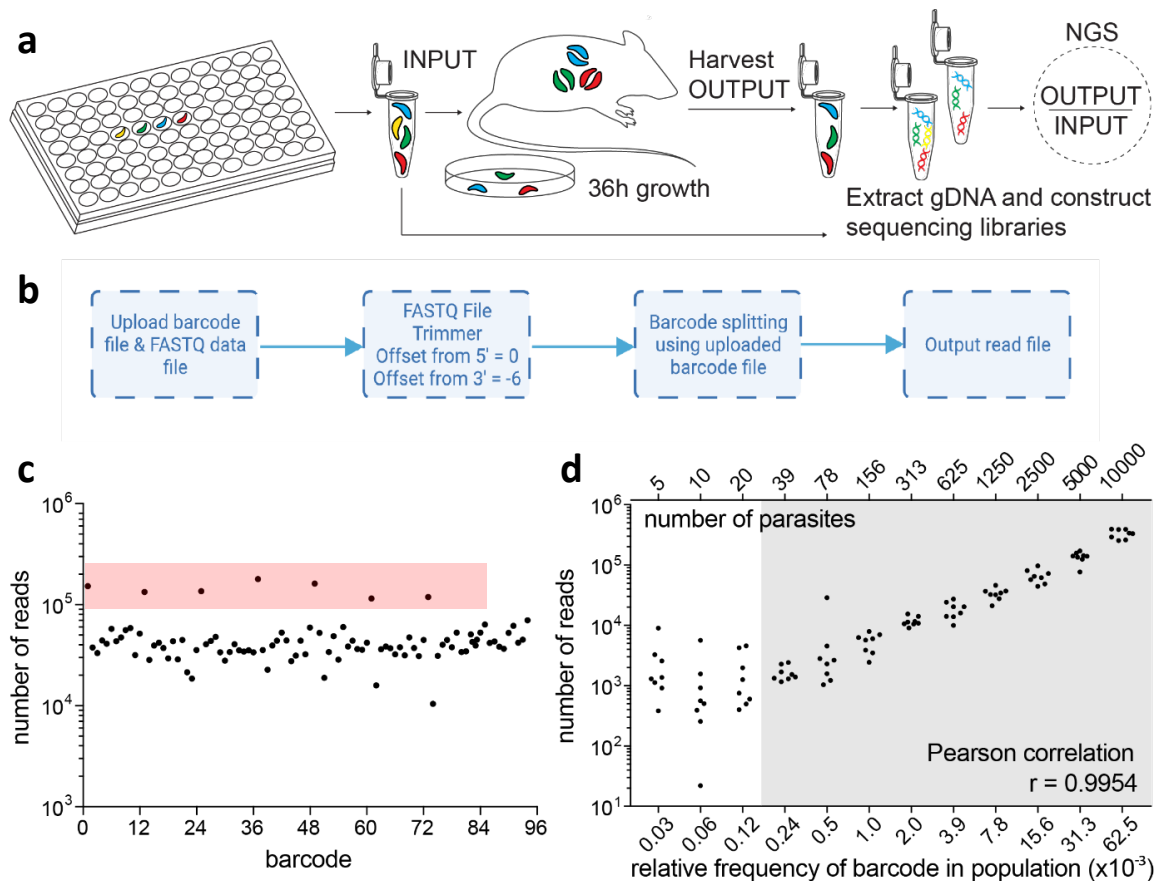
Figure 11 | Individual barcodes can be identified from a complex population of barcoded strains. Data presented in this figure was collected by members of the Child Lab prior to my PhD. (a) Schematic depicting approach to multiplex the barcoding workflow. Transfections were conducted in a plate-based format, with each well containing a unique barcoded parasite population. These populations were pooled, propagated *in vitro* or *in vivo* before being harvested and gDNA extracted for NGS. (b) Workflow for processing sequencing files using the Galaxy web application (c) Scatter plot showing number of NGS reads for each barcode. Highlighted in red are over-represented barcodes as a result of a pipetting error. (d) A serial dilution of known parasite numbers was conducted to assess correlation between parasite number and number of NGS reads. PCC analysis of number of parasites to number of NGS reads, $p < 0.001$ of samples in the shaded grey box.

parasite library, the barcoded region of the *UPRT* locus was amplified for Illumina sequencing.

To optimise the efficiency of the workflow I constructed my own NGS libraries as follows. Amplicons encompassing the barcode were generated through targeted PCR. The primers used in this reaction were not only integration specific but attached an adapter sequence to either end of the amplicon to facilitate indexing. Samples (e.g. replicates) were indexed using Illumina Nextera XT v2 primers, followed by magnetic bead purification. Purified samples were subsequently quantified using a Qubit fluorometer and their fragment sizes precisely determined using Agilent TapeStation. Indexed libraries were pooled at equal concentrations and sequenced on an Illumina MiSeq or NextSeq. NGS read data was initially demultiplexed, and then Fastq NGS read files manipulated and barcode representation was quantified using the Galaxy web application (Figure 11b)[133]. Using this pipeline, I was able to successfully identify all 96 uniquely barcoded strains within the pooled population. Due to pipetting error,

wells from one row of the 96-well plate over-represented in the pooled population. This illustrates the technical challenge associated with manual plate-based transfection methods (Figure 11c). Fortuitously, this error facilitated the observation that the NGS readout was highly sensitive to differences in barcode frequency in the pooled population.

We next tested if we could discriminate between groups of barcodes present at different frequencies within a complex population. To do this we conceived an experiment where a two-fold dilution series of a pooled parasite population was conducted. This facilitated comparison between sequencing read number and the known number of barcoded parasites in each sample and establish the relationship between the two. A positive correlation (PCC r = 0.9954) was observed between the number of reads (read output) and the number of parasites in the input to a resolution of 39 parasites/barcode or a relative frequency of 0.00024% within the population (Figure 11d). These data confirmed that barcode alleles could be identified and quantified, using NGS, in complex populations and establish the lower boundary for which NGS read number can be used as a reliable proxy for parasite number.
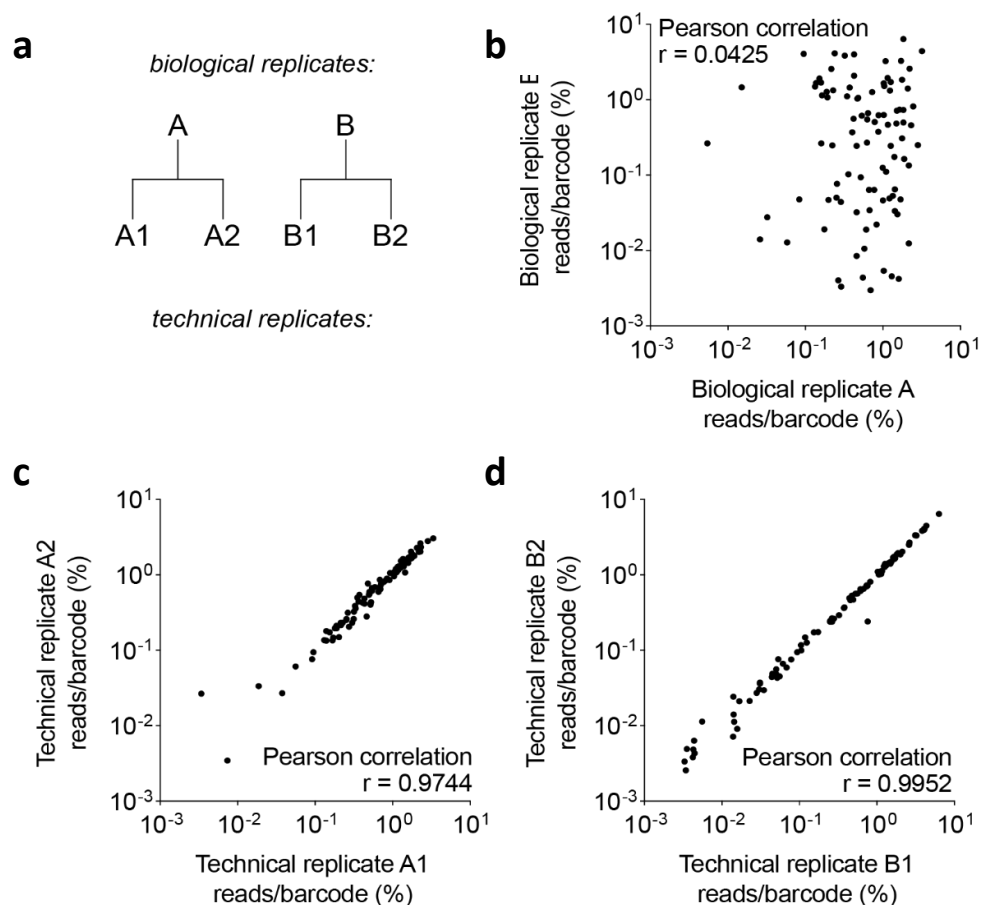

Figure 12 | Our NGS pipeline is sensitive to differences in barcode representation. (a) Schematic of strategy to probe variance between technical and biological replicates. (b) PCC analysis of NGS output reads from biological replicates A and B, r = 0.0425, n = 96, p < 0.001. (c) PCC analysis of NGS output reads from technical replicates A1 and A2, r = 0.9744, n = 96, p < 0.001. (d) PCC analysis of NGS output reads from technical replicates B1 and B2, r = 0.9952, n = 96, p < 0.001.
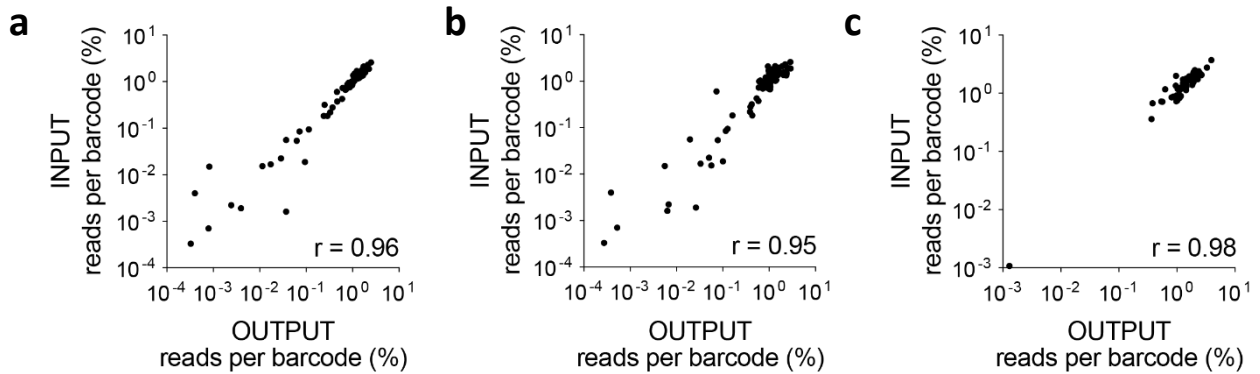
Figure 13 | Assessment of the stability of barcoded populations both *in vitro* and *in vivo*. (a) PCC analysis of NGS reads from input and output samples following one lytic cycle *in vitro*, r = 0.96, n = 96, p < 0.01. (b) PCC analysis of NGS reads from input and output samples following six lytic cycles *in vitro*, r = 0.95, n = 96, p < 0.01. (c) PCC analysis of NGS reads from input and output samples following 36 hours propagation *in vivo*, r = 0.98, n = 96, p < 0.01.

To assess the variation between various replicate types I compared NGS reads obtained from biological and technical replicates. I observed greater variation in barcode representation between biological replicate transfections than between technical replicates (Figure 12). This underscores the reproducibility of the NGS pipeline and highlights the variability of barcode representation across separate transfection experiments.

To test whether the relative frequencies of each of the 96 barcodes within a complex library was stably maintained *in vitro* the pooled library of barcoded parasites was serially passed through HFFs. Lysed-out parasite cultures were passaged and sampled every 36 hours for a period of six passages, equivalent to six lytic growth cycles (invasion, replication, egress), and barcode frequencies from each sample was quantified by NGS. Relative to the input, the genetic complexity of the barcode population *in vitro* remained stable throughout the experiment (Figure 13a and b, respectively). We next tested whether we could reliably propagate and recover the pooled barcode library *in vivo*. An inoculum of ~20,000 parasites was injected IP into C57BL/6 mice. After 36 hours, parasites were isolated from the peritoneal cavity by lavage. gDNA from harvested parasites was processed according to our library preparation protocol and sequence reads were and compared with the input population. A strong positive correlation (PCC r = 0.98, n = 96, p < 0.01) was observed between the input and output populations, demonstrating that the genetic complexity of the multiplex barcode

library was stable over the first 36 hours of *in vivo* infection (Figure 13c). A similar *in vitro* experiment was repeated over 28 days. Populations of parasites were passaged, and samples harvested once a week. These samples were processed and subsequently analysed by NGS (Figure 14). The population remained stable after 28 days of growth *in vitro* indicating that there were no subtle fitness defects associated with the integration of specific barcodes, with low-frequency barcodes stably maintained. Together these data indicated that barcoded
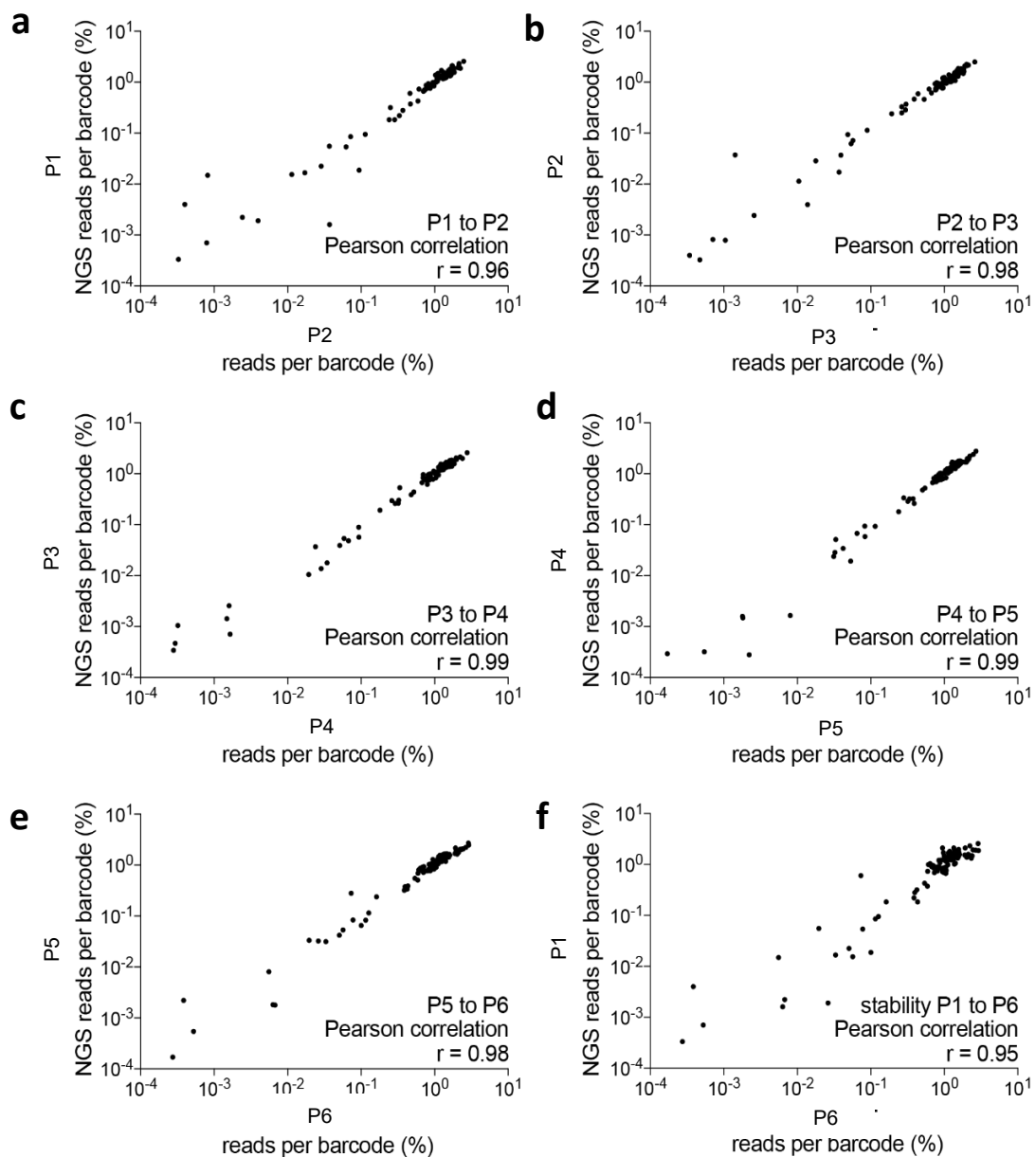

Figure 14 | Barcoded populations are stable *in vitro* over 28 days. (a) PCC analysis of output NGS reads following passage 1 and passage 2 *in vitro*, r = 0.96, n = 96, p < 0.01. (b) PCC analysis of output NGS reads following passage 2 and passage 3 *in vitro*, r = 0.98, n = 96, p < 0.01. (c) PCC analysis of output NGS reads following passage 3 and passage 4 *in vitro*, r = 0.99, n = 96, p < 0.01. (d) PCC analysis of output NGS reads following passage 4 and passage 5 *in vitro*, r = 0.99, n = 96, p < 0.01. (e) PCC analysis of output NGS reads following passage 5 and passage 6 *in vitro*, r = 0.98, n = 96, p < 0.01. (f) PCC analysis of output NGS reads following passage 1 and passage 6 *in vitro*, r = 0.95, n = 96, p < 0.01.

parasite populations remain stable *in vitro* and *in vivo*. Relative barcoded populations also remain consistent when harvested *in vivo* and expanded in *in vitro* culture.

3.2.3. Simultaneous transfection of many barcodes in a 'pool':

As noted in Figure 12b, a challenge associated with this plate-based transfection strategy was the variability in transfection efficiency between individual wells. To ensure the integration of a single barcode into the UPRT locus, I designed our CRISPR-Cas9 strategy to simultaneously delete the protospacer and PAM motifs recognised by the CRISPR sgRNA. Therefore, I hypothesised that a barcoded library of parasites could be generated through a single transfection using the Cas9-sgRNA plasmid and a pooled library of oligonucleotide repair templates (Figure 15a). This "one-pot" method was tested on type I RHΔ*ku80* and type II PruΔ*ku80* parasite strains in parallel using a mixed pool of the 96 barcode oligo repair templates. Enriched FUDR-resistant parasite populations were harvested, gDNA was isolated, UPRT locus amplicons were prepared, and NGS libraries were sequenced. From the sequence reads generated, all 96 barcodes were identified in the UPRT locus amplicon. Upon comparison, the two independently transfected parasite strains displayed correlated frequency distributions for all barcodes, except the least abundant ones, with a PCC of r = 0.70 (Figure 15b). This demonstrates the improved replicability of this method across transfection experiments, as compared to the previously described plate-based methodology. The use of this new approach enhances the consistency and reliability of results obtained from different transfection experiments, ensuring more robust and accurate data.

3.2.4. Improving Next Generation Sequencing Quality Through Dark-Cycling

 The Illumina NGS technology is optimised for diverse genetic libraries. Sequence diversity assists with differentiation between cluster fluorescence, particularly at the introductory nucleotides (Figure 9). Despite high diversity between the six nucleotide DNA barcodes, the
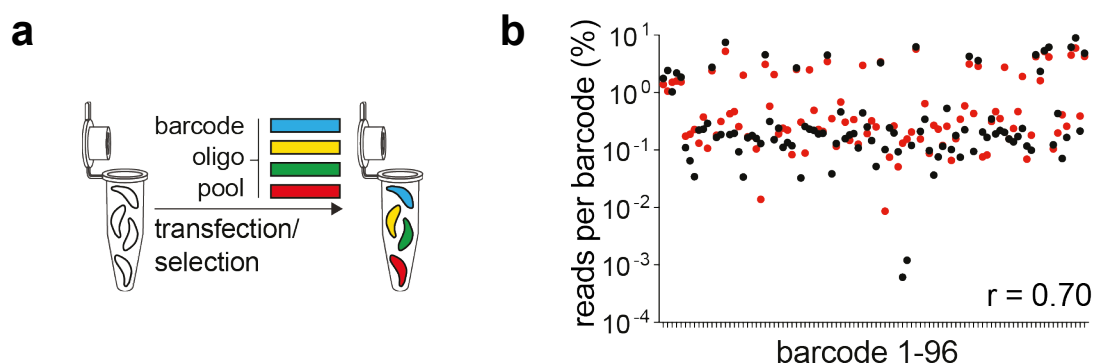
Figure 15 | Barcoded libraries can be generated via one-pot transfections. (a) Schematic of the one-pot transfection method to generate barcoded parasite libraries. (b) Overlay of the relative percent frequency of each barcode in an RHΔ*ku80* population (black dots) or a PruΔ*ku80* population (red dots) generated from a common pool of barcode oligos by one-pot transfection. The scatter plot is distributed according to barcode identifier (1–96), and PCC analysis of the two populations, r = 0.70, n = 96, p < 0.01.

homology regions and stop codons flanking the barcode are identical between samples. This led to poorly defined clusters and lower read depth in sequencing runs (~2,000,000 reads/sample). I worked in collaboration with Dr Laurence Game, MRC Genomics Lab and Illumina application specialists to employ a dark cycling technique to combat this. Dark cycling does not record the fluorescent signals of nucleotide clusters (Figure 9) until a user designated cycle. In essence, no data is recorded with regards to which nucleotides are present up until this point. Introducing 29 dark cycles in sequencing when reading from the 5′ of the amplicon, data is collected from the first nucleotide of the barcode. This region is highly diverse leading to better resolution of clusters and therefore more reads per run (Appendix 8.4).

3.3. Adapting CRISPR-Mediated Cellular Barcoding to Other Eukaryotic Pathogens

The population genetics of Protozoa are understudied and so to address this and investigate the wider utility of our tool, I sought to apply our barcoding method to other pathogens. To easily adapt our method, I identified criteria I considered to be instrumental to its success: 1) the cell system of interest should have a non-essential locus amenable to genetic manipulation using CRISPR-Cas9; 2) an NHEJ deficient system to drive efficient barcode integration by HDR; 3) an endogenously encoded negative selection marker to facilitate the selection of barcoded cells following integration. With this in mind, I identified systems that fulfilled these criteria.

3.3.1. *Trypanosome brucei brucei*

In collaboration with the Tiengwe Lab (Dept. Life Science, Imperial College), we used an established CRISPR-Cas9 system in *T. brucei* trypomastigotes targeted to the Tb*AAT6* locus, a single-copy non-essential gene that confers sensitivity to eflornithine[134] (Figure 16a). This technique has been used to induce point mutations indicating its capacity to be used from precision genetic mutations. Notably, *T. brucei* lacks the NHEJ machinery[135], leading to Cas9-
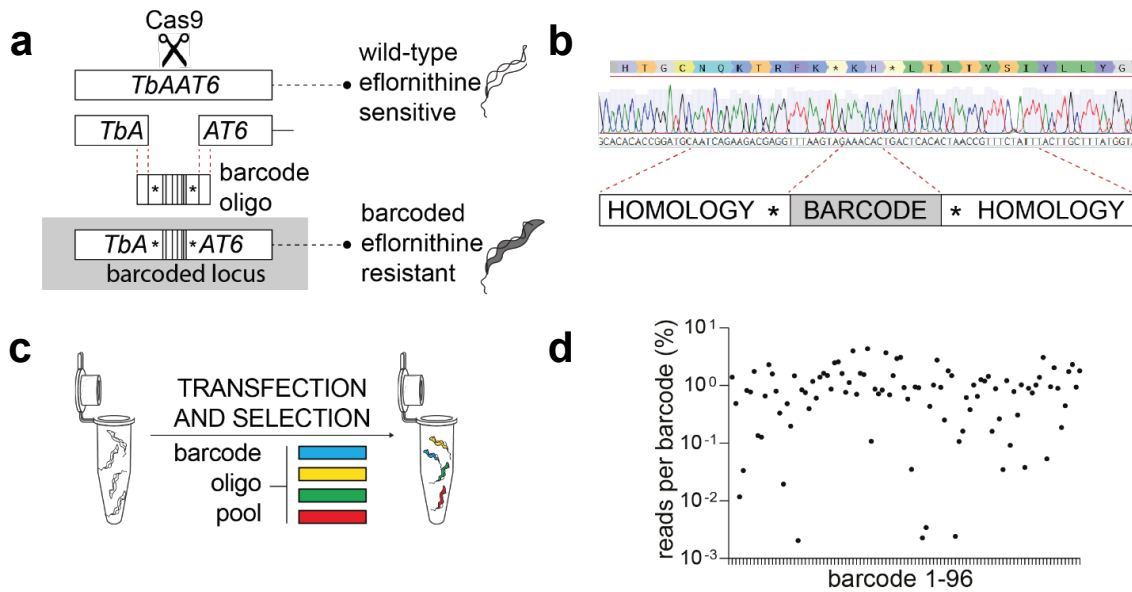
Figure 16 | *T. brucei* can be barcoded using the 'pooled' transfection technique. Transfections and cell culture in *T. brucei* were conducted by the Tiengwe lab. (a) A stable Cas9 expression cell line was established and transfected with a PCR amplicon encoding the guide to target Cas9 to the *TbAAT6* locus, along with the barcode oligomer library. Wild type *T. brucei* parasites are sensitive to the drug eflornithine however once the *TbAAT6* locus is disrupted they become resistant, enabling positive selection of successfully transfected parasites. (b) Sanger sequencing chromatogram of a PCR amplicon derived from the *TbAAT6* locus following barcode integration by HDR. The 'barcode' is six nucleotides long and flanked by two stop codons. There are homology regions of 24 nucleotides to stimulate integration of the oligomer into the targeted locus by HDR. (c) Schematic demonstrating a pooled transfection approach where 96 barcodes are combined at equal concentration and a sample of this is used in parasite transfections. (d) All 96 barcodes were represented in NGS data from a pooled transfection.

induced DSBs repaired through homologous recombination mechanisms. Our barcoding technique utilises this to achieve predictable integration of oligos.

To establish barcoding conditions in *T. brucei*, parasites stably expressing Cas9 were co-transfected with *AAT6*-targeting sgDNAs and single-stranded oligos with homology regions of 24, 40, 51 and 81 nucleotides. No integration was seen at the *AAT6* locus following these transfections (appendix). I hypothesised this lack of integration was due to using single-stranded repair oligos. To investigate this, we trialled identical transfection conditions with double-stranded oligos with homology regions of 24, 40, 51 and 81 base pairs[136,137]. Sanger sequencing confirmed the anticipated genomic rearrangement and barcoding of the *T. brucei* genome with double-stranded oligos with all homology arm lengths. Going forward double-stranded oligos with 24 base pairs of homologies will be used in *T. brucei* with experiments to minimise cost of barcode libraries (Figure 16b). Once confirmed that *T. brucei* was amenable to our barcoding methodology we conducted proof-of-concept studies to confirm suitability of the one-pot transfection technique. 96 double-stranded barcoding oligos were combined at equal concentrations (Figure 16c). A sample of this pool was used to transfect bloodstream *T. brucei*, gDNA was extracted and amplicons analysed by NGS. All 96 barcodes were

successfully identified by sequencing confirming that our barcoding methodology is transferable between these pathogens[1] (Figure 16d). As with *T. gondii*, barcodes represented at low frequencies are likely a result of lower representation in the pool used for transfection. This is unlikely to confound future studies as representation relative to input samples is calculated.

3.3.2. *Cryptosporidium parvum*

In collaboration with the Sateriale Lab (Francis Crick Institute, London), we attempted to adapt our method to *Cryptosporidium parvum* (*C. parvum*). Double- and single-stranded oligos were co-transfected with sgDNA targeting the thymidine kinase (*Cp*TK) locus. Similarly to *T. brucei*, *C. parvum* naturally lacks the NHEJ machinery[138]. Maintaining *in vitro* cultures of *C. parvum* is challenging, as the parasite can only be sustained in a non-replicating state in cell culture for approximately 72 hours. Additionally, the *Cp*TK locus is currently the only identified
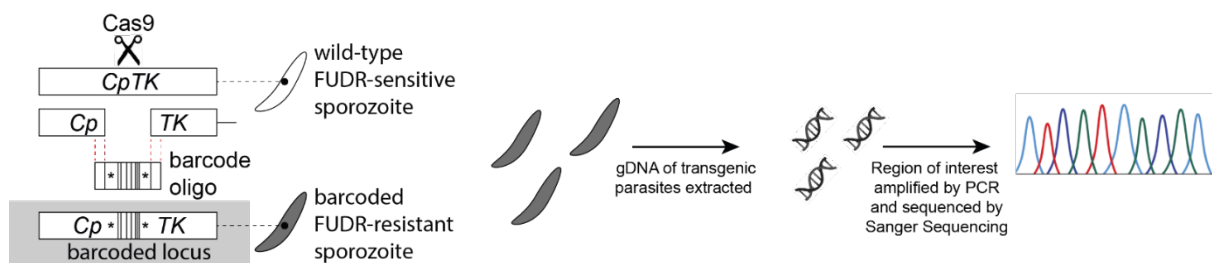

Figure 17 | Barcoded parasites could not be isolated following transfection in *C. parvum*. Transfections and cell culture in *C. parvum* were conducted by the Sateriale lab. Using a CRISPR-Cas9 system targeted to the Cp*TK* locus in *C. parvum* a double-stranded break was induced and using electroporation we introduce our barcoding oligonucleotide which is integrated through homology-directed repair (HDR). Wild type *C. parvum* parasites are sensitive to the drug FUDR however once the Cp*TK* locus is disrupted they become resistant, enabling positive selection of successfully transfected parasites.

selectable marker in *C. parvum*, its disruption providing resistance to paromomycin[139,140]. Unfortunately, parasites used in the Sateriale Lab were previously under paromomycin selection for other experiments. However, subsequent work in the lab showed that disrupting this locus also provides resistance to FUDR, similar to *T. gondii* (Figure 17). There are several limitations to this approach when working with *C. parvum*, including its limited, 72 hours *in vitro* viability, and the inability to use FUDR for *in vivo* selection due to host toxicity, unlike paromomycin. We attempted to apply our barcoding technique to *C. parvum* using double- and single-stranded oligos with homology regions ranging from 24 to 81 base pairs, but no integration was detected by Sanger sequencing or NGS.

3.4. Discussion

Cellular barcoding has provided critical insights into the infection biology of viruses such as poliovirus[112,141] and bacteria such as *Salmonella*[113–115,142] and *Escherichia coli*[143]. I anticipate

barcoded *T. gondii* strains will have similarly broad applications. While barcode sequencing (bar-seq) strategies have been extensively used for phenotypic screens of eukaryotic pathogens[144–149], the use of cellular barcodes to investigate the within-host infectious population structure of wild-type eukaryotic pathogens has so far been limited. A notable example is the use of eight uniquely barcoded *T. brucei* strains to investigate colonisation of the tsetse fly and subsequent survival of these strains within the bloodstream of a murine host[150]. Our simple approach allows for an increased number of cellular barcodes, providing further opportunities for *T. brucei* researchers to investigate colonisation within specific infectious niches, such as the subcutaneous fat and bone marrow[151].

An intriguing observation in these initial studies was the distinct oligonucleotide requirements for successful barcode integration in *T. gondii* and *T. brucei*. While both species only required 24nt homology arms, *T. brucei* necessitated the use of double-stranded templates, whereas in *T. gondii* I employed single-stranded templates. The success of relatively short homology arms on the donor template can be attributed to the fact that the barcode insert itself is only six nucleotides long. Previous studies have demonstrated that donor templates with approximately 30nt homology arms exhibit enhanced recombination efficiency for short insertions[152]. Interestingly, single-stranded repair templates generally display higher efficiency in HDR compared to their double-stranded counterparts[153,154]. Single-stranded DNA templates are less stable than double-stranded DNA templates[155] and this discrepancy in stability could account for the difference in success between *T. brucei* and *T. gondii*. It is possible that if the HDR process takes longer in *T. brucei* than in *T. gondii*, the donor templates may begin to degrade before their successful insertion can occur.

Our approach allows for the creation of multiple independently barcoded parasite populations, enabling researchers to track the effects of specific treatments or conditions on parasite fitness. Experiments can be multiplexed, and quantitative data deconvoluted from the sequencing readout. Our proof-of-concept data demonstrates the stable representation of barcodes within a population across multiple lytic cycles in the absence of selection pressures, confirming that barcodes have no impact on parasite fitness and can function as neutral alleles. The versatility of our approach permits researchers to work with individual barcoded strains in isolation before pooling them via plate-based library generation, or with complex libraries of barcoded strains generated through our one-pot approach. This study limited the number of barcodes used to 96 to simplify the workflow, reduce cost and facilitate experiments being conducted in a plate-based format. To determine the true limit on barcode number, reducing dilutions of barcoding oligo could be trialled in transfections until integration is no longer detected. From this, it would be possible to extrapolate the number of barcodes that may be pooled and successfully simultaneously transfected.

Collaborating with the sequencing lab, I optimised the NGS process for our sample type to obtain high-quality and representative data. The simple oligo barcoding strategy I developed can be applied to systems accessible to CRISPR that have endogenous or engineered negative selection markers. I believe this novel technique has the potential for a wide range of applications, from the interrogation of within-host population dynamics[110,114] to multiplexed *in vivo* drug screening[117].

4. Using Cellular Barcoding to Interrogate Population Dynamics in *T. gondii*

4.1. Introduction

A vast array of factors can influence the within-host dynamics of pathogenic infections, including but not limited to the availability of suitable infection niches, host immune response, and the pathogen's ability to replicate and distribute itself[156]. Interrogating the within-host molecular pathways and processes that enable the initiation and proliferation of pathogenic infections is a challenging logistical task. Although bottlenecks experienced by viral populations are well studied[116,157], quantifying the effective population size *in vivo T. gondii* infections of has proved challenging. The absence of effective methods for tracking genetic lineage has limited our ability to not only quantify absolute parasite numbers in infections but also to determine if the genetic diversity in the initial inoculum populations remains accurately represented throughout the course of chronic infection. It is unclear whether there is a stochastic reduction and subsequent expansion in all pathogen genotypes during an infection, or whether a limited number of genotypes successfully survive the bottleneck, giving rise to a clonal expanding population[158,159]. WITS have been used to create distinguishable pathogens that are genetically equivalent and has been successfully applied to study bacterial infections. However, while informative, studies that seek to use WITS are often limited by a small number of genetic markers. The limited number of markers available in these studies reduces their resolution, as bottlenecks can only be accurately quantified if they fall within the range of available markers. This limitation is unproblematic in infections where the effective population size is small; however, in diverse infections, bottlenecks cannot be determined[114,142].

In studies with *Vibrio cholerae* (*V. cholerae*), population analysis frameworks, structured sets of methodologies, tools, and statistical techniques used to study and analyse populations, were combined with cellular barcoding technologies to generate the sequence tag-based analysis of microbial populations (STAMP) methodology[110]. This method quantifies the relative abundance of each artificial allele within an inoculum population, establishing a baseline allele frequency. The abundance of alleles can then be quantified in samples taken from different niches at varying infection timepoints and compared to this baseline, allowing for the estimation of how many bacteria from the inoculum have progressed the infection. This

number is referred to as the founding or bottleneck population size (N_b) and gives an indication of the stringency of host barriers to infection. Further analysis of the difference between the infection niche and inoculum populations facilitates the calculation of chord distance (Section 2.4), which measures the genetic distance between a sample and a reference population. It provides a measurement of genetic divergence between two populations expressed between zero and one, where 'zero' indicates that the genetic structures of the two populations are identical and 'one' indicates maximum genetic divergence[160].

Although the tissue distribution of the *T. gondii* acute and chronic infection is well documented and discussed in length in section 1.1.2. It is unclear whether colonisation of tissues throughout the host results in a genetic bottleneck. Understanding *in vivo* bottlenecks can be advantageous in developing drug targeting strategies. Exploiting the knowledge of bottlenecks can help identify critical stages of parasite growth and replication where they are subject to enhanced host immune pressure. By targeting these specific stages, inhibition of parasite growth becomes more effective, increasing the likelihood of successful elimination of the infection. This approach allows for the design of interventions that take advantage of the host immune response to maximise the efficacy of drug treatments. Transmission blocking therapies as an elimination strategy for malaria have become widely investigated[161–163]. The lack of genetic diversity in the sexual and mosquito mid-gut phase combined with significantly lower parasitaemia than in the human blood-stage form make this a more tractable target for complete inhibition[164,165]. By using techniques similar to those mentioned above, particularly analyses carried out by Abel et al., combined with our barcoding methodology, I investigated whether parasites encounter stringent bottlenecks as they colonise the host.

4.2. Results

4.2.1. Cellular barcodes reveal the population structure of a *T. gondii* infection *in vivo*

The transition from acute to chronic *T. gondii* infection corresponds with the spatial redistribution of parasites to skeletal muscle and the CNS[166]. Tachyzoite differentiation into cyst-forming bradyzoites accompanies this transition, and these bradyzoites are necessary for parasite transmission[167]. The restrictive nature of the BBB is well-documented in other infection models. For instance, in poliovirus infections, the BBB serves as the most stringent genetic bottleneck[141]. The BBB is hypothesised to be one of the principal bottlenecks experienced by *T. gondii* during the *in vivo* host infections[168].

We first tested whether diluting the PruΔ*ku80* barcode library to a non-lethal inoculum dose influenced the population structure in a way that could confound the interpretation of *in vivo* experiments. In order to achieve chronic *in vivo* infections, an appropriate inoculum dose is essential. If the dose of parasites is too low, the murine innate immune system will clear the infection, while if it is too high, the acute infection can result in fatality[169].

To do this, a sample of ~2,000,000 viable tachyzoites in a 2 mL suspension was diluted 2-fold to give 500,000 parasites/1 mL suspension. Three samples of 80 µL (~37,000 parasites) were taken from this stock solution (confirmed by plaque assay) and plated on HFFs to re-expanded for five days in tissue culture. Following this period of growth parasites were harvested, their gDNA was extracted, and the barcoded Tg*UPRT* locus was isolated and amplified using PCR. The resulting amplicons were then analysed using NGS. In each inoculum sample, all 96 barcodes were detected by NGS, and pairwise comparisons of each sample were strongly correlated (PCC r = 0.98). This confirms that dilution and expansion of
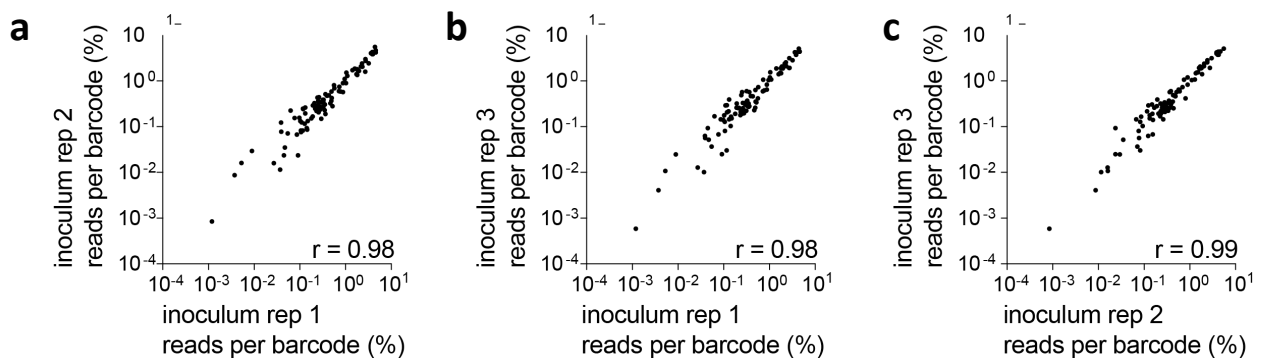

Figure 18 | *In vitro* expansion of parasites harvested from mice does not affect relative percentage representation of barcodes. (a–c) One-pot transfections using PruΔ*ku80* were diluted to a founder population of 37,000 parasites and re-expanded in HFFs prior to NGS. All 96 barcodes were identified in each sample, and the relative percentage frequency of barcodes was highly correlated in pairwise comparison for each inoculum sample: (a) inoculum 1 versus 2, PCC r = 0.98, n = 96, p < 0.0001 (two-tailed); (b) inoculum 1 versus 3, PCC r = 0.98, n = 96, p < 0.0001 (two-tailed); (c) inoculum 2 versus 3, PCC r = 0.99, n = 96, p < 0.0001 (two-tailed). PCC values provided on scatter plots indicate degree of correlation between populations being compared.

barcoded parasite lines results in consistent representation of each barcode and doesn't change their relative representation within the population as a whole (Figure 18a-c).

To investigate the colonisation of the murine host by *T. gondii*, an *in vivo* experiment was designed to observe the progression from the initial acute infection in the peritoneum to the chronic infection in the brain. gDNA extracted from 2 million parasites was sequenced to ascertain barcode representation prior to infection (Figure 19b). In order to assess the stability of the parasite population during the early acute infection, parasites were isolated from the peritoneal cavity of three mice at 48 hours post-infection and subsequently expanded *in vitro*.

The relative frequency of barcodes in each peritoneal isolate was distinct for all individual animals (Figure 19c). This observation reinforces a model in which the initial selective immune pressure encountered during the onset of acute infection is unpredictable and likely stochastic. It highlights the importance of acknowledging each host organism as a distinct environment. Following a 28-day period, the brains of 14 mice with chronic infection were isolated, and parasites were expanded *in vitro*. Surprisingly, the majority of barcodes were detected in each host brain, indicating minimal founder effects on the genetic diversity of the parasite population colonising the CNS (Figure 19d). The cumulative extinction frequency of barcodes across all
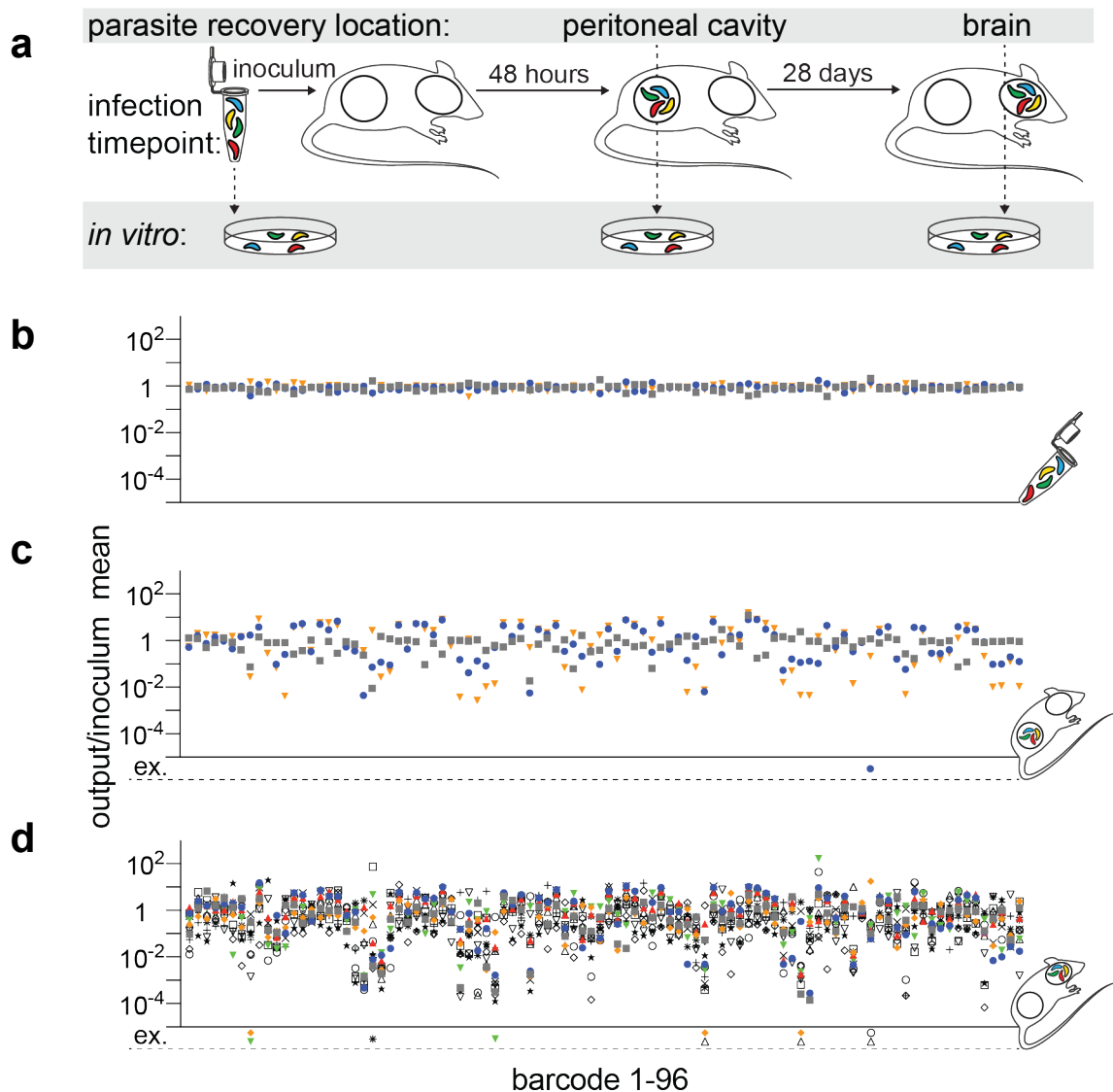
Figure 19 | The murine host brain is permissively colonised by *T. gondii*. (a) CBA/J mice were infected with an inoculum of 37,000 tachyzoites injected intraperitoneally. The infectious population structure was monitored in peritoneal exudates at 48 h post-infection (early acute) or in the brain at the onset of the chronic phase (28 days post-infection). Parasites were re-expanded in tissue culture prior to NGS sequencing. (b–d) Scatterplots represent individual barcode frequencies in each sample relative to the mean of the inoculum (b) in the inoculum n = 3 replicates, (c) the peritoneal cavity at 48 h n = 3 mice, and (d) the brain at 28 days n = 14 mice. Barcode extinctions (ex.) are indicated below the x axis in the corresponding position to the absent barcoded strain. Each symbol represents an individual inoculum or mouse.

14 mice was 0.007 (10 out of 1,344 total barcodes), and the lost barcodes were typically found at low abundance within the initial inoculum. It is noteworthy that these extinction events were observed in the NGS runs with the lowest total read counts, implying that the extinctions might be attributed to reduced sequence sampling depth rather than the actual absence of the specific barcoded parasite(s) in the brain. These findings collectively suggest that any selection bottleneck faced by parasites during brain niche colonisation must either be absent or broader in scope than can be measured with the quantity of markers used in this experiment.

4.2.2. How host genetic background influences the dynamics of parasite brain colonisation over time

To understand how the brain colonising population changed over time, we conducted a further *in vivo* infection study to assess barcode representation at 72 hours, one month and three months post-infection. To do this, we infected Swiss Webster mice, which are outbred to maximise genetic diversity and genetic heterogeneity, and inbred CBA/J mice[169] (Figure 20a). The mean parasite cyst burden at one-month post-infection did not show a significant difference between CBA/J mice and Swiss Webster mice. This suggests that a comparable number of parasites were able to access the brain niche, and the genetic background of the host had minimal impact (Figure 20b). As is documented in other studies, cyst burden declined over time, but the reduction in mean cyst number was similar between mouse genotypes[169,170]. To evaluate the parasite population structure within these hosts, I used Cavalli Sforza's chord distance calculation (Section 2.3)[160]. Chord distance is derived from allele frequency data. It represents the genetic divergence between populations or individuals by comparing the frequencies of genetic markers or alleles across different populations Although it does not allow for precise quantification of the exact width of bottlenecks, as N_b does, it can be used to deduce alterations in the genetic composition of a population from point A to point B. In CBA/J mice, the inoculum and peritoneal exudate populations, harvested 72 hours post-infection, in CBA/J were separated by a smaller genetic distance (~0.5) than the the one-month brain samples (~0.7). This distance continued to grow over time with the greatest distance observed between the inoculum and parasites harvested from the brains of CBA/J mice three months post-infection. This trend was also observed in the outbred Swiss Webster mice (Figure 20c and d). In contrast, the parasite population structure showed higher variability across individual Swiss mice at each chronic time point compared to CBA/J mice. Unlike CBA/J mice, Swiss mice exhibited similar median chord distances between the inoculum and the brains at one- or three-month post-infection. Collectively, our data revealed that the murine brain exhibits an unforeseen permissiveness to colonisation following intraperitoneal (IP) infection.
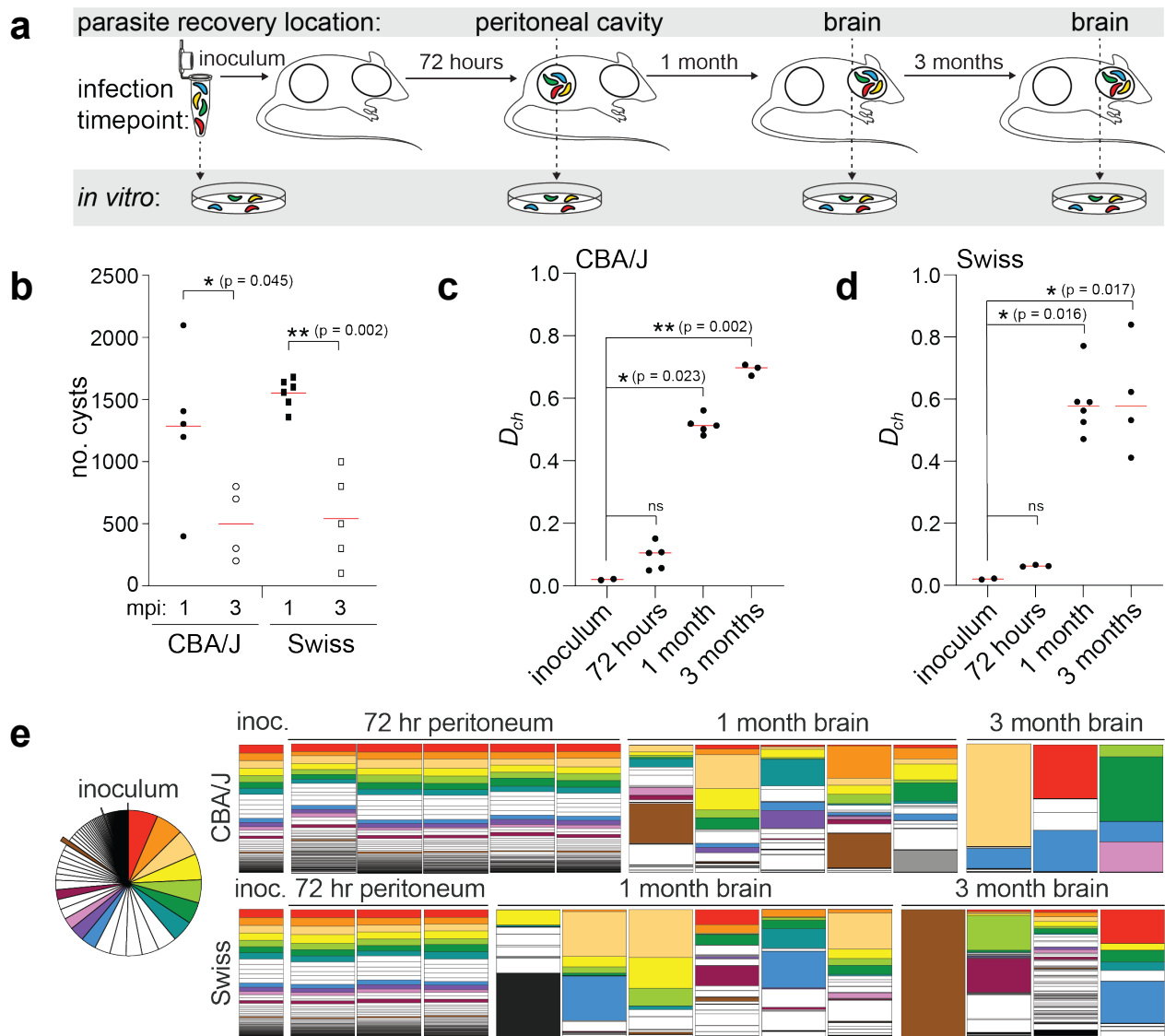
Figure 20 | The population structure of the *T. gondii* chronic infection is dynamic within the murine host brain niche. (a) Inbred CBA/J and outbred Swiss Webster mice were intraperitoneally injected with 12,000 tachyzoites and the *T. gondii* population structure was evaluated at an early acute (72 hours in peritoneal exudate), early chronic (one month in the brain), and late chronic (three months in the brain) infection time points. (b) Bradyzoite cyst counts from *T. gondii* infected brains isolated from CBA/J and Swiss Webster mice at one and three-months post-infection (mpi). Significance was tested using one-way ANOVA. (c and d) Scatterplot of chord distances (D_{ch}) calculated from barcode representation in the inoculum relative to 48 hour peritoneal samples, 1-month post-infection brains, or 3-month post-infection brains in (c) CBA/J or (d) Swiss Webster mice. Significance relative to inoculum was tested using Kruskal-Wallis one-way ANOVA, with the Mann-Whitney test used for pairwise comparisons. (e) Parts-of-whole charts representing the relative frequency of each barcode within the mouse samples. Barcodes are ranked in descending order of abundance in the inoculum mean (inoculum mean n = 3, represented as a pie chart and bar charts labelled "inoc."). A colour was assigned to a barcode if it was the dominant barcode in any brain sample or if it represented greater than 20% of all reads in any brain sample. Each colour is unique to one barcode. Any barcode that dominated a mouse brain and was represented in the bottom 25% of inoculum reads is outset in the inoculum pie chart for clarity and shaded brown, grey, or black in parts-of-whole charts.

Furthermore, the parasite population structure demonstrated dynamic changes within individual hosts.

58

I conducted an analysis of our data to determine whether the most abundant barcoded lineages in the inoculum had a competitive advantage for long-term persistence. In CBA/J mice, the dominant barcode alleles at three months post-infection corresponded to parasite populations that were among the top 35% of reads (top six most abundant barcodes) in the inoculum. While a similar trend was observed at one month post-infection, it is noteworthy that several CBA/J mice harboured a dominant barcode lineage that was represented in the bottom 25% of reads in the inoculum. Consistent with the findings from the chord distance analysis, individual Swiss mice exhibited diverse dominant population structures at three months post-infection. While the dominant barcodes in the Swiss brains often originated from lineages represented within the top 75% of reads in the inoculum, there were instances where this was not the case (Figure 20e). As an illustration, one mouse was primarily dominated by a barcode that had a low frequency in the inoculum, whereas another mouse displayed a population structure characterised by a more even distribution of barcode frequencies compared to the inoculum or peritoneal parasite isolates. This heterogeneity was similarly observed in the dominant barcodes within the Swiss brains at one-month post-infection, indicating that even strains with relatively low frequencies can establish and sustain persistent infections in a manner that is likely influenced by unpredictable factors and host genotype.
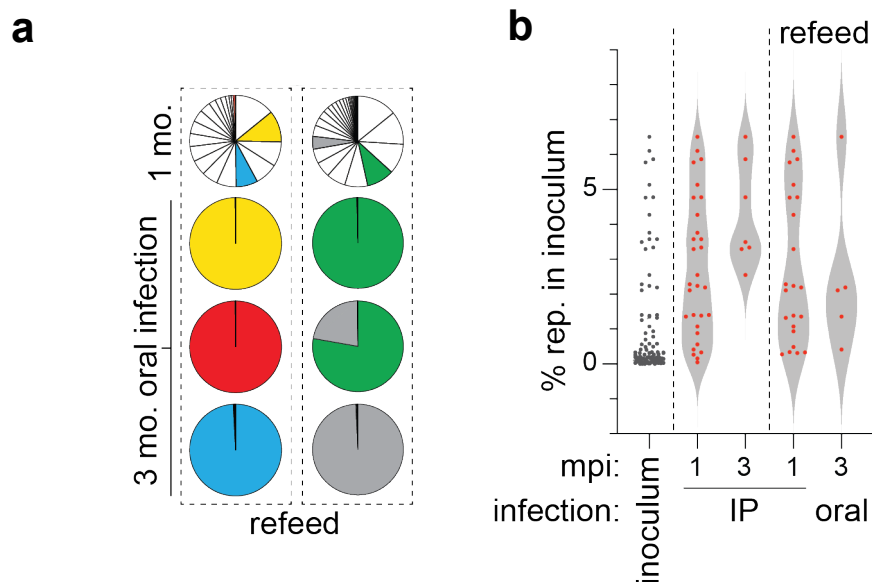
Figure 21 | The population structure of *T. gondii* chronic infection in the murine host has reduced genetic diversity following oral infection vs. inoculation via IP injection. Brain isolates harvested in the experiment in Figure 20 at one month post infection (mpi) were used to orally infect CBA/J mice (n=6). Brains from these mice were harvested three months post infection and relative barcode representation was quantified by NGS. (a) Coloured segments reference a unique barcode identity so representation in the inoculum can be compared to that at three mpi. (b) Individual barcodes are represented as red dots to demonstrate how persisting barcodes from various mpi and infection initiation routes were represented in the inoculum.

As an auxiliary experiment we looked to track population dynamics through a pseudo-natural infection. To do this, a sample of brains harvested at one month were used to orally infect CBA/J mice. In nature, one route of infection is the ingestion of tissue infected with encysted bradyzoites. At three months post-infection in five out of six orally infected mice, only one barcode was significantly represented with the populations being almost clonal (Figure 21a). Three months after oral infection, a variety of barcodes from the original inoculum, irrespective of their percentage representation, were still detectable, showing no bias towards those with high representation. This differs from the results seen three months after IP infection, where the persistent barcodes were those that had a high representation in the inoculum (Figure 21b). These data suggest that in orally initiated infections parasites encounter stringent bottlenecks to genetic diversity before colonising the CNS.

4.3. Discussion

While it is appreciated that knowledge of within-host pathogen dissemination patterns can aid in the understanding and treatment of disease, these studies have predominantly been confined to the study of viruses and bacteria[171,172]. In this work we expand *in vivo* population genetics studies to protozoa.

T. gondii strategically exploits brain residency as a means to evade the host immune response and enhance the likelihood of transmission through predation. Feline consumption of high-fat, energy-rich, infected neuronal tissue from prey, such as mice, offers *T. gondii* a pathway back into the definitive host[19,173]. Employing a method to avoid a stringent genetic bottleneck when colonising the brain niche would be in line with this evolutionary strategy, allowing for maximum transmission of genetic diversity into the feline host and contributing to recombination in the subsequent sexual cycle. In order to infect a wide range of warm-blooded animals[174], *T. gondii* must adapt to frequent and unpredictable selection environments within intermediate hosts. The diverse set of environments encountered may favour different phenotypes at different times. Consequently, maintaining genetic diversity and phenotypic plasticity is likely advantageous for the parasite, ensuring its survival across various future host species.

A key role of tissue barriers is to restrict pathogen access to the underlying tissue niche[20]. Surprisingly, our research findings indicate the occurrence of multiple distinct colonisation events by *T. gondii*, where the parasite enters the brain parenchyma from circulation. This contradicts the notion of rare brain invasions followed by expansion within the niche. If few parasites breached the BBB and then multiplied, I would anticipate observing genetically identical bradyzoites. However, our observation of diverse barcodes suggests the existence of multiple instances of BBB traversal by parasites. Chord distance analysis revealed a change in the genetic structure of the parasite population in the CNS at one-month

post-infection. In both CBA/J mice and Swiss mice, the dominant barcodes in the brain following IP infection tended to be those most frequently observed in the inoculum, but most barcodes were still detected, indicating limited selective pressure imparted by the BBB.

I sought to more precisely quantify the founder population of the CNS infectious niche using the STAMP methodology (Section 2.4)[110]. However, I encountered several obstacles when attempting to apply STAMP to our experiments. I used substantially fewer markers than were used in the original STAMP publication. Although this did not seem immediately problematic, on further interrogation it became clear that the maths employed in the calculation of N_b equated a total number of barcodes of ≤50, to zero. As I began with only 96 markers this meant anything more than an approximately 52% reduction in barcode representation could not be quantified. It is also plausible that the bottleneck imposed by the BBB is broader than what I could measure with the limited number of markers in this study. By increasing the number of markers used, I would have the opportunity to test this hypothesis and more precisely determine the extent to which the BBB influences the genetic diversity of infectious populations within the brain parenchyma.

Further considerations when interpreting this study include that the disruption of the Tg*UPRT* locus has been shown to decrease cyst burden[175]. It is important to note that this may have led to an underrepresentation of barcode diversity in the CNS in our study. However, insertion of barcodes at the same genetic locus in all parasites acts as an internal control. Additionally, it is important to acknowledge that our data represent two differentiations: tachyzoite-to-bradyzoite *in vivo*, followed by bradyzoite-to-tachyzoite during *in vitro* expansion. This presents the opportunity for an artificial bottleneck to be imposed, however in preliminary experiments illustrated in figure 18, incidence of this was negligible. We used IP infection to precisely quantify barcode frequency and coverage across the inoculum dose. The inoculum dose was selected to ensure coverage of all barcodes, accounting for possible differences in tachyzoite viability *in vitro* versus *in vivo*. Future studies could include more in-depth spatial analysis of brain colonisation. Although I have learned that the BBB exhibits relatively permissive characteristics during *T. gondii* infection, it remains unknown whether certain brain regions are more susceptible to invasion. By micro-dissecting brains obtained from subsequent experiments, it may be feasible to ascertain whether barcodes cluster in specific niches or are evenly distributed throughout the infected brain, thus shedding light on the spatial dynamics of infection[176,177].

In future studies, the use of fluorescent barcodes could be considered to enhance spatial dynamics analysis. This approach would enable real-time visualisation of parasite dissemination within the host, providing valuable insights into the routes of infection and identifying specific locations that are particularly susceptible to parasite colonisation. By tagging parasites with fluorescent barcodes, imaging techniques could be used to track their

movement and distribution in live hosts. This real-time visualisation would offer a dynamic understanding of the spatial dynamics of infection, allowing for the identification of key anatomical sites or tissues where the pathogen tends to localise or replicate. This information can be valuable for understanding the mechanisms of pathogen spread, identifying potential targets for intervention, and gaining insights into host-pathogen interactions. Overall, the use of fluorescent barcodes would enable investigation of the spatial dynamics of infection in a more dynamic and visual manner, providing a deeper understanding of how pathogens disseminate throughout the host and identifying locations that may be particularly susceptible to infection.

The etiological agents of human trypanosomiasis, *T. brucei rhodesiense* and *T. brucei gambiense,* are closely related to *T. brucei brucei*, the other pathogen for which we have employed our cellular barcoding methodology. It is worth noting that both *T. brucei rhodesiense* and *T. brucei gambiense*, also invade the brain and other tissue niches[178]. In future studies, it would be intriguing to investigate the within-host population dynamics of *T. brucei* and determine whether the BBB exhibits a similar level of permissiveness in this case. It is important to highlight that human blood serves as a major reservoir for *T. brucei*, so the colonisation of the brain would not be considered within the same context of bet-hedging as observed with *T. gondii*.

5. Cellular Barcoding Facilitates Multiplexed *In vivo* Drug Screening

5.1. Introduction

Phenotypic Drug Discovery (PDD) involves evaluating drug candidates by observing changes in disease phenotype associated with a specific disease. While target-based approaches have been traditionally favoured, PDD has gained popularity in recent years. However, early-stage *in vivo* drug screening remains challenging. The evaluation of a relatively small number of compounds requires a large number of animals, making the process time-consuming and resource intensive[179]. However, the study of host-pathogen interactions in infectious diseases is crucial for drug development[180–182]. Therefore, there is high demand for techniques that enable cost-effective and resource-efficient interrogation of these interactions. PDD interrogates the target system in its biological context. In target-based drug discovery compounds are screened against the target in isolation which may not be comparable to the target in its *in vivo* environment.

With the cellular barcoding technique developed in this study, our objective was to create a platform for multiplexed *in vivo* phenotypic drug screening. Previous studies have highlighted the significance of certain *T. gondii* proteins exclusively in *in vivo* survival[183]. Based on this knowledge, I hypothesise that critical host-pathogen interactions, which cannot be effectively studied *in vitro*, necessitate *in vivo* drug screening. Precedent for the use of barcoded cells in multiplexed drug screening was set in work with pancreatic cancer cell lines with drugs interrogated with regards to their effectiveness at preventing metastasis[117].

I hypothesised that through iterative screening and non-selective elaboration of hit molecules I could develop a potent drug molecule with demonstratable phenotypic effect. To do this I first screened compounds from our ~80-compound library sampling diverse chemical space. Following identification of hit molecules in this screen, similar but elaborated compounds were identified in a library of ~10,000 covalent fragments purchased from Enamine. These would then be screened for increased efficacy and specificity.

5.2. Results

5.2.1. *In vitro* & *in vivo* Controls Development

To demonstrate the ability to distinguish between killed and healthy parasites in an NGS readout it was first important to establish a suitable positive control (Figure 22a). To do this I conducted plate-based transfections to generate 96 barcoded populations. I treated a third of the plate with 4% paraformaldehyde (PFA), a commonly used fixative for *T. gondii*, as the positive control for successfully killed parasites, a third of the plate with DMSO, as the vehicle control, and left a third of the plate untreated, as the negative control. Following treatment, parasites were pooled, grown for three days, and then isolated for gDNA extraction, and

amplicon PCR as described for the population genetic studies in chapters 3 and 4. The relative frequencies of barcodes representing the different treatments were quantified. It was not possible to distinguish between treatment groups, with all falling within $\pm 3x$ the SD of the vehicle treatment. Two possible causes of this observation were 1) 4% PFA treatment was not effectively killing parasites or 2) successfully killed parasites were equally represented compared to untreated in the NGS readout. To determine which of these may be responsible I conducted a plaque assay to assess a range of killing methods. I trialled a range of killing conditions in this plaque assay, including ice cold methanol for 20 minutes, 4% PFA, 1 mM iodoacetamide (IAA) a broad-spectrum cysteine reactive electrophile, and heating at 94°C for 10 minutes. DMSO was included as a vehicle control and one sample was untreated as a negative control. Methanol, PFA and heat treatment are all well documented *T. gondii* fixatives, I included IAA as it is known to have broad spectrum toxicity as a cysteine alkylating agent and as a covalent modifier would be a fitting control in screening of covalent fragments. Only IAA and heat-treatment successfully prevented all parasite replication, confirming that 4% PFA treatment was not a suitable positive control for our drug screening platform (Figure 22b). As heat treatment is not convenient in a plate-based format, I opted to use IAA going forward.

As IAA is not a commonly used killing agent for *T. gondii* there is little information in the literature regarding effective concentrations. To assess this, I conducted a further plaque assay, treating extracellular parasites with IAA at a range of concentrations, and propagating on HFF monolayers for seven days undisturbed. From this I established that 1 mM treatment was the lowest concentration at which growth and replication were entirely inhibited (Figure 22c).
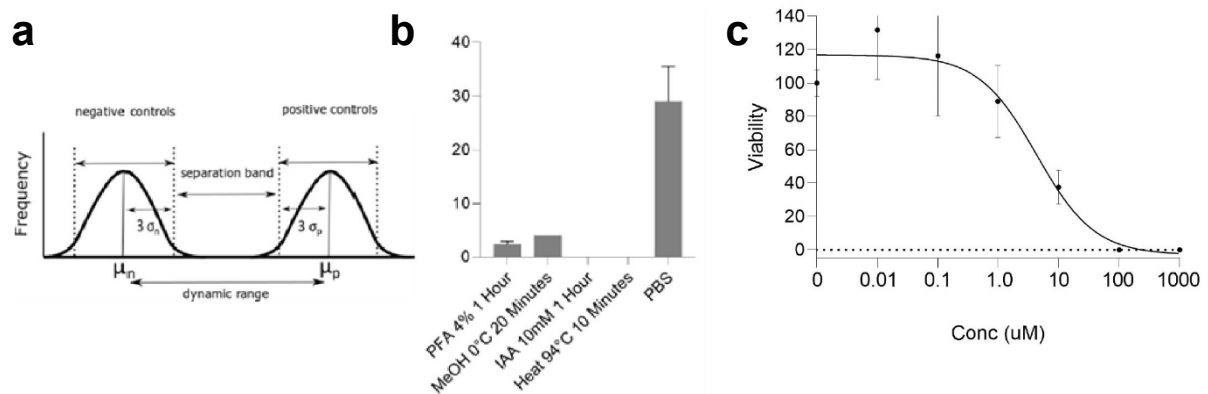
Figure 22 | IAA acts as a suitable positive control for multiplexed drug screening. (a) graph demonstrating z factor derivation. (b) Bar chart illustrating plaque numbers after treatment with a variety of fixative agents to identify a suitable positive control. (c) IC_{50} curve of iodoacetamide treatments to determine effective concentration.
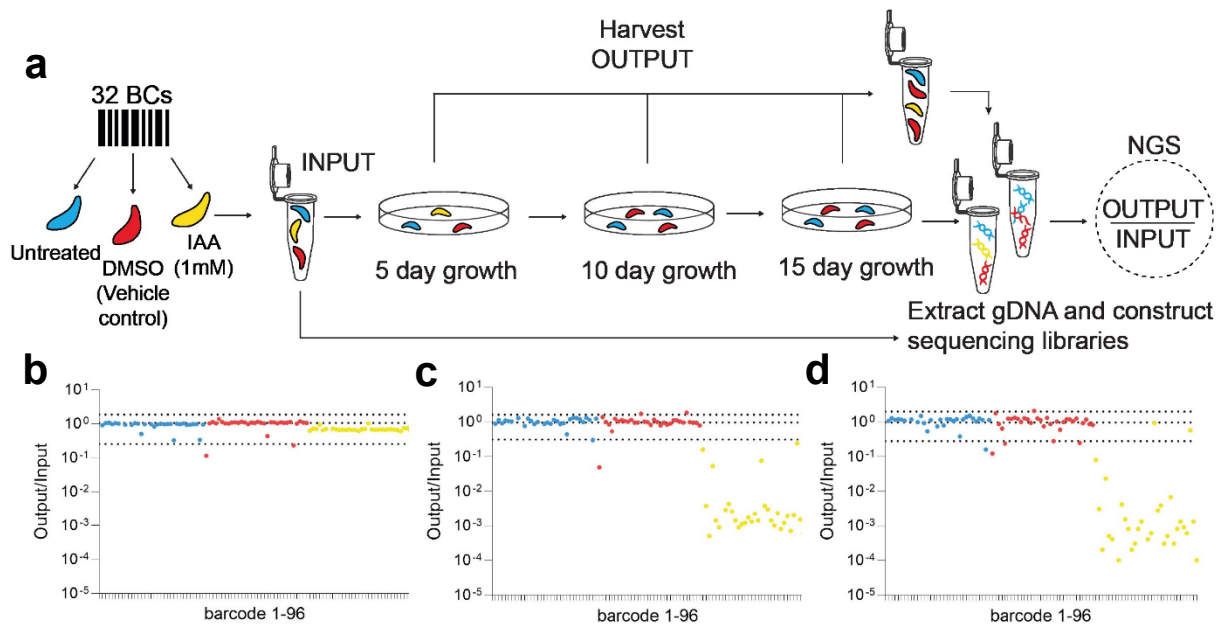
Figure 23 | Live and dead parasites can only be distinguished from one another *in vitro* after serial passage. (a) Schematic depicting the experimental workflow used to establish conditions in which live and dead parasites can be distinguished *in vitro*. Briefly, three pools of barcoded strains, each representing a mini-library of 32 barcoded strains, were generated and each exposed to different treatment conditions. These were then cultured *in vitro* and sampled and passaged at 5-day intervals until final harvest at 15 days post treatment. (b) Scatter plot illustrating percentage representation of barcode populations at five days post infection. Lines displayed on graph illustrate the mean of the untreated and DMSO treated samples and the mean minus ~3x the SD of these controls. (c) Scatter plot illustrating percentage representation of barcode populations at 10 days post infection. Lines displayed on graph illustrate the mean of the untreated and DMSO treated samples and the mean minus ~3x the SD of these controls. (d) Scatter plot illustrating percentage representation of barcode populations at 15 days post infection. Lines displayed on graph illustrate the mean of the untreated and DMSO treated samples and the mean minus ~3x the SD of these controls.

With conditions now established for the reproducible killing of extracellular parasites, I undertook an experiment to test our ability to discriminate between live and IAA-treated parasites using our NGS readout. This was still unsuccessful despite having confirmation from plaque assays that IAA at this concentration does kill parasites. At this stage the experimental set-up involved treatment, followed by propagation *in vitro* for one lytic cycle of around 48-72 hours. I hypothesised that, despite parasites being killed, *in vitro* dead parasites remained in-tact and their gDNA was ultimately still contributing to the final dataset. To interrogate this theory, the experiment was modified, with three samples taken at five day intervals and the remaining parasites passaged onto a new monolayer (Figure 23a). I also included a wash step prior to parasite harvesting to remove extracellular 'zombie' parasites to increase the likelihood of passaging viable parasites. Samples taken at later timepoints are therefore likely to only include parasites which survived treatment. For this experiment I

also lowered the initial infectious burden to facilitate comparisons with planned future *in vivo* studies, where inoculum number would be limited to prevent host fatality. With this reduction
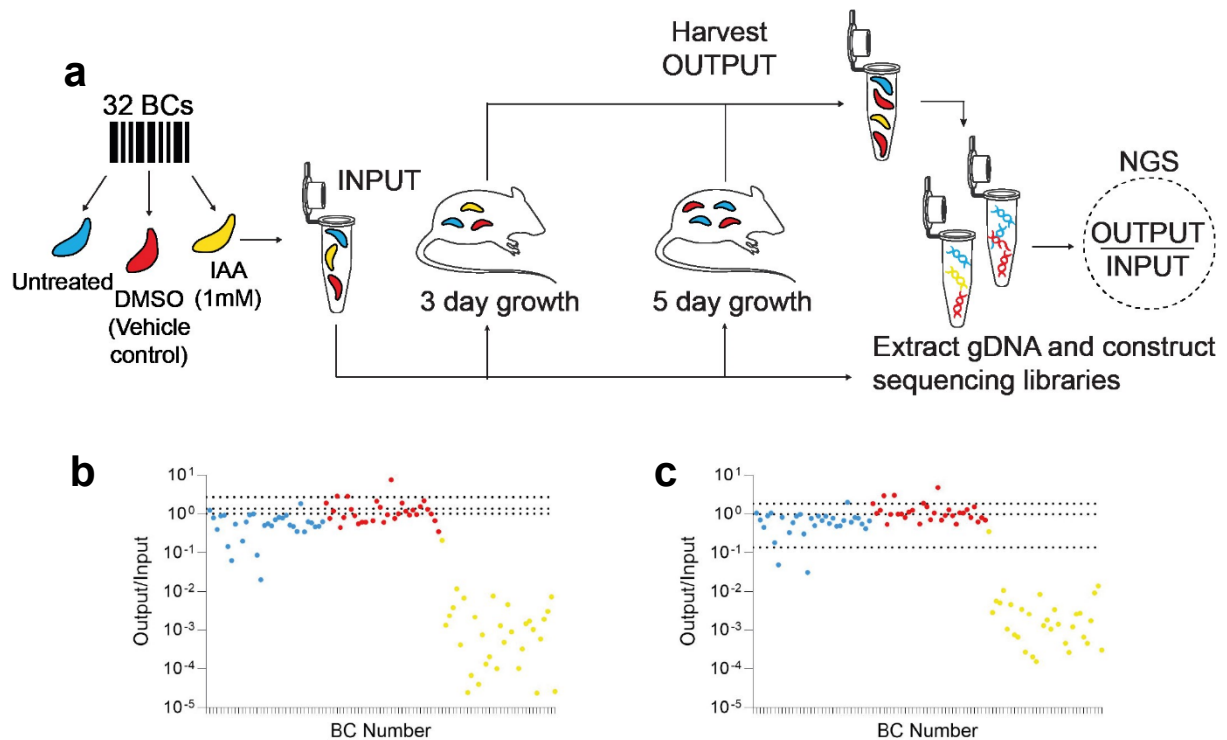
Figure 24 | Live and dead parasites are distinguishable after short proliferation *in vivo*. (a) Schematic depicting the experimental workflow used to establish conditions in which live and dead parasites can be distinguished in vitro. Briefly, three pools of barcoded strains, each representing a mini-library of 32 barcoded strains, were generated and each exposed to different treatment conditions. These were then cultured *in vivo* for five-days with samples taken at three- and five-days post treatment. (b) Scatter plot illustrating percentage representation of barcode populations at three days post infection. Lines displayed on graph illustrate the mean of the untreated and DMSO treated samples, the mean minus ~3x the SD of these controls, and the mean plus ~3x the SD of these controls. (c) Scatter plot illustrating percentage representation of barcode populations at five days post infection. Lines displayed on graph illustrate the mean of the untreated and DMSO treated samples, the mean minus ~3x the SD of these controls, and the mean plus ~3x the SD of these controls.

in parasite number, I increased the interval between passages to provide a sufficient replication window.

Following five days growth, there was little separation between abundance of barcodes associated with the IAA and DMSO treated parasites in the NGS readout (Figure 23b). However, in the samples taken at 10- and 15-days post infection the two groups could be distinguished from one another, with the abundance of barcodes associated with IAA treated parasites was reduced relative to the other treatment groups (Figure 23c&d).

Having established *in vitro* conditions that allowed us to eliminate chemically killed parasites from sample collection, I then sought to establish equivalent conditions for the *in vivo* studies that were planned to run in parallel. I replicated the *in vitro* experimental set-up with subsequent propagation *in vivo* rather than *in vitro*. Extracellular parasites were treated with drug *ex-vivo* for one hour. The parasites were then washed three times with PBS before I resuspended in ice-cold PBS/4% FBS. Intraperitoneal infections were initiated in five animals with two culled at three days post infection and the remaining three at five days post infection

(Figure 24a). As reported in chapter 3 and 4, I have data confirming that there is little host impact on relative frequency distribution of barcodes in the early acute phase of infection. As a result, any changes in barcode representation can be attributed to drug treatment. Positive and negative controls could be detected earlier *in vivo* compared to *in vitro*. Clear differentiation was observed at both three- and five-day post-infection, with the IAA-treated samples showing tighter grouping at the later timepoint (Figure 24 b and c).

I concluded that longer periods of time are required to separate live and dead parasites *in vitro* compared to *in vivo,* which I hypothesise is likely to be a consequence of the effective host clearance system being absent in cell culture. In the *in vitro* scenario, despite parasites being dead, it is possible to harvest their gDNA, hence the requirement of an artificial clearance step (washing and passaging). For *in vivo* experiments, it is also conceivable that the additional *in vitro* expansion step following collection by peritoneal wash improves separation as only live parasites will continue to replicate. It is conceivable with long-acting drugs, one of the hallmarks of those with a covalent MoA, that continued therapeutic action may be observed during this time. I would hypothesise that, though there are growth and infection pathways which are irrelevant *in vitro* but contributory *in vivo*, the reverse of this effect is less prevalent and therefore doesn't confound the experimental results.

5.2.2. Multiplexed Drug Screening

After confirmation that our assay could reproducibly differentiate between live and chemically-killed parasites, I proceeded with multiplexed *in vivo* drug screening as depicted in Figure 25. The mapping of covalent chemical treatments onto the barcoded strain library provides a genetic tracer of the chemical treatment. As barcode identities (and therefore compound identities) can be mapped using the bar-seq approach described in chapter 3, I am able to pool fragment-treated parasites. I can then de-multiplex this information from the NGS data and investigate the phenotypic consequences of individual fragment treatments. Using this
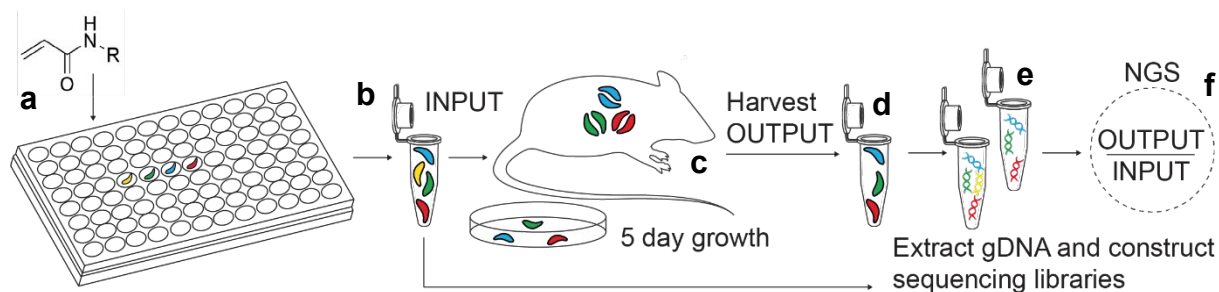
Figure 25 | Using cellular barcoding, *in vivo* drug discovery can be multiplexed. (a) Plate-based transfections are used to generate unique parasite populations which are treated with a range of potential drug compounds *ex vivo*. (b) These parasites are then washed, pooled, and (c) propagated concurrently *in vivo* and *in vitro*. (d) Parasites are harvested by peritoneal wash and their (e) gDNA is extracted and genotyped by NGS. (f) Comparison between input and output samples facilitates assessment of anti-parasitic activity of screened fragments.

approach, I am able to study the impact of many unique fragments on the *T. gondii* lytic cycle at the same time. The chemical fragments used in these experiments operate through a covalent MoA. This results in a considerably prolonged duration of action compared to traditional reversible inhibitors. Consequently, it allows for the extended observation of their effects on growth and infectivity following treatment.

5.2.3. First-round Screen

In the first round of the screen, the library of 84 covalent fragments (2.1.4) was screened along with 12 independent DMSO treatments as negative controls. The experiment was run in parallel both *in vitro* and *in vivo* to facilitate observation of any distinction between the two conditions. Two replicates were conducted *in vitro* and five *in vivo*. The *in vitro* samples were propagated for five days, at which point they were harvested and washed. Of this harvested sample ~20,000 were plated onto fresh HFFs to continue propagating and the remaining parasites were retained for NGS. The final harvest was conducted at ten days post infection (Figure 26a). For the *in vivo* experiment, mice were infected via an IP injection of ~20,000
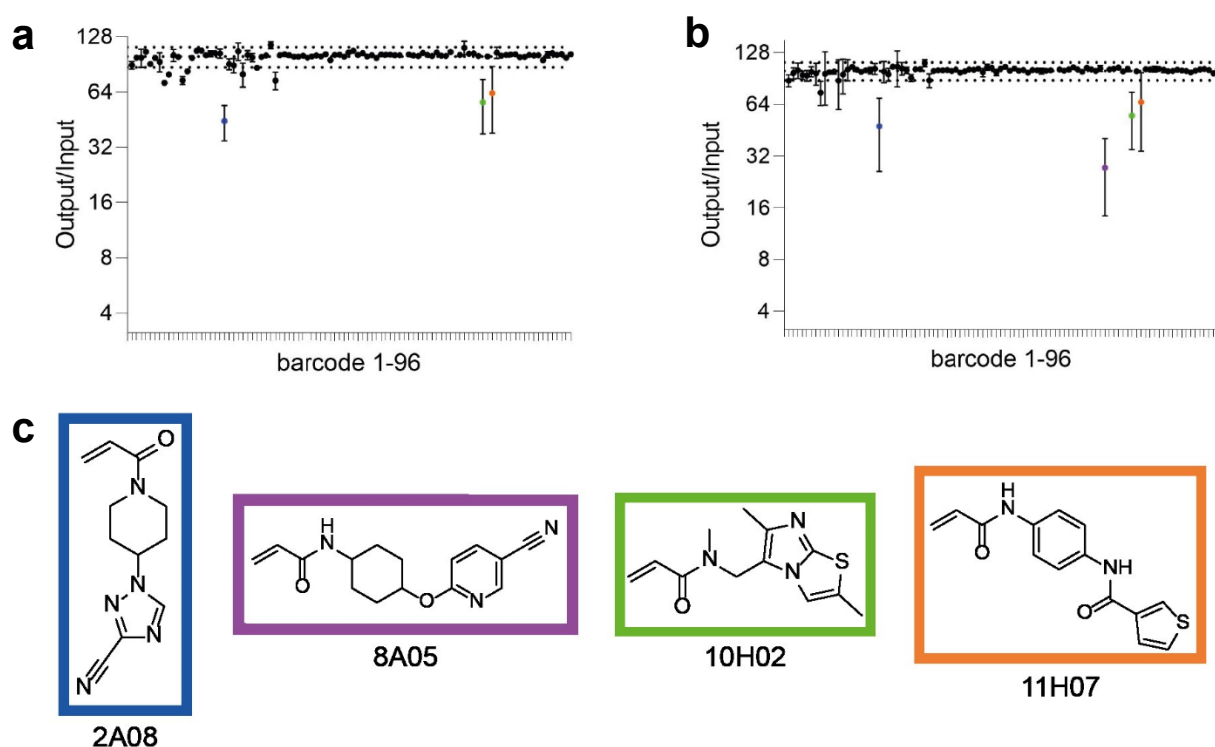
Figure 26 | Multiplexed drug discovery identifies compounds which inhibit parasite growth and replication *in vivo* and *in vitro*. (a) Barcode representation at harvest following 10-days *in vitro* growth. n=2, point plotted is mean value with error bars indicating spread of points. Mean of DMSO control and ±3xSD are plotted. Compounds used to treat barcoded parasite populations which fall below the mean minus 3xSD can be considered to have had an anti-parasitic effect. (b) Barcode representation at harvest following 5-days *in vivo* growth. n=5, point plotted is mean value with error bars indicating spread of points. Mean of DMSO control and ±3xSD are plotted. Compounds used to treat barcoded parasite populations which fall below the mean minus 3xSD can be considered to have had an anti-parasitic effect. (c) Chemical structures of hits identified through screening. Compounds are highlighted to correspond with point colour on graph. Compound highlighted in purple only had an anti-parasitic effect when screened *in vivo*.
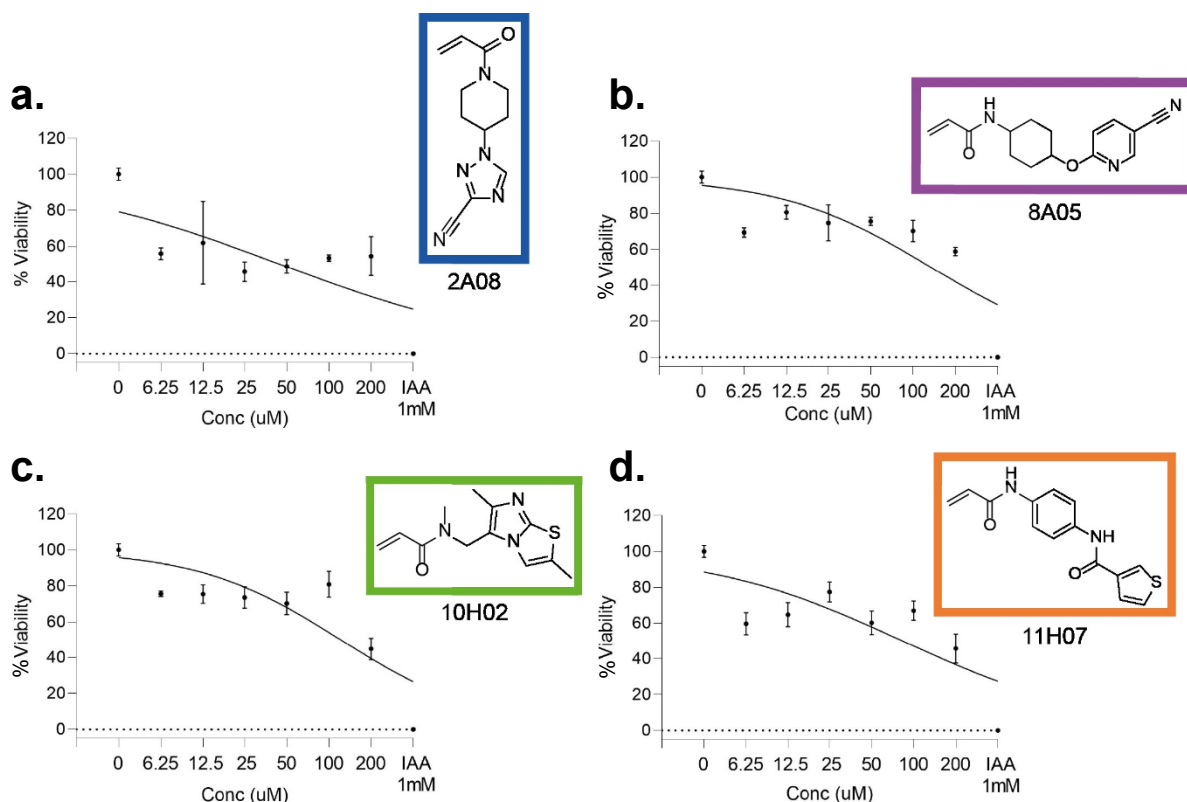
Figure 27 | (a-d) IC$_{50}$ curves depicting the activity of hit compounds identified in the initial drug screen. Plaque assay experiments were performed to determine the IC$_{50}$ value for each hit compound. Extracellular parasites were treated with varying concentrations of the drug and subsequently used to infect confluent HFF monolayers at low concentrations. After a seven-day incubation, crystal violet staining was applied to visualize areas of lytic clearance.

parasites which were propagated for five days before harvesting via peritoneal wash (Figure 26b). Retrieved parasites were subsequently expanded *in vitro* for five days and then harvested. gDNA was isolated and NGS samples prepared as described in chapters 3 and 4 and materials and methods (3.1.3), with data processed in Galaxy prior to analysis. Three hit compounds (2A08, 10H02 and 11H07) were identified in both the *in vitro* and *in vivo* experiments with one hit (8A05) identified solely in the *in vivo* experiment (Figure 26c). This validated the hypothesis that *in vivo* interrogation could reveal novel targets/pathways which are not required for growth in cell culture models.

Plaque assay experiments were conducted to establish an IC$_{50}$ value for each of the hit compounds. Extracellular parasites were treated with a range of drug concentrations and used to inoculate confluent HFF monolayers at low concentrations. These were incubated for seven days before staining with crystal violet to visualise areas of lytic clearance. The observed results did not demonstrate the anticipated gradient of inhibition relative to concentration (Figure 27a-d). Compound 8A05 showed comparable effect *in vitro* to the other hits in this experiment. This could be a result of continuous treatment compared to a finite one-hour treatment in the drug screening protocol. This is comparable to observations made by

Gruner et al.[117] where a similar plateau effect was witnessed. Upon completion of the first round of screening I concluded that the platform could be used to identify novel anti-parasitic lead molecules.

5.2.4. Secondary Screen

Following on from the success of the first round of the screen, I began a second round. In the second iteration, I bioinformatically interrogated the ~10,000-member Enamine covalent fragment library to identify compounds which exhibited structural and/or chemical similarity to first-round hits. Similarity of enamine compounds to first-round hit molecules was assessed using the FragFp diversity algorithm in DataWarrior v5.5.0[122] to identify compounds that shared molecular features with our lead hits from the first round of the screen. Those with a similarity score of >0.8 were taken forward for secondary screening. For the hit which had been shown to only inhibit parasite growth *in vivo,* no similar compounds were identified within the 10,000-member library of covalent fragments (Enamine). Compounds with both acrylamide and nitrile moieties were identified as having similarity to other hits. Both are cysteine modifiers; however, nitriles have a milder electrophilic character reducing the
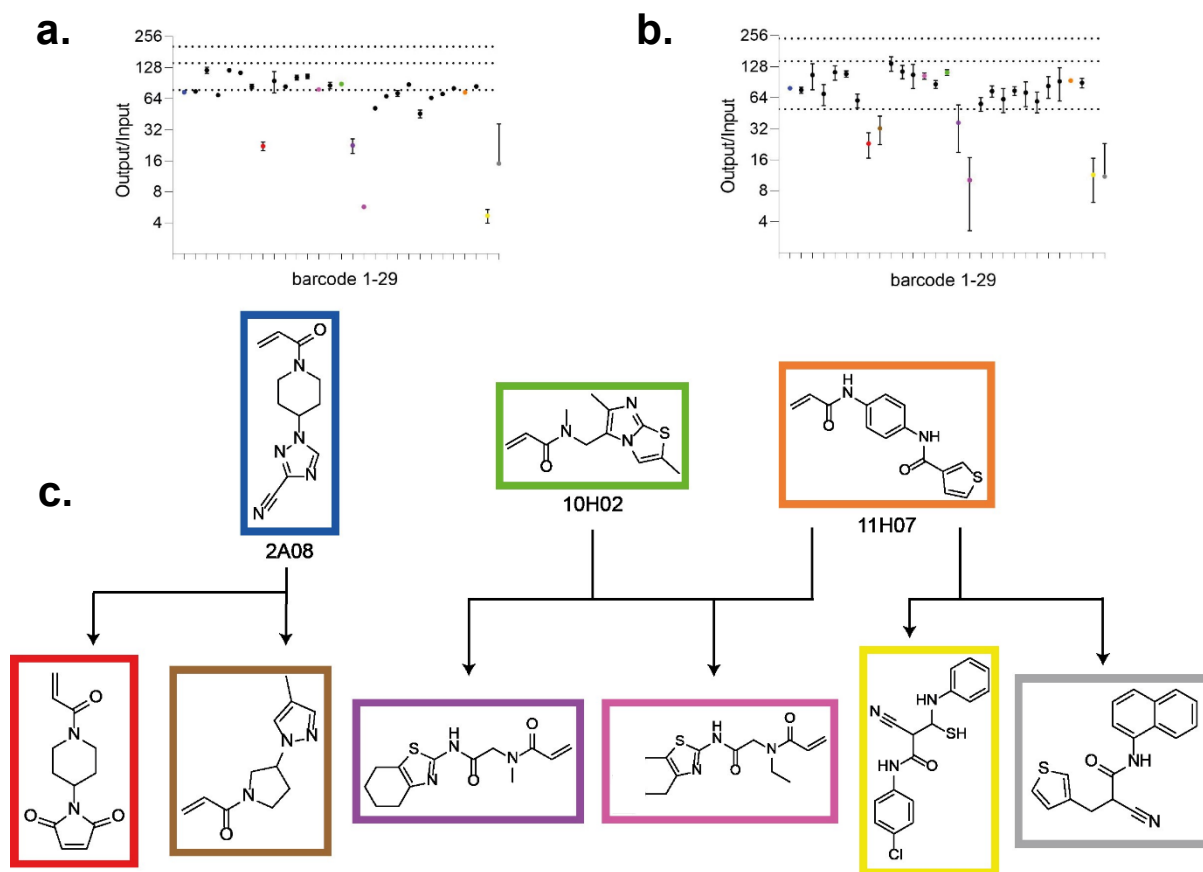
Figure 28 | A second iteration of multiplexed drug screening with next generation compounds identified fragments with enhanced efficacy. (a) Scatterplot illustrating barcode representation following secondary screen *in vitro*. n=2, point plotted is mean value with error bars indicating spread of points Also plotted is the average of the DMSO treated barcodes ±3xSD and the average of the IAA treated barcodes. (b) Scatterplot illustrating barcode representation following secondary screen *in vivo*. n=5, point plotted is mean value with error bars indicating spread of points. Also plotted is the average of the DMSO treated barcodes ±3xSD and the average of the IAA treated barcodes. (c) Second-generation hit molecules and their lineage relative to first-generation compounds. Some second-generation molecules were identified as similar to two first-generation compounds indicating multiple potential targets or and overlap in targets of different hits.

likelihood of off-target effects. Due to this reduced electrophilicity in comparison to acrylamides, nitriles typically rely on highly reactive nucleophiles to bind however, the reactivity of nitriles may be increased by attachment to electron withdrawing moieties[184,185]. The screening protocol was as previously described with the addition of 1 mM IAA serving as a high-potency positive control. For each of the preliminary hits there were second generation molecules which exhibited enhanced efficacy, demonstrating the tractability of this iterative approach (Figure 28 a and b). Compounds with both acrylamide and nitrile warheads were identified as having increased potency over first-round hits (Figure 28 c).

5.2.5. Inactive Fragment Screen

To test whether enhanced efficacy observed in the secondary screen was a result of elaboration of primary hits I conducted a similar process with compounds from the original

screen that had exhibited no growth inhibitory effects. I searched our Enamine library for compounds with similarity to four randomly selected members of the original screening library. As before, up to eight similar compounds were then selected using DataWarrior's FragFp diversity algorithm and these were screened in the same format as previously described. On examination of NGS data it became clear that there were very few matched reads providing limited coverage of barcode frequency impeding analysis. I believe this is a result of poor parasite growth in cell culture prior to the experiment. I had been experiencing some unusual morphology in our host cell line and as *T. gondii* is an obligate intracellular pathogen it is dependent on the host for nutrition and growth. I hypothesise that this had a significant impact on transfection efficiency and therefore confounded any meaningful analysis of the results.

5.2.6. Characterisation of 11H07

To assess host-cell toxicity of identified first-round hits, I conducted an MTT viability assay. This viability assay measures metabolic activity, using NAD(P)H-dependent cellular oxidoreductase enzymes. The presence of these enzymes under defined conditions is proportional to the number of viable cells. The enzymes reduce the tetrazolium dye MTT producing an insoluble formazan salt which is purple in colour. By measuring the absorbance of solution incubated with treated cells relative to an untreated control confirmation on host cell toxicity can be obtained (Figure 29a)[186]. Using IAA as a positive control for cell death along with a 'no HFF' negative control I confirmed that the identified hits did not exhibit host cell toxicity (Figure 29b).
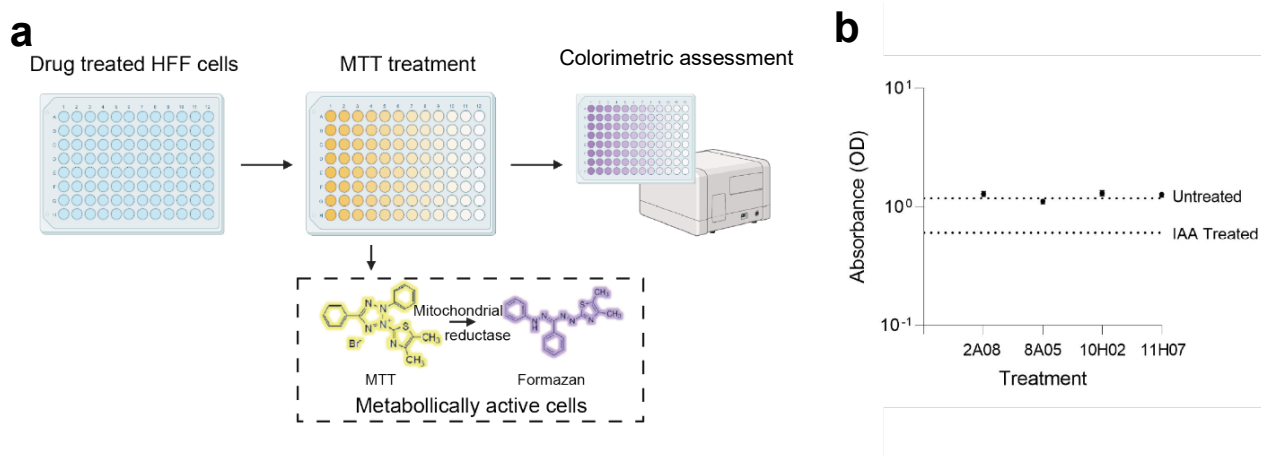
Figure 29 | Primary screen hits are not metabolically toxic to host cells. (a) Schematic depicting workflow for MTT cytotoxicity assay. Briefly, HFFs are incubated with the control and experimental treatment and subsequently DMEM and 50 µg MTT (Roche). Cells were treated with 10% SDS (Roche) overnight. A colorimetric assessment then quantifies the relative presence of formazan determining metabolic activity. (b) Scatter plot demonstrating absorbance of samples subjected to varying drug treatments, n = 4. Lines represent average absorbance untreated cells and those treated with 1mM IAA.
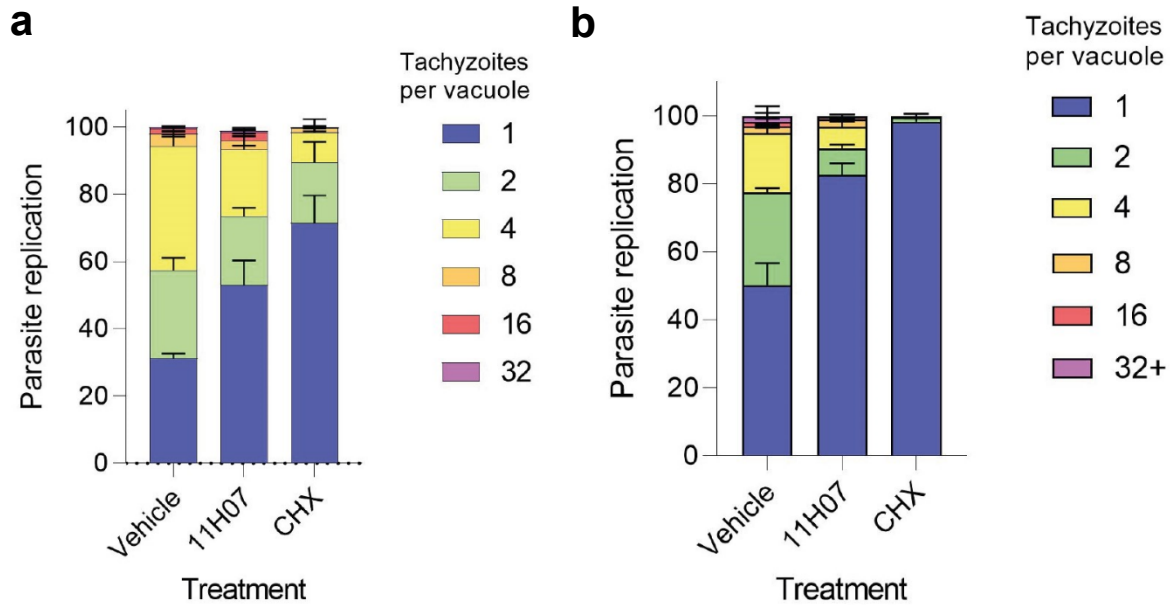
Figure 30 | 11H07 is more inhibitive to parasite replication when administered as a continuous treatment rather than a single dose. Parasites were treated with either 11H07, Cycloheximide (CHX) or 1% DMSO, (a) continuously throughout this experiment or (b) for two hours. Invasion of confluent HFFs seeded in a 24 well plate was synchronised by infecting chilled host cells and allowed to proceed for 4 hrs. Cells were then washed to remove any uninvaded parasites. Parasites were allowed to replicate for 18hrs and stained with anti-SAG1. Parasites per vacuole were counted for each treatment and normalised against the DMSO control. Bars represent mean of 4 technical replicate ± SD. *Work in this figure was conducted by Rebecca Oxtoby, a Master's student, under my direct supervision

In concurrent work in the Child Lab[187], 11H07 was identified as a potential translation inhibitor in *T. gondii*. As a result, further studies were conducted to characterise its effects on parasite growth and replication. As an obligate intracellular parasite *T. gondii* can only replicate within the host cell environment. In an experiment conceived to determine the effect of 11H07 on parasite replication, parasites were treated with either DMSO (vehicle control), cycloheximide (CHX) (positive control) or 11H07. CHX is a translation inhibitor so if our hypothesis regarding the mechanism of action of 11H07 is correct they should exhibit similar phenotypes. Small molecules were applied as continuous treatments (Figure 30a) and as a single dose (Figure 30b). Treated parasites were incubated for 18 hours before staining with fluorescent anti-SAG-1 antibodies for visualisation. The majority of those treated with 11H07 and CHX only had one parasite/PV compared to between two and four parasite/PV in those treated with DMSO. Continuous treatment with 11H07 results in a more robust effect, this may indicate the target is highly expressed and can be rapidly replenished. As 11H07 is a covalent modifier it binds irreversibly so compound from a one-off dose will be in complex with the target and cannot engage newly synthesised protein.

5.3. Discussion

I have successfully developed a multiplexed *in vivo* drug screening platform using cellular barcoding. In the current format this technique facilitates screening of up to 96 potential drug fragments in a single experiment. However, with further development it may be possible to increase this number using techniques such as automated liquid handling or development of a larger barcode library. Whilst all screening experiments in this work were conducted in *T. gondii*, given the successes had with regards to barcoding *T. brucei* this technique should be readily transferable into this parasite system.

When establishing a suitable positive control for use in drug-screening experiments I encountered unanticipated findings. I did not predict that dead parasites would meaningfully contribute to NGS data without removal by mechanical clearance following an *in vitro* infection. While HFFs are primary human cells and so provide conditions somewhat comparable to clinical infections[188], *T. gondii* does not routinely infect skin-cells and so they may lack optimised clearance mechanisms. Consequently, dead parasites persist in cell culture as "zombies," with their DNA remaining intact enough for extraction and sequencing. To overcome this issue, I successfully introduced a mechanical clearance step and extended the duration of parasite culture. It is worth noting that I did not find documentation of this challenge in the literature.

In vivo screening is a necessary step in the drug-discovery process[189]. Current strategies facilitate testing of one condition per animal, making the process resource and time intensive. There are also reproducibility implications which are compounded further in complex infection models. The technique reported in this body of work effectively reduces the number of animals used in early-stage discovery, while still generating an amount of data and replicates that are deemed meaningful. To quantify the reduction in animal usage facilitated by this technique, using traditional strategies assessing a library of 84 compounds would require 84 animals. To correct for inter-individual variation this would be repeated five times, giving a total of 420 animals. I generated equivalent information in five animals, a 98.8% reduction. An interesting point for consideration is whether data points obtained from different animals are biological or technical replicates. This somewhat hard to distinguish as they originate from the same inoculum, and therefore have the same biological origins, but were propagated in independently grown cells/individual mice, and so have been subjected to unique biological pressures and environments.

With regards to the results when seeking to determine the IC_{50} of first-round hits, further interrogation of the literature presented a possible explanation. Covalent binding, as described

$$E + I \rightleftharpoons E{\bullet}I \rightarrow E{\bullet}I*$$

$$K_i \qquad K_{inact}$$
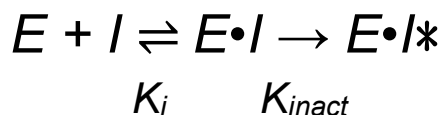
Figure 31 | Calculation describing the kinetics of covalent binding where E = enzyme and I = inhibitor. The equation describes a two-stage process where the inhibitor reversibly binds the enzyme generating an inhibitor-enzyme complex. In the second stage the inhibitor renders the enzyme inactive.

by the equation in Figure 31, is a two-step process. The first stage of binding is reversible association with the target, impacted by both efficacy of drug molecule and the target's association constant. The second stage is irreversible biding which inactivates the target[190]. A consequence of this two-step process, and the non-continuous nature of treatment, both concentration and incubation time act as limiting factors. To establish the true IC_{50} of each fragment, parasites could be exposed to continuous treatment or experiments to determine optimal treatment time could be conducted.

I noted that compound 8A05 displayed some level of inhibition on parasite growth and replication during the *in vitro* IC_{50} experiment. This observation raises a concern regarding the potential for false negatives within the screening protocol. The disparity observed can likely be attributed to differences in treatment duration between the initial screening and the IC_{50} experiments.

During the preliminary screening, parasites were exposed to the treatment for only one hour before undergoing a wash step. Consequently, for a fragment to demonstrate inhibitory effects, it would need to possess an extended duration of action. Conversely, in the IC_{50} experiment, parasites were co-incubated with the treatment for a continuous seven-day period. It is plausible that prolonged exposure to 8A05 does indeed result in an inhibition of growth and replication *in vitro*.

This insight suggests the possibility that there could be other compounds exhibiting a similar pattern, requiring sustained treatment for their inhibitory properties to manifest either *in vitro*, *in vivo*, or in both contexts. While acknowledging the susceptibility of this screening platform to false negatives, the substantial number of compounds screened and the favourable hit rate reassure us that the protocol efficiently identifies potential lead molecules. This efficiency is instrumental in reducing both the time and resources required, as well as minimising the need for animal testing in the drug discovery process.

Our proof-of-concept data supports the hypothesis that by iteratively developing screened compounds at the molecular level, it is possible to generate increasingly effective hits. By maximising diversity in our initial screening library and using FragFp diversity algorithm in DataWarrior v5.5.0 to assess similarity and therefore eligibility for second-round screening,

limiting human bias, I successfully identified hits with increased efficacy. By establishing accurate IC$_{50}$'s of hit molecules, I would be able to also confirm whether iterative hits had enhanced efficacy at lower concentrations, improving their eligibility for further drug discovery studies.

By effectively conducting a screen using presumed inactive fragments, I can evaluate whether the improved efficacy of second-generation hits stems from enhancements made to first-generation hits rather than random selection of new effective molecules from the fragment library. Although using extensive libraries of fragments and software-based similarity assessments has been valuable for modelling unbiased iterative lead compound development, it is important to acknowledge that both approaches are to some extent influenced by human selection biases. The construction of commercial libraries is informed by previously successful drug molecules. To move towards and entirely unbiased medicinal chemistry strategy it is necessary to remove any reliance on past data. I propose a novel method of evolution-directed synthesis (EDS). A library of compound fragments would be synthesised with the sole aim of sampling a diverse range of chemical space. These fragments would be screened in the workflow described in this work. Any identified hit molecule would then be taken forward for development. To do this in an unbiased way, hits would be exposed to a range of chemical conditions, similar to studies by Townley et al.[64]. In this study researchers started with complex three-dimensional intermediates that could be converted into various molecular scaffolds through scaffold-hopping techniques. The study successfully yielded 21 distinct scaffolds, which may be further developed to generate lead-like screening compounds with potential for bioactive molecular discovery. This process of fragment elaboration could continue iteratively until a viable drug molecule is produced. I believe that this combined with the cellular barcoding methodology described in this work would facilitate discovery of novel anti-parasitic agents unbiased by traditional drug discovery conventions.

6. Conclusions & Future Directions

This work presents the development of a novel cellular barcoding tool for eukaryotic pathogens, specifically applied to *T. gondii* and *T. brucei*. I propose that this technique can be adapted for other systems with suitable loci for genetic manipulation. Through proof-of-concept studies, I confirmed that barcoded parasites did not exhibit impaired growth phenotypes. Additionally, I demonstrated that no barcodes resulted in preferential growth of parasite strains, and all barcodes could be stably maintained *in vitro* and *in vivo*. These combined results validate the use of these cellular barcodes as neutral tags for genetically tracking parasite populations, similar to previously employed strategies in bacteria and viruses[110,114,191].

Cellular barcoding was employed in *T. gondii* to examine the population dynamics of *in vivo* infection. The findings revealed that *T. gondii* readily colonises the brain parenchyma, and contrary to expectations, the BBB does not pose a restrictive bottleneck. Additionally, I established the dynamic nature of brain colonisation, as evidenced by a gradual decline in cyst numbers over several months. This decline correlated with ongoing genetic divergence from the initial parasite population, indicating that *T. gondii* chronic infection is more dynamic than previously presumed. Interestingly, the pattern of genetic divergence varied among hosts, with replicates from inbred mouse strains clustering closely together. In contrast, replicates from outbred mouse strains showed a broader range of genetic divergence compared to the initial parasite population. These findings highlight the importance of the host environment in the study of parasitic infections, providing further evidence for the necessity of *in vivo* investigations. In future studies, the cellular barcoding methodology employed in this research could be applied to explore the population dynamics of *T. brucei* within the host. Similar to its application in *T. gondii*, it could be used to identify preferred infectious niches and determine whether there are any significant genetic bottlenecks during their colonisation.

I have demonstrated that cellular barcoding enables multiplexed, *in vivo* drug screening, thereby reducing the number of animals used in early-stage research while generating meaningful data. To build upon these findings, target identification studies, particularly with 11H07 and 8A05, could provide further insights into potential pathways for inhibiting *T. gondii's in vivo* proliferation. It may also be of interest to modify 11H07 to function as a chemical probe. If 11H07 does in fact inhibit translation this would be a first-in-class tool to interrogate this process in *T. gondii*.

To expand on the hit molecules identified in this study, two approaches can be considered. Firstly, the interrogation of molecular databases to identify similar compounds for screening, aiming to identify more efficacious hit molecules. Secondly, the implementation of the principle

of EDS outlined in Section 5.5. This approach would reduce dependence on commercially available compounds and minimise bias toward 'drug-like' features. However, implementing this method would require expertise in organic synthesis and the necessary infrastructure, yet it holds the potential to yield results that are otherwise unattainable through alternative strategies.

In conclusion, this study showcases the development and application of a novel cellular barcoding methodology for eukaryotic pathogens. This innovative approach enables *in vivo* evaluation of parasitic infections and potential treatments. Importantly, this work contributes to the 3R's principles by reducing the number of animals required in early-stage research.

7. References:

1. Wincott, C. J. *et al.* Cellular barcoding of protozoan pathogens reveals the within-host population dynamics of Toxoplasma gondii host colonization. *Cell Reports Methods* **2**, (2022).

2. Kim, K. & Weiss, L. M. Toxoplasma gondii: The model apicomplexan. *Int J Parasitol* **34**, 423–432 (2004).

3. Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria.

4. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. Malaria: Biology and Disease. *Cell* **167**, 610–624 (2016).

5. Gerace, E., Presti, V. D. M. Lo & Biondo, C. Cryptosporidium Infection: Epidemiology, Pathogenesis, and Differential Diagnosis. *Eur J Microbiol Immunol (Bp)* **9**, 119 (2019).

6. Thomson, S. *et al.* Bovine cryptosporidiosis: impact, host-parasite interaction and control strategies. *Veterinary Research 2017 48:1* **48**, 1–16 (2017).

7. Gondim, L. F. P. Neospora caninum in wildlife. *Trends Parasitol* **22**, 247–252 (2006).

8. Fatoba, A. J. & Adeleke, M. A. Transgenic Eimeria parasite: A potential control strategy for chicken coccidiosis. *Acta Trop* **205**, (2020).

9. Dubey, J. P. The history of Toxoplasma gondii - The first 100 years. *Journal of Eukaryotic Microbiology* **55**, 467–475 (2008).

10. Sabin, A. B. & Olitsky, P. K. Toxoplasma and obligate intracellular parasitism. *Science (1979)* **85**, 336–338 (1937).

11. Wolf, A., Cowen, D. & Paige, B. Human toxoplasmosis: Occurrence in infants as an encephalomyelitis verification by transmission to animals. *Science (1979)* **89**, 226–227 (1939).

12. Sabin, A. B. TOXOPLASMIC ENCEPHALITIS IN CHILDREN. *J Am Med Assoc* **116**, 801–807 (1941).

13. Innes, E. A. A brief history and overview of Toxoplasma gondii. *Zoonoses Public Health* **57**, 1–7 (2010).

14. Miller, N. L., Frenkel, J. K. & Dubey, J. P. *Oral Infections with Toxoplasma Cysts and Oocysts in Felines, Other Mammals, and in Birds*. *Source: The Journal of Parasitology* vol. 58 https://www.jstor.org/stable/3286588 (1972).

15. Di Genova, B. M., Wilson, S. K., Dubey, J. P. & Knoll, L. J. Intestinal delta-6-desaturase activity determines host range for Toxoplasma sexual reproduction. *PLoS Biol* **17**, e3000364 (2019).

16. *Toxoplasma gondii The Model Apicomplexan: Perspectives and Methods*. (2014).

17. Frenkel, J. K., Ruiz, A. & Chinchilla, M. Soil survival of Toxoplasma oocysts in Kansas and Costa Rica. *American Journal of Tropical Medicine and Hygiene* **24**, 439–443 (1975).

18. Dubey, J. P., Miller, N. L. & Frenkel, J. K. The Toxoplasma gondii oocyst from cat feces. *J Exp Med* **132**, 636–662 (1970).

19. Dubey, J. P., Lindsay, D. S. & Speer, C. A. Structures of Toxoplasma gondii Tachyzoites, Bradyzoites, and Sporozoites and Biology and Development of Tissue Cysts. **11**, 267–299 (1998).

20. Kim, K. S. Mechanisms of microbial traversal of the blood–brain barrier. *Nat Rev Microbiol* **6**, 625–634 (2008).

21. Meissner, M., Schlüter, D. & Soldati, D. Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion. *Science* **298**, 837–840 (2002).

22. Opitz, C. & Soldati, D. 'The glideosome': a dynamic complex powering gliding motion and host cell invasion by Toxoplasma gondii. *Mol Microbiol* **45**, 597–604 (2002).

23. Barragan, A. & David Sibley, L. Transepithelial Migration of Toxoplasma gondii Is Linked to Parasite Motility and Virulence. *Journal of Experimental Medicine* **195**, 1625–1633 (2002).

24. Konradt, C. *et al.* Endothelial cells are a replicative niche for entry of Toxoplasma gondii to the central nervous system. *Nat Microbiol* **1**, 1–8 (2016).

25. Lambert, H., Hitziger, N., Dellacasa, I., Svensson, M. & Barragan, A. Induction of dendritic cell migration upon Toxoplasma gondii infection potentiates parasite dissemination. *Cell Microbiol* **8**, 1611–1623 (2006).

26. Black J C., M. W. B. Lytic Cycle of Toxoplasma gondii. *Microbiology and Molecular Biology Reviews* **64**, 607–623 (2000).

27. Blader, I. J., Coleman, B. I., Chen, C. T. & Gubbels, M. J. Lytic Cycle of Toxoplasma gondii: 15 Years Later. *Annu Rev Microbiol* **69**, 463–485 (2015).

28. Meissner, M., Schlüter, D. & Soldati, D. Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion. *Science* **298**, 837–840 (2002).

29. Pavlou, G. *et al.* Toxoplasma Parasite Twisting Motion Mechanically Induces Host Cell Membrane Fission to Complete Invasion within a Protective Vacuole. *Cell Host Microbe* **24**, 81-96.e5 (2018).

30. Clough, B. & Frickel, E. M. The Toxoplasma Parasitophorous Vacuole: An Evolving Host-Parasite Frontier. *Trends Parasitol* **33**, 473–488 (2017).

31. Jerome, M. E., Radke, J. R., Bohne, W., Roos, D. S. & White, M. W. Toxoplasma gondii bradyzoites form spontaneously during sporozoite-initiated development. *Infect Immun* **66**, 4838–4844 (1998).

32. Alday, P. H. & Doggett, J. S. Drugs in development for toxoplasmosis: advances, challenges, and current status. *Drug Des Devel Ther* **11**, 273–293 (2017).

33. Montoya, J. G. & Liesenfeld, O. Toxoplasmosis. *The Lancet* **363**, 1965–1976 (2004).

34. Remington, J. S. & Israelski, D. M. Toxoplasmosis in patients with cancer. *Clin Infect Dis* **17 Suppl 2**, S423–S435 (1993).

35. Wang, Z. D. *et al.* Prevalence and burden of Toxoplasma gondii infection in HIV-infected people: a systematic review and meta-analysis. *Lancet HIV* **4**, e177–e188 (2017).

36. Pappas, G., Roussos, N. & Falagas, M. E. Toxoplasmosis snapshots: global status of Toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis. *Int J Parasitol* **39**, 1385–1394 (2009).

37.   Grover, C. M., Thulliez, P., Remington, J. S. & Boothroyd, J. C. Rapid prenatal diagnosis of congenital Toxoplasma infection by using polymerase chain reaction and amniotic fluid. *J Clin Microbiol* **28**, 2297–2301 (1990).

38.   Stelzer, S. *et al.* Toxoplasma gondii infection and toxoplasmosis in farm animals: Risk factors and economic impact. *Food Waterborne Parasitol* **15**, (2019).

39.   Dubey, J. P. Toxoplasmosis in sheep—The last 20 years. *Vet Parasitol* **163**, 1–14 (2009).

40.   Dubey, J. P. Toxoplasmosis in pigs—The last 20 years. *Vet Parasitol* **164**, 89–103 (2009).

41.   EYLES, D. E. & COLEMAN, N. Tests of 2,4-diaminopyrimidines on toxoplasmosis. *Public Health Reports* **67**, 249 (1952).

42.   Piketty, C., Derouin, F., Rouveix, B. & Pocidalo, J. J. In vivo assessment of antimicrobial agents against Toxoplasma gondii by quantification of parasites in the blood, lungs, and brain of infected mice. *Antimicrob Agents Chemother* **34**, 1467–1472 (1990).

43.   Dunay, I. R., Gajurel, K., Dhakal, R., Liesenfeld, O. & Montoya, J. G. Treatment of toxoplasmosis: Historical perspective, animal models, and current clinical practice. *Clin Microbiol Rev* **31**, (2018).

44.   Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov* **10**, 428–438 (2011).

45.   Scannell, J. W. & Bosley, J. When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. *PLoS One* **11**, e0147215 (2016).

46.   Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* **9**, 203–214 (2010).

47.   Lindsay, M. A. Target Discovery. *Nat Rev Drug Discov* **2**, 831–838 (2003).

48.   Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5**, 821–834 (2006).

49.   Sams-Dodd, F. Target-based drug discovery: is something wrong? *Drug Discov Today* **10**, 139–147 (2005).

50.   Arrowsmith, J. & Miller, P. Phase II and Phase III attrition rates 2011–2012. *Nature Reviews Drug Discovery 2013 12:8* **12**, 569 (2013).

51.   Vincent, F. *et al.* Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nature Reviews Drug Discovery 2022 21:12* **21**, 899–914 (2022).

52.   Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat Rev Drug Discov* **10**, 507–519 (2011).

53.   Munos, B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov* **8**, 959–968 (2009).

54.   Vincent, F. *et al.* Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nature Reviews Drug Discovery 2022 21:12* **21**, 899–914 (2022).

55.   Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* **1**, 337–341 (2004).

56.   Jia, X. *et al.* Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. *Nature* **573**, 251–255 (2019).

57.   Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688-702 e13 (2020).

58.   Galloway, W. R. & Spring, D. R. Is synthesis the main hurdle for the generation of diversity in compound libraries for screening? . *Expert Opinion in Drug Discovery* **4**, 467–472 (2009).

59.   Galloway, W. R., Isidro-Llobet, A. & Spring, D. R. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nat Commun* **1**, 80 (2010).

60.   Burke, M. D., Berger, E. M. & Schreiber, S. L. Generating diverse skeletons of small molecules combinatorially . *Science (1979)* **302**, 613–618 (2003).

61.   Welsch, M. E., Snyder, S. A. & Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr Opin Chem Biol* **14**, 347–361 (2010).

62.   Wetzel, S., Bon, R. S., Kumar, K. & Waldmann, H. Biology-oriented synthesis. *Angew Chem Int Ed Engl* **50**, 10800–10826 (2011).

63.   van Hattum, H. & Waldmann, H. Biology-oriented synthesis: harnessing the power of evolution. *J Am Chem Soc* **136**, 11853–11859 (2014).

64.   Townley, C., McMurray, L., Marsden, S. P. & Nelson, A. A unified "top-down" approach for the synthesis of diverse lead-like molecular scaffolds. *Bioorg Med Chem Lett* **62**, 128631 (2022).

65.   Weber, L. Multi-component reactions and evolutionary chemistry. *Drug Discov Today* **7**, 143–147 (2002).

66.   Weber, L. Applications of genetic algorithms in molecular diversity. *Curr Opin Chem Biol* **2**, 381–385 (1998).

67.   Weber, L., Wallbaum, Dr. S., Broger, Dr. C. & Gubernator, Dr. K. Optimization of the Biological Activity of Combinatorial Compound Libraries by a Genetic Algorithm. *Angew Chem Int Ed Engl* **34**, 2280–2282 (1995).

68.   Ghosh, A. K., Samanta, I., Mondal, A. & Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **14**, 889 (2019).

69.   Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nature Reviews Drug Discovery 2011 10:4* **10**, 307–317 (2011).

70.   Chandrasekharan, N. V. & Simmons, D. L. The cyclooxygenases. *Genome Biol* **5**, 1–7 (2004).

71.   Vane, J. R. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nat New Biol* **231**, 232–235 (1971).

72.   Nicola, G., Tomberg, J., Pratt, R. F., Nicholas, R. A. & Davies, C. Crystal structures of covalent complexes of β-lactam antibiotics with E. coli penicillin-binding protein 5: toward an understanding of antibiotic specificity. *Biochemistry* **49**, 8094 (2010).

73.   Kahan, F. M., Kahan, J. S., Cassidy, P. J. & Kropp, H. The mechanism of action of fosfomycin (phosphonomycin). *Ann N Y Acad Sci* **235**, 364–386 (1974).

74.   Danenberg, P. V., Langenbach, R. J. & Heidelberger, C. Structures of reversible and irreversible complexes of thymidylate synthetase and fluorinated pyrimidine nucleotides. *Biochemistry* **13**, 926–933 (1974).

75.   Xu, H., Faber, C., Uchiki, T., Racca, J. & Dealwis, C. Structures of eukaryotic ribonucleotide reductase I define gemcitabine diphosphate binding and subunit assembly. *Proc Natl Acad Sci U S A* **103**, 4028–4033 (2006).

76.   Ou, S. H. I. Second-generation irreversible epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs): a better mousetrap? A review of the clinical evidence. *Crit Rev Oncol Hematol* **83**, 407–421 (2012).

77.   Paez, J. G. *et al.* EGFR mutations in lung, cancer: Correlation with clinical response to gefitinib therapy. *Science (1979)* **304**, 1497–1500 (2004).

78.   Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129–2139 (2004).

79.   Harvey, R. D., Adams, V. R., Beardslee, T. & Medina, P. Afatinib for the treatment of EGFR mutation-positive NSCLC: A review of clinical findings. *J Oncol Pharm Pract* **26**, 1461–1474 (2020).

80.   Finlay, M. R. V. *et al.* Discovery of a potent and selective EGFR inhibitor (AZD9291) of both sensitizing and T790M resistance mutations that spares the wild type form of the receptor. *J Med Chem* **57**, 8249–8267 (2014).

81.   Walter, A. O. *et al.* Discovery of a mutant-selective covalent inhibitor of EGFR that overcomes T790M-mediated resistance in NSCLC. *Cancer Discov* **3**, 1404–1415 (2013).

82.   Leonetti, A. *et al.* Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *British Journal of Cancer 2019 121:9* **121**, 725–737 (2019).

83.   Pal Singh, S., Dammeijer, F. & Hendriks, R. W. Role of Bruton's tyrosine kinase in B cells and malignancies. *Molecular Cancer 2018 17:1* **17**, 1–23 (2018).

84.   Arneson, L. C., Carroll, K. J. & Ruderman, E. M. Bruton's Tyrosine Kinase Inhibition for the Treatment of Rheumatoid Arthritis. *Immunotargets Ther* **10**, 333 (2021).

85.   Pan, Z. *et al.* Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem* **2**, 58–61 (2007).

86.   Honigberg, L. A. *et al.* The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc Natl Acad Sci U S A* **107**, 13075–13080 (2010).

87.   Herman, S. E. M. *et al.* Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood* **117**, 6287–6296 (2011).

88.   Burger, J. A. & Buggy, J. J. Bruton tyrosine kinase inhibitor ibrutinib (PCI-32765). *Leuk Lymphoma* **54**, 2385–2391 (2013).

89.   Twenty-One Billion Dollars. Really. | Science | AAAS. https://www.science.org/content/blog-post/twenty-one-billion-dollars-really.

90.   Munir, T. *et al.* Final analysis from RESONATE: Up to six years of follow-up on ibrutinib in patients with previously treated chronic lymphocytic leukemia or small lymphocytic lymphoma. *Am J Hematol* **94**, 1353–1363 (2019).

91.   Goebel, L., Müller, M. P., Goody, R. S. & Rauh, D. KRasG12C inhibitors in clinical trials: a short historical perspective. *RSC Med Chem* **11**, 760–770 (2020).

92.   Moore, A. R., Rosenberg, S. C., McCormick, F. & Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nature Reviews Drug Discovery 2020 19:8* **19**, 533–552 (2020).

93.   Huang, L., Guo, Z., Wang, F. & Fu, L. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduction and Targeted Therapy 2021 6:1* **6**, 1–20 (2021).

94.   Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature 2013 503:7477* **503**, 548–551 (2013).

95.   Lito, P., Solomon, M., Li, L. S., Hansen, R. & Rosen, N. Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. *Science* **351**, 604–608 (2016).

96.   Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature 2019 575:7781* **575**, 217–223 (2019).

97.   Patricelli, M. P. *et al.* Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. *Cancer Discov* **6**, 316–329 (2016).

98.   Skoulidis, F. *et al.* Sotorasib for Lung Cancers with KRAS p.G12C Mutation . *New England Journal of Medicine* **384**, 2371–2381 (2021).

99.   Boike, L., Henning, N. J. & Nomura, D. K. Advances in covalent drug discovery. *Nature Reviews Drug Discovery 2022 21:12* **21**, 881–898 (2022).

100.   Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols 2013 8:11* **8**, 2281–2308 (2013).

101.   Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences* **95**, 5172–5177 (1998).

102.   Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology 2017 18:8* **18**, 495–506 (2017).

103.   Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 2008 456:7218* **456**, 53–59 (2008).

104.   Sequencing Technology | Sequencing by synthesis. https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html.

105.   Tu, J. *et al.* Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. *BMC Genomics* **13**, 1–9 (2012).

106. Smith, A. M. *et al.* Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res* **38**, e142 (2010).

107. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

108. Young, J. *et al.* A CRISPR platform for targeted in vivo screens identifies Toxoplasma gondii virulence factors in mice. *Nature Communications 2019 10:1* **10**, 1–11 (2019).

109. Porter, S. N., Baker, L. C., Mittelman, D. & Porteus, M. H. Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. *Genome Biol* **15**, R75 (2014).

110. Abel, S. *et al.* Sequence tag-based analysis of microbial population dynamics. *Nat Methods* **12**, 223–226 (2015).

111. Pfeiffer, J. K. & Kirkegaard, K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* **1**, 102–110 (2005).

112. Pfeiffer Karla, J. K. and K. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *PNAS* **103**, 5520–5525 (2006).

113. Grant, A. J. *et al.* Modelling within-Host Spatiotemporal Dynamics of Invasive Bacterial Disease. *PLoS Biol* **6**, 757–770 (2008).

114. Lim, C. H. *et al.* Independent Bottlenecks Characterize Colonization of Systemic Compartments and Gut Lymphoid Tissue by Salmonella. *PLoS Pathog* **10**, e1004270 (2014).

115. Lam, L. H. & Monack, D. M. Intraspecies Competition for Niches in the Distal Gut Dictate Transmission during Persistent Salmonella Infection. *PLoS Pathog* **10**, e1004527 (2014).

116. Abel, S., Abel zur Wiesch, P., Davis, B. M. & Waldor, M. K. Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog* **11**, (2015).

117. Gruner, B. M. *et al.* An in vivo multiplexed small-molecule screening platform. *Nat Methods* **13**, 883–889 (2016).

118. Binan, L., Drobetsky, E. A. & Costantino, S. Exploiting Molecular Barcodes in High-Throughput Cellular Assays. *SLAS Technol* **24**, 298–307 (2019).

119. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods 2012 9:7* **9**, 743–748 (2012).

120. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods 2012 9:7* **9**, 743–748 (2012).

121. Krutzik, P. O., Clutter, M. R., Trejo, A. & Nolan, G. P. Fluorescent Cell Barcoding for Multiplex Flow Cytometry. *Current protocols in cytometry / editorial board, J. Paul Robinson, managing editor ... [et al.]* **CHAPTER**, Unit (2011).

122. Sander, T., Freyss, J., Von Korff, M. & Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J Chem Inf Model* **55**, 460–473 (2015).

123. Giaever, G. *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002).

124. Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat Biotechnol* **34**, 419–423 (2016).

125. Baldwin, L. A. *et al.* DNA barcoding reveals ongoing immunoediting of clonal cancer populations during metastatic progression and immunotherapy response. *Nature Communications 2022 13:1* **13**, 1–18 (2022).

126. Rico, E., Jeacock, L., Kovářová, J. & Horn, D. Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. *Sci Rep* **8**, 1–10 (2018).

127. Shen, B., Brown, K. M., Lee, T. D. & Sibley, L. D. Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9. *mBio* **5**, 1–11 (2014).

128. Sidik, S. M., Hackett, C. G., Tran, F., Westwood, N. J. & Lourido, S. Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9. *PLoS One* **9**, e100450 (2014).

129. Donald, R. G. K. & Roos, D. S. Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. *Proc Natl Acad Sci U S A* **92**, 5749–5753 (1995).

130. Markovic, M., Ben-Shabat, S. & Dahan, A. Prodrugs for Improved Drug Delivery: Lessons Learned from Recently Developed and Marketed Products. *Pharmaceutics* **12**, 1–12 (2020).

131. Huynh, M. H. & Carruthers, V. B. Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryot Cell* **8**, 530–539 (2009).

132. Yamamoto, Y. & Gerbi, S. A. Making ends meet: Targeted integration of DNA fragments by genome editing. *Chromosoma* **127**, 405 (2018).

133. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537–W544 (2018).

134. Vincent, I. M. *et al.* A molecular mechanism for eflornithine resistance in African trypanosomes. *PLoS Pathog* **6**, e1001204 (2010).

135. Genois, M.-M. *et al.* DNA Repair Pathways in Trypanosomatids: from DNA Repair to Drug Resistance. *Microbiol Mol Biol Rev* **78**, 40 (2014).

136. Wu, Y., Lu, J. & Kang, T. Human single-stranded DNA binding proteins: guardians of genome stability. *Acta Biochim Biophys Sin (Shanghai)* **48**, 671–677 (2016).

137. Rauth, S. *et al.* Transfection and homologous recombination involving single-stranded DNA substrates in mammalian cells and nuclear extracts. *Proc Natl Acad Sci U S A* **83**, 5587 (1986).

138. Vinayak, S. *et al.* Genetic modification of the diarrhoeal pathogen Cryptosporidium parvum. *Nature* **523**, 477–480 (2015).

139. Mochizuki, K. High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. *Gene* **425**, 79–83 (2008).

140. Pawlowic, M. C., Vinayak, S., Sateriale, A., Brooks, C. F. & Striepen, B. Generating and maintaining transgenic Cryptosporidium parvum parasites. *Curr Protoc Microbiol* **46**, 20B.2.1 (2017).

141.   Kuss, S. K., Etheredge, C. A. & Pfeiffer, J. K. Multiple host barriers restrict poliovirus trafficking in mice. *PLoS Pathog* **4**, e1000082 (2008).

142.   Kaiser, P., Slack, E., Grant, A. J., Hardt, W. D. & Regoes, R. R. Lymph Node Colonization Dynamics after Oral Salmonella Typhimurium Infection in Mice. *PLoS Pathog* **9**, e1003532 (2013).

143.   Hullahalli, K., Pritchard, J. R. & Waldor, M. K. Refined Quantification of Infection Bottlenecks and Pathogen Dissemination with STAMPR. *mSystems* **6**, 1–16 (2021).

144.   Sidik, S. M. *et al.* A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423-1435.e12 (2016).

145.   Bushell, E. *et al.* Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260-272.e8 (2017).

146.   Beneke, T. & Gluenz, E. Bar-seq strategies for the LeishGEdit toolbox. *Mol Biochem Parasitol* **239**, (2020).

147.   Beneke, T. *et al.* Genetic dissection of a Leishmania flagellar proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathog* **15**, (2019).

148.   Baker, N. *et al.* Systematic functional analysis of Leishmania protein kinases identifies regulators of differentiation or survival. *Nat Commun* **12**, (2021).

149.   Alsford, S. *et al.* High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res* **21**, 915–924 (2011).

150.   Oberle, M., Balmer, O., Brun, R. & Roditi, I. Bottlenecks and the Maintenance of Minor Genotypes during the Life Cycle of Trypanosoma brucei. *PLoS Pathog* **6**, e1001023 (2010).

151.   Mugnier, M. R., Cross, G. A. M. & Papavasiliou, F. N. The in vivo dynamics of antigenic variation in Trypanosoma brucei. *Science* **347**, 1470–1473 (2015).

152.   Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J Biotechnol* **241**, 136–146 (2017).

153.   Schubert, M. S. *et al.* Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Scientific Reports 2021 11:1* **11**, 1–15 (2021).

154.   Wright, W. D., Shah, S. S. & Heyer, W. D. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry* **293**, 10524–10535 (2018).

155.   Pal, A. & Levy, Y. Structure, stability and specificity of the binding of ssDNA and ssRNA with proteins. *PLoS Comput Biol* **15**, 1–32 (2019).

156.   Levin, B. R., Lipsitch, M. & Bonhoeffer, S. Population biology, evolution, and infectious disease: Convergence and synthesis. *Science (1979)* **283**, 806–809 (1999).

157.   Moxon, R. & Kussell, E. The Impact of Bottlenecks on Microbial Survival, Adaptation and Phenotypic Switching in Host-Pathogen Interactions. doi:10.1111/evo.13370.

158.   Gutiérrez, S., Michalakis, Y. & Blanc, S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol* **2**, 546–555 (2012).

159. Pollak, E. A new method for estimating the effective population size from allele frequency changes. *Genetics* **104**, 531–548 (1983).

160. Luigi Luca Cavalli-Sforza, A. W. F. E. Phylogenetic Analysis Models and Estimation Procedures. *The American Journal of Human Genetics* **19**, 233–257 (1967).

161. Yusuf, Y. *et al.* A Viral-Vectored Multi-Stage Malaria Vaccine Regimen With Protective and Transmission-Blocking Efficacies. *Front Immunol* **10**, 2412 (2019).

162. Draper, S. J. *et al.* Malaria Vaccines: Recent Advances and New Horizons. *Cell Host Microbe* **24**, 43–56 (2018).

163. Wang, J. *et al.* Characterization of Pb51 in Plasmodium berghei as a malaria vaccine candidate targeting both asexual erythrocytic proliferation and transmission. *Malaria Journal 2017 16:1* **16**, 1–10 (2017).

164. Richie, T. L. & Saul, A. Progress and challenges for malaria vaccines. *Nature* **415**, 694–701 (2002).

165. Carter, R., Mendis, K. N., Miller, L. H., Molineaux, L. & Saul, A. Malaria transmission-blocking vaccines—how can their development be supported? *Nature Medicine 2000 6:3* **6**, 241–244 (2000).

166. Wohlfert, E. A., Blader, I. J. & Wilson, E. H. Brains and Brawn: Toxoplasma Infections of the Central Nervous System and Skeletal Muscle. *Trends Parasitol* **33**, 519–531 (2017).

167. Barrett, M. P., Kyle, D. E., Sibley, L. D., Radke, J. B. & Tarleton, R. L. Protozoan persister-like cells and drug treatment failure. *Nat Rev Microbiol* **17**, 607–620 (2019).

168. Profaci, C. P., Munji, R. N., Pulido, R. S. & Daneman, R. The blood-brain barrier in health and disease: Important unanswered questions. *J Exp Med* **217**, (2020).

169. Watson, G. F. & Davis, P. H. Systematic review and meta-analysis of variation in Toxoplasma gondii cyst burden in the murine model. *Exp Parasitol* **196**, 55–62 (2019).

170. McGovern, K. E., Cabral, C. M., Morrison, H. W. & Koshy, A. A. Aging with Toxoplasma gondii results in pathogen clearance, resolution of inflammation, and minimal consequences to learning and memory. *Sci Rep* **10**, 1–13 (2020).

171. Vargas, R. *et al.* In-host population dynamics of mycobacterium tuberculosis complex during active disease. *Elife* **10**, 1–37 (2021).

172. González-Candelas, F., Patiño-Galindo, J. Á. & Valiente-Mullor, C. Population Genomics of Human Viruses. *Population Genomics: Microorganisms* 267 (2019) doi:10.1007/13836_2018_31.

173. Plantinga, E. A., Bosch, G. & Hendriks, W. H. Estimation of the dietary nutrient profile of free-roaming feral cats: possible implications for nutrition of domestic cats. *Br J Nutr* **106 Suppl 1**, (2011).

174. Tenter, A. M., Heckeroth, A. R. & Weiss, L. M. Toxoplasma gondii: from animals to humans. *Int J Parasitol* **30**, 1217 (2000).

175. Pan, M. *et al.* The determinants regulating Toxoplasma gondii bradyzoite development. *Front Microbiol* **13**, 4488 (2022).

176. Merritt, E. F. *et al.* Transcriptional Profiling Suggests T Cells Cluster around Neurons Injected with Toxoplasma gondii Proteins. *mSphere* **5**, (2020).

177. Meyerhoff, J. *et al.* Microdissection of Mouse Brain into Functionally and Anatomically Different Regions. *J Vis Exp* **2021**, 1–14 (2021).

178. Ponte-Sucre, A. An Overview of Trypanosoma brucei Infections: An Intense Host–Parasite Interaction. *Front Microbiol* **7**, (2016).

179. Kiriiri, G. K., Njogu, P. M. & Mwangi, A. N. Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future Journal of Pharmaceutical Sciences 2020 6:1* **6**, 1–12 (2020).

180. Queiroz, E., Wolfender, J.-L. & Hostettmann, K. Modern approaches in the search for new lead antiparasitic compounds from higher plants. *Curr Drug Targets* **10**, 202–211 (2009).

181. Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* **307**, 223–227 (2005).

182. Ma, H. *et al.* Characterization of the metabolic activation of hepatitis C virus nucleoside inhibitor beta-D-2'-Deoxy-2'-fluoro-2'-C-methylcytidine (PSI-6130) and identification of a novel active 5'-triphosphate species. *J Biol Chem* **282**, 29812–29820 (2007).

183. Butterworth, S. *et al.* Toxoplasma gondii virulence factor ROP1 reduces parasite susceptibility to murine and human innate immune restriction. *PLoS Pathog* **18**, e1011021 (2022).

184. Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J Med Chem* **62**, 5673–5724 (2019).

185. Brogi, S. *et al.* Covalent Reversible Inhibitors of Cysteine Proteases Containing the Nitrile Warhead: Recent Advancement in the Field of Viral and Parasitic Diseases. *Molecules* **27**, (2022).

186. Ghasemi, M., Turnbull, T., Sebastian, S. & Kempson, I. The mtt assay: Utility, limitations, pitfalls, and interpretation in bulk and single-cell analysis. *Int J Mol Sci* **22**, (2021).

187. Benns, H. J. *et al.* CRISPR-based oligo recombineering prioritizes apicomplexan cysteines for drug discovery. *Nature Microbiology 2022 7:11* **7**, 1891–1905 (2022).

188. Roos, D. S., Donald, R. G. K., Morrissette, N. S. & Moulton, A. L. C. Chapter 3: Molecular Tools for Genetic Dissection of the Protozoan Parasite Toxoplasma gondii. *Methods Cell Biol* **45**, 27–63 (1995).

189. Ciallella, J. R. & Reaume, A. G. In vivo phenotypic screening: clinical proof of concept for a drug repositioning approach. *Drug Discov Today Technol* **23**, 45–52 (2017).

190. Thorarensen, A. *et al.* The advantages of describing covalent inhibitor in vitro potencies by IC50 at a fixed time point. IC50 determination of covalent inhibitors provides meaningful data to medicinal chemistry for SAR optimization. *Bioorg Med Chem* **29**, (2021).

191. Sarkar, I., Hauber, I., Hauber, J. & Buchholz, F. HIV-1 proviral DNA excision using an evolved recombinase. *Science (1979)* **316**, 1912–1915 (2007).

8. Appendices

8.1. Primers & Oligonucleotides

8.1.1. Barcoding Oligonucleotides for *T. gondii*

>Barcode 1	GATGTGTCATACCATGGAGTTTCCtgaAAACACtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 2	GATGTGTCATACCATGGAGTTTCCtgaAAACCGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 3	GATGTGTCATACCATGGAGTTTCCtgaAAACTCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 4	GATGTGTCATACCATGGAGTTTCCtgaAAAGACtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 5	GATGTGTCATACCATGGAGTTTCCtgaAAAGGCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 6	GATGTGTCATACCATGGAGTTTCCtgaAAAGTGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 7	GATGTGTCATACCATGGAGTTTCCtgaAACGTCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 8	GATGTGTCATACCATGGAGTTTCCtgaAACTGAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 9	GATGTGTCATACCATGGAGTTTCCtgaAAGAGAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 10	GATGTGTCATACCATGGAGTTTCCtgaAAGCACtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 11	GATGTGTCATACCATGGAGTTTCCtgaAAGCTCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 12	GATGTGTCATACCATGGAGTTTCCtgaAAGGTGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 13	GATGTGTCATACCATGGAGTTTCCtgaAATACCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 14	GATGTGTCATACCATGGAGTTTCCtgaAATAGCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 15	GATGTGTCATACCATGGAGTTTCCtgaACAAAGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 16	GATGTGTCATACCATGGAGTTTCCtgaACAACCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 17	GATGTGTCATACCATGGAGTTTCCtgaACAGCAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 18	GATGTGTCATACCATGGAGTTTCCtgaACATGTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 19	GATGTGTCATACCATGGAGTTTCCtgaACCATAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 20	GATGTGTCATACCATGGAGTTTCCtgaACCTTTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 21	GATGTGTCATACCATGGAGTTTCCtgaACGGTTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 22	GATGTGTCATACCATGGAGTTTCCtgaACTCAAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 23	GATGTGTCATACCATGGAGTTTCCtgaACTTCCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 24	GATGTGTCATACCATGGAGTTTCCtgaAGAAACtagGGCGAGTCGATGGAA AGCGGCTTG

>Barcode 25	GATGTGTCATACCATGGAGTTTCCtgaAGAAGGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 26	GATGTGTCATACCATGGAGTTTCCtgaAGACGTtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 27	GATGTGTCATACCATGGAGTTTCCtgaAGGAAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 28	GATGTGTCATACCATGGAGTTTCCtgaAGGTAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 29	GATGTGTCATACCATGGAGTTTCCtgaAGTGGAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 30	GATGTGTCATACCATGGAGTTTCCtgaAGTTACtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 31	GATGTGTCATACCATGGAGTTTCCtgaATCAACtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 32	GATGTGTCATACCATGGAGTTTCCtgaATCTGAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 33	GATGTGTCATACCATGGAGTTTCCtgaATCTTCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 34	GATGTGTCATACCATGGAGTTTCCtgaATGCACtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 35	GATGTGTCATACCATGGAGTTTCCtgaATGTTGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 36	GATGTGTCATACCATGGAGTTTCCtgaCAAACAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 37	GATGTGTCATACCATGGAGTTTCCtgaCAATACtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 38	GATGTGTCATACCATGGAGTTTCCtgaCACAAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 39	GATGTGTCATACCATGGAGTTTCCtgaCACATTtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 40	GATGTGTCATACCATGGAGTTTCCtgaCACGAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 41	GATGTGTCATACCATGGAGTTTCCtgaCAGTTGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 42	GATGTGTCATACCATGGAGTTTCCtgaCATCTAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 43	GATGTGTCATACCATGGAGTTTCCtgaCCAAATtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 44	GATGTGTCATACCATGGAGTTTCCtgaCCGTATtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 45	GATGTGTCATACCATGGAGTTTCCtgaCCTAATtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 46	GATGTGTCATACCATGGAGTTTCCtgaCGCAAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 47	GATGTGTCATACCATGGAGTTTCCtgaCGTTTCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 48	GATGTGTCATACCATGGAGTTTCCtgaCTAAAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 49	GATGTGTCATACCATGGAGTTTCCtgaCTATGGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 50	GATGTGTCATACCATGGAGTTTCCtgaCTCATGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 51	GATGTGTCATACCATGGAGTTTCCtgaCTCCTTtagGGCGAGTCGATGGAAAGCGGCTTG

>Barcode 52	GATGTGTCATACCATGGAGTTTCCtgaCTGAAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 53	GATGTGTCATACCATGGAGTTTCCtgaCTGGTAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 54	GATGTGTCATACCATGGAGTTTCCtgaCTTCCAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 55	GATGTGTCATACCATGGAGTTTCCtgaCTTCGAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 56	GATGTGTCATACCATGGAGTTTCCtgaCTTGACtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 57	GATGTGTCATACCATGGAGTTTCCtgaCTTGTTtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 58	GATGTGTCATACCATGGAGTTTCCtgaCTTTAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 59	GATGTGTCATACCATGGAGTTTCCtgaGAAACCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 60	GATGTGTCATACCATGGAGTTTCCtgaGAAAGGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 61	GATGTGTCATACCATGGAGTTTCCtgaGAAATCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 62	GATGTGTCATACCATGGAGTTTCCtgaGAACAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 63	GATGTGTCATACCATGGAGTTTCCtgaGAAGAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 64	GATGTGTCATACCATGGAGTTTCCtgaGACAATtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 65	GATGTGTCATACCATGGAGTTTCCtgaGAGATAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 66	GATGTGTCATACCATGGAGTTTCCtgaGATACCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 67	GATGTGTCATACCATGGAGTTTCCtgaGATATGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 68	GATGTGTCATACCATGGAGTTTCCtgaGATGAAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 69	GATGTGTCATACCATGGAGTTTCCtgaGATTTCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 70	GATGTGTCATACCATGGAGTTTCCtgaGCACTAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 71	GATGTGTCATACCATGGAGTTTCCtgaGCGTTTtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 72	GATGTGTCATACCATGGAGTTTCCtgaGCTAAGtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 73	GATGTGTCATACCATGGAGTTTCCtgaGCTCTTtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 74	GATGTGTCATACCATGGAGTTTCCtgaGCTGTAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 75	GATGTGTCATACCATGGAGTTTCCtgaGGTCTAtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 76	GATGTGTCATACCATGGAGTTTCCtgaGGTTTCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 77	GATGTGTCATACCATGGAGTTTCCtgaGTAATCtagGGCGAGTCGATGGAAAGCGGCTTG
>Barcode 78	GATGTGTCATACCATGGAGTTTCCtgaGTACAAtagGGCGAGTCGATGGAAAGCGGCTTG

>Barcode 79	GATGTGTCATACCATGGAGTTTCCtgaGTAGAGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 80	GATGTGTCATACCATGGAGTTTCCtgaGTGATCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 81	GATGTGTCATACCATGGAGTTTCCtgaGTTAGTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 82	GATGTGTCATACCATGGAGTTTCCtgaGTTCACtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 83	GATGTGTCATACCATGGAGTTTCCtgaGTTGCTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 84	GATGTGTCATACCATGGAGTTTCCtgaGTTTGGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 85	GATGTGTCATACCATGGAGTTTCCtgaTAACCCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 86	GATGTGTCATACCATGGAGTTTCCtgaTAAGCCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 87	GATGTGTCATACCATGGAGTTTCCtgaTAAGGCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 88	GATGTGTCATACCATGGAGTTTCCtgaTACAGAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 89	GATGTGTCATACCATGGAGTTTCCtgaTACTCAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 90	GATGTGTCATACCATGGAGTTTCCtgaTAGAACtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 91	GATGTGTCATACCATGGAGTTTCCtgaTAGCGAtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 92	GATGTGTCATACCATGGAGTTTCCtgaTAGTCTtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 93	GATGTGTCATACCATGGAGTTTCCtgaTATGCCtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 94	GATGTGTCATACCATGGAGTTTCCtgaTATTGGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 95	GATGTGTCATACCATGGAGTTTCCtgaTCAAAGtagGGCGAGTCGATGGAA AGCGGCTTG
>Barcode 96	GATGTGTCATACCATGGAGTTTCCtgaTCAGTCtagGGCGAGTCGATGGAA AGCGGCTTG

8.1.2. Barcoding Oligonucleotides for *T. brucei*

>Barcode 1	TGCAATCAGAAGACGAGGTTTAAGtagAAACACtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 2	TGCAATCAGAAGACGAGGTTTAAGtagAAACCGtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 3	TGCAATCAGAAGACGAGGTTTAAGtagAAACTCtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 4	TGCAATCAGAAGACGAGGTTTAAGtagAAAGACtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 5	TGCAATCAGAAGACGAGGTTTAAGtagAAAGGCtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 6	TGCAATCAGAAGACGAGGTTTAAGtagAAAGTGtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 7	TGCAATCAGAAGACGAGGTTTAAGtagAACGTCtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 8	TGCAATCAGAAGACGAGGTTTAAGtagAACTGAtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 9	TGCAATCAGAAGACGAGGTTTAAGtagAAGAGAtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 10	TGCAATCAGAAGACGAGGTTTAAGtagAAGCACtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 11	TGCAATCAGAAGACGAGGTTTAAGtagAAGCTCtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 12	TGCAATCAGAAGACGAGGTTTAAGtagAAGGTGtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 13	TGCAATCAGAAGACGAGGTTTAAGtagAATACCtgaCTCACACTAACCGTTTCGATTTAC

>B arc ode 14	TGCAATCAGAAGACGAGGTTTAAGtagAATAGCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 15	TGCAATCAGAAGACGAGGTTTAAGtagACAAAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 16	TGCAATCAGAAGACGAGGTTTAAGtagACAACCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 17	TGCAATCAGAAGACGAGGTTTAAGtagACAGCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 18	TGCAATCAGAAGACGAGGTTTAAGtagACATGTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 19	TGCAATCAGAAGACGAGGTTTAAGtagACCATtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 20	TGCAATCAGAAGACGAGGTTTAAGtagACCTTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 21	TGCAATCAGAAGACGAGGTTTAAGtagACGGTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 22	TGCAATCAGAAGACGAGGTTTAAGtagACTCAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 23	TGCAATCAGAAGACGAGGTTTAAGtagACTTCCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 24	TGCAATCAGAAGACGAGGTTTAAGtagAGAAACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 25	TGCAATCAGAAGACGAGGTTTAAGtagAGAAGGtgaCTCACACTAACCGTTTCGA TTTAC
>B arc ode 26	TGCAATCAGAAGACGAGGTTTAAGtagAGACGTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc	TGCAATCAGAAGACGAGGTTTAAGtagAGGAAAtgaCTCACACTAACCGTTTCGAT TTAC

ode 27	
>B arc ode 28	TGCAATCAGAAGACGAGGTTTAAGtagAGGTAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 29	TGCAATCAGAAGACGAGGTTTAAGtagAGTGGAtgaCTCACACTAACCGTTTCGA TTTAC
>B arc ode 30	TGCAATCAGAAGACGAGGTTTAAGtagAGTTACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 31	TGCAATCAGAAGACGAGGTTTAAGtagATCAACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 32	TGCAATCAGAAGACGAGGTTTAAGtagATCTGAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 33	TGCAATCAGAAGACGAGGTTTAAGtagATCTTCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 34	TGCAATCAGAAGACGAGGTTTAAGtagATGCACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 35	TGCAATCAGAAGACGAGGTTTAAGtagATGTTGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 36	TGCAATCAGAAGACGAGGTTTAAGtagCAAACAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 37	TGCAATCAGAAGACGAGGTTTAAGtagCAATACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 38	TGCAATCAGAAGACGAGGTTTAAGtagCACAAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 39	TGCAATCAGAAGACGAGGTTTAAGtagCACATTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 40	TGCAATCAGAAGACGAGGTTTAAGtagCACGAAtgaCTCACACTAACCGTTTCGAT TTAC

>B arc ode 41	TGCAATCAGAAGACGAGGTTTAAGtagCAGTTGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 42	TGCAATCAGAAGACGAGGTTTAAGtagCATCTAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 43	TGCAATCAGAAGACGAGGTTTAAGtagCCAAATtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 44	TGCAATCAGAAGACGAGGTTTAAGtagCCGTATtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 45	TGCAATCAGAAGACGAGGTTTAAGtagCCTAATtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 46	TGCAATCAGAAGACGAGGTTTAAGtagCGCAAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 47	TGCAATCAGAAGACGAGGTTTAAGtagCGTTTCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 48	TGCAATCAGAAGACGAGGTTTAAGtagCTAAAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 49	TGCAATCAGAAGACGAGGTTTAAGtagCTATGGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 50	TGCAATCAGAAGACGAGGTTTAAGtagCTCATGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 51	TGCAATCAGAAGACGAGGTTTAAGtagCTCCTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 52	TGCAATCAGAAGACGAGGTTTAAGtagCTGAAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 53	TGCAATCAGAAGACGAGGTTTAAGtagCTGGTAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc	TGCAATCAGAAGACGAGGTTTAAGtagCTTCCAtgaCTCACACTAACCGTTTCGAT TTAC

ode 54	
>B arc ode 55	TGCAATCAGAAGACGAGGTTTAAGtagCTTCGAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 56	TGCAATCAGAAGACGAGGTTTAAGtagCTTGACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 57	TGCAATCAGAAGACGAGGTTTAAGtagCTTGTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 58	TGCAATCAGAAGACGAGGTTTAAGtagCTTTAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 59	TGCAATCAGAAGACGAGGTTTAAGtagGAAACCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 60	TGCAATCAGAAGACGAGGTTTAAGtagGAAAGGtgaCTCACACTAACCGTTTCGA TTTAC
>B arc ode 61	TGCAATCAGAAGACGAGGTTTAAGtagGAAATCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 62	TGCAATCAGAAGACGAGGTTTAAGtagGAACAGtgaCTCACACTAACCGTTTCGA TTTAC
>B arc ode 63	TGCAATCAGAAGACGAGGTTTAAGtagGAAGAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 64	TGCAATCAGAAGACGAGGTTTAAGtagGACAATtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 65	TGCAATCAGAAGACGAGGTTTAAGtagGAGATAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 66	TGCAATCAGAAGACGAGGTTTAAGtagGATACCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 67	TGCAATCAGAAGACGAGGTTTAAGtagGATATGtgaCTCACACTAACCGTTTCGAT TTAC

>B arc ode 68	TGCAATCAGAAGACGAGGTTTAAGtagGATGAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 69	TGCAATCAGAAGACGAGGTTTAAGtagGATTTCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 70	TGCAATCAGAAGACGAGGTTTAAGtagGCACTAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 71	TGCAATCAGAAGACGAGGTTTAAGtagGCGTTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 72	TGCAATCAGAAGACGAGGTTTAAGtagGCTAAGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 73	TGCAATCAGAAGACGAGGTTTAAGtagGCTCTTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 74	TGCAATCAGAAGACGAGGTTTAAGtagGCTGTAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 75	TGCAATCAGAAGACGAGGTTTAAGtagGGTCTAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 76	TGCAATCAGAAGACGAGGTTTAAGtagGGTTTCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 77	TGCAATCAGAAGACGAGGTTTAAGtagGTAATCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 78	TGCAATCAGAAGACGAGGTTTAAGtagGTACAAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 79	TGCAATCAGAAGACGAGGTTTAAGtagGTAGAGtgaCTCACACTAACCGTTTCGA TTTAC
>B arc ode 80	TGCAATCAGAAGACGAGGTTTAAGtagGTGATCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc	TGCAATCAGAAGACGAGGTTTAAGtagGTTAGTtgaCTCACACTAACCGTTTCGAT TTAC

ode 81	
>B arc ode 82	TGCAATCAGAAGACGAGGTTTAAGtagGTTCACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 83	TGCAATCAGAAGACGAGGTTTAAGtagGTTGCTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 84	TGCAATCAGAAGACGAGGTTTAAGtagGTTTGGtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 85	TGCAATCAGAAGACGAGGTTTAAGtagTAACCCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 86	TGCAATCAGAAGACGAGGTTTAAGtagTAAGCCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 87	TGCAATCAGAAGACGAGGTTTAAGtagTAAGGCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 88	TGCAATCAGAAGACGAGGTTTAAGtagTACAGAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 89	TGCAATCAGAAGACGAGGTTTAAGtagTACTCAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 90	TGCAATCAGAAGACGAGGTTTAAGtagTAGAACtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 91	TGCAATCAGAAGACGAGGTTTAAGtagTAGCGAtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 92	TGCAATCAGAAGACGAGGTTTAAGtagTAGTCTtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 93	TGCAATCAGAAGACGAGGTTTAAGtagTATGCCtgaCTCACACTAACCGTTTCGAT TTAC
>B arc ode 94	TGCAATCAGAAGACGAGGTTTAAGtagTATTGGtgaCTCACACTAACCGTTTCGAT TTAC

>Barcode 95	TGCAATCAGAAGACGAGGTTTAAGtagTCAAAGtgaCTCACACTAACCGTTTCGATTTAC
>Barcode 96	TGCAATCAGAAGACGAGGTTTAAGtagTCAGTCtgaCTCACACTAACCGTTTCGATTTAC
24 nt homology trial	TGCAATCAGAAGACGAGGTTTAAGtagAAACACtgaCTCACACTAACCGTTTCGATTTAC
40 nt homology trial	CCCTGCACACACCGGATGCAATCAGAAGACGAGGTTTAAGtagAAACACtgaCTCACACTAACCGTTTCGATTTACTTGCTTTATGGTATCA
51 nt homology trial	GCGATCGTCCTTCCTGCACACACCGGATGCAATCAGAAGACGAGGTTTAAGtagAAAGTGtgaCTCACACTAACCGTTTCGATTTACTTGCTTTATGGTATCACAGGGTACCTT
81 nt homology trial	GCGAGTAACCTGATGGGACTTGAGGGAATTGCGATCGTCCTTCCTGCACACACCGGATGCAATCAGAAGACGAGGTTTAAGtagAAAGTGtgaCTCACACTAACCGTTTCGATTTACTTGCTTTATGGTATCACAGGGTACCTTGCATACGGAACGTCAATCAATACGAGCATT

8.1.3. Barcoding Oligonucleotides for *C. parvum*

24 nt homology trial	GGAAATTTATTTGAAGGAAGTAAAtagAAACACtgaGATAAACTTACTGAAATCAAAACT
40 nt homology trial	GAGAACAGACTTTAAGGGAAATTTATTTGAAGGAAGTAAAtagAAACACtgaGATAAACTTACTGAAATCAAAACTATTTGTCGCTGTGGCA
51 nt homology trial	TGCTATGGTTTGAGAACAGACTTTAAGGGAAATTTATTTGAAGGAAGTAAAtagAAACACtgaGATAAACTTACTGAAATCAAAACTATTTGTCGCTGTGGCAAAAAAGCTACC

8.1.4. PCR Primers

Tg*UPRT* Specific FWD Primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtggatgtgtcataccatggagtttcctg
Tg*UPRT* Specific REV Primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgttttagtgtaacaaagtggacagcagc
T. gondii Amplicon Specific FWD Primer	ATATGTAGATATGAGTTTGAACGCG
T. gondii Amplicon Specific REV Primer	TTGCCTATCGCTCTTCACC
Tb*AAT6* Specific FWD Primer	ATGGAGATAAACAAGAACGG
Tb*AAT6* Specific REV Primer	ACTGCACGGGGTATGTGC
T. brucei Amplicon Specific FWD Primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGCAATCAGAAGACGAGGTTTAAGtag
T. brucei Amplicon Specific REV Primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCGAGTTCAATACTAACAATAACCAGG
Cp*TK* Specific FWD Primer	AGGCTCAATTTGCTCTAGAATTGG

Cp*TK* Specific REV Primer	GCATACAGAAGTATAAATAGAATTGTCACC
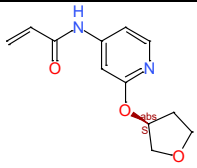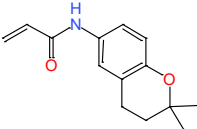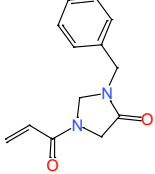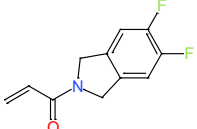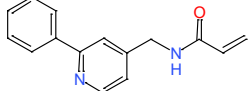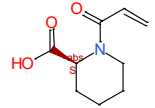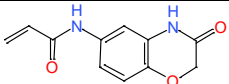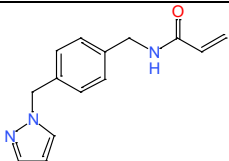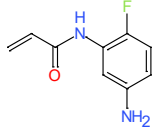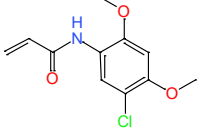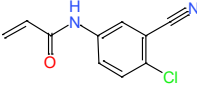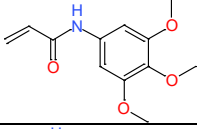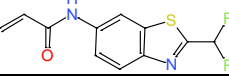
8.1.5. NGS Indexing Primers
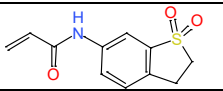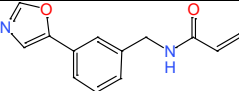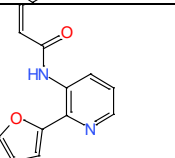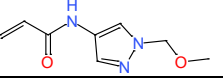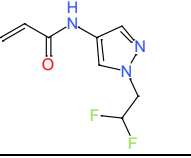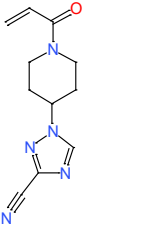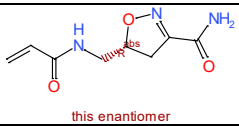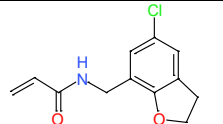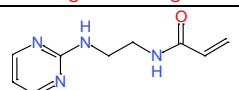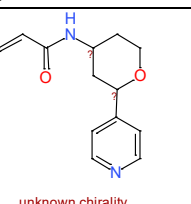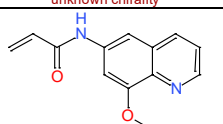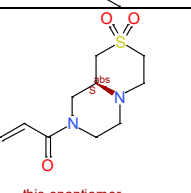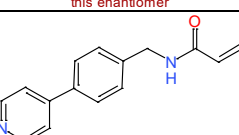
TCGCCTTA	N701
CTAGTACG	N702
TTCTGCCT	N703
GCTCAGGA	N704
AGGAGTCC	N705
CATGCCTA	N706
GTAGAGAG	N707
CAGCCTCG	N710
TGCCTCTT	N711
TCCTCTAC	N712
TCATGAGC	N714
CCTGAGAT	N715
TAGCGAGT	N716
GTAGCTCC	N718
TACTACGC	N719
AGGCTCCG	N720
GCAGCGTA	N721
CTGCGCAT	N722
GAGCGCTA	N723
CGCTCAGT	N724
GTCTTAGG	N726
ACTGATCG	N727
TAGCTGCA	N728
GACGTCGA	N729
CTCTCTAT	S502
TATCCTCT	S503
GTAAGGAG	S505
ACTGCATA	S506
AAGGAGTA	S507
CTAAGCCT	S508
CGTCTAAT	S510
TCTCTCCG	S511
TCGACTAG	S513
TTCTAGCT	S515
CCTAGAGT	S516
GCGTAAGA	S517
CTATTAAG	S518
AAGGCTAT	S520
GAGCCTTA	S521
TTATGCGA	S522
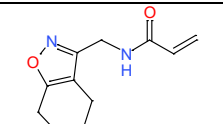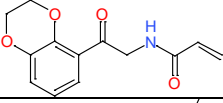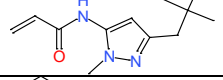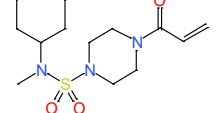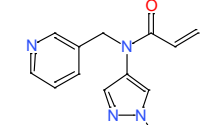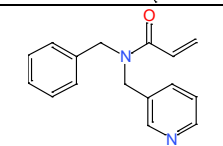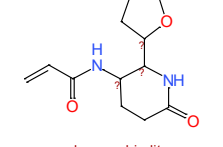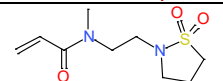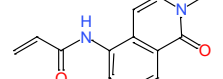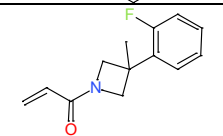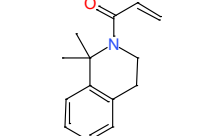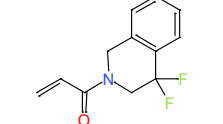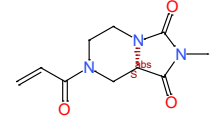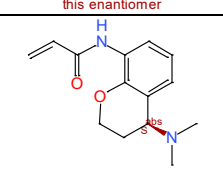
8.2. Tg*UPRT* Targeting Plasmid

Plasmid map: **pSAG1::CAS9-U6::sgUPRT** 9674 bp

(9635 .. 9651) pBluescriptSK
(9640) **SpeI**
(9608) **HpaI**
(9602) **AgeI**
(9408) **AarI**
(9318) **PacI**
(9247 .. 9268) EGFP-C
(8639 .. 8660) EGFP-N

NotI (9653)
AvrII (290)
BbsI (363)
BstEII (483)
M13 Reverse (697 .. 713)
M13/pUC Reverse (710 .. 732)
L4440 (928 .. 945)
pBR322ori-F (1179 .. 1198)

(7735) **KflI**
(7147) **PmlI**
(6211) **EcoRV**
(6127) **PasI**

BsaI (1992)
Amp-R (2481 .. 2500)
F1ori-F (2986 .. 3007)
F1ori-R (3198 .. 3217)
M13/pUC Forward (3431 .. 3453)
M13 Forward (3445 .. 3462)
BssHII (3465)
T7 (3472 .. 3491)
Acc65I (3499)
KpnI (3503)
AbsI - **PaeR7I** - **PspXI** - **XhoI** (3514)
pBluescriptKS (3515 .. 3531)
RsrII (3563)
AfeI (3886)
NheI (3887)
BmtI (3891)
NsiI (4347)
BglII (4608)

Labels: SiriusGFP, gRNA scaffold, CAP binding site, M13 rev, ori, HA, SV40 NLS, Cas9, AmpR, AmpR promoter, f1 ori, T7 promoter

8.3. Initial Covalent Fragment Library
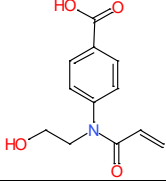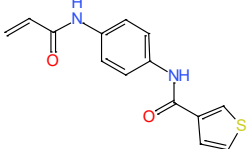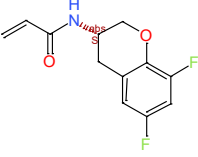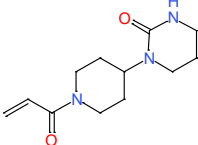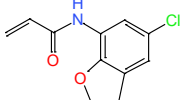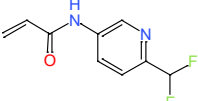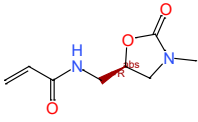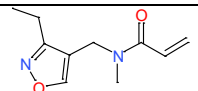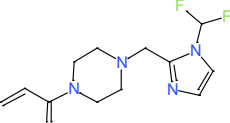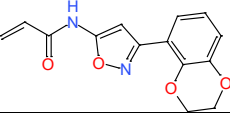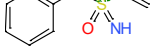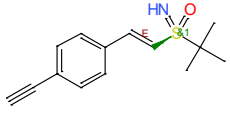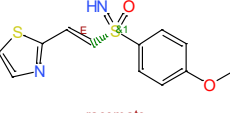
Structure No	Smiles	Structure
1	C=CC(Nc1cc(O[C@@H]2COCC2)ncc1)=O	
2	CC(C)(CC1)Oc(cc2)c1cc2NC(C=C)=O	
3	C=CC(N(C1)CN(Cc2ccccc2)C1=O)=O	
4	C=CC(N(Cc1c2)Cc1cc(F)c2F)=O	
5	C=CC(NCc1cc(-c2ccccc2)ncc1)=O	
6	C=CC(N(CCCC1)[C@@H]1C(O)=O)=O	
7	C=CC(Nc(cc1)cc(N2)c1OCC2=O)=O	
8	C=CC(NCc1ccc(Cn2nccc2)cc1)=O	
9	C=CC(Nc(cc(cc1)N)c1F)=O	
10	COc(c(NC(C=C)=O)c1)cc(OC)c1Cl	
11	C=CC(Nc(cc1)cc(C#N)c1Cl)=O	
12	COc1cc(NC(C=C)=O)cc(OC)c1OC	
13	C=CC(Nc(cc1)cc2c1nc(C(F)F)s2)=O	
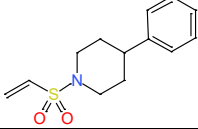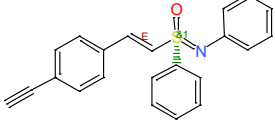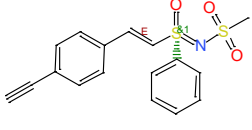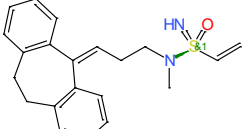
14	C=CC(Nc1cc(S(CC2)(=O)=O)c2cc1)=O	
15	C=CC(NCc1cc(-c2cnco2)ccc1)=O	
16	C=CC(Nc1c(-c2ccco2)nccc1)=O	
17	COCn1ncc(NC(C=C)=O)c1	
18	C=CC(Nc1cn(CC(F)F)nc1)=O	
19	C=CC(N(CC1)CCC1n1nc(C#N)nc1)=O	
20	C=CC(NC[C@@H](C1)ON=C1C(N)=O)=O	 this enantiomer
21	C=CC(NCc1c2OCCc2cc(Cl)c1)=O	
22	C=CC(NCCNc1ncccn1)=O	
23	C=CC(NC1CC(c2ccncc2)OCC1)=O	 unknown chirality
24	COc1cc(NC(C=C)=O)cc2cccnc12	
25	C=CC(N(CC1)C[C@@H](C2)N1CCS2(=O)=O)=O	 this enantiomer
26	C=CC(NCc(cc1)ccc1-c1ccncc1)=O	

| 27	C=CC(NCc1noc2c1CCCC2)=O	
28	C=CC(NCC(c1c2OCCOc2ccc1)=O)=O	
29	CC(C)(C)Cc1nn(C)c(NC(C=C)=O)c1	
30	CN(C1CCCCC1)S(N(CC1)CCN1C(C=C)=O)(=O)=O	
31	Cn1ncc(N(Cc2cnccc2)C(C=C)=O)c1	
32	C=CC(N(Cc1ccccc1)Cc1cnccc1)=O	
33	C=CC(NC(CC1)C(C2OCCC2)NC1=O)=O	unknown chirality
34	CN(CCN(CCC1)S1(=O)=O)C(C=C)=O	
35	CN(C=Cc1c2cccc1NC(C=C)=O)C2=O	
36	CC(C1)(CN1C(C=C)=O)c(cccc1)c1F	
37	CC(C)(c1c(CC2)cccc1)N2C(C=C)=O	
38	C=CC(N(CC1(F)F)Cc2c1cccc2)=O	
39	CN(C([C@H](C1)N2CCN1C(C=C)=O)=O)C2=O	this enantiomer
40	CN(C)[C@@H]1c2cccc(NC(C=C)=O)c2OCC1	this enantiomer

41	C=CC(Nc1cc(C2CCOCC2)ccc1)=O	
42	C=CC(N1c2cccc(C(N)=O)c2CCC1)=O	
43	C=CC(Nc1cnc(C(F)(F)F)nc1)=O	
44	C=CC(N(C1)CC1(C1)CS1(=O)=O)=O	
45	COc1cc(CCN(CC2)C(C=C)=O)c2c(OC)c1OC	
46	CN(C=CN=C1N(CC2)CCN2C(C=C)=O)C1=O	
47	C=CC(N(CC1)CCN1C(CN1CCCC1)=O)=O	
48	C=CC(N(CCC1)c2c1c(OCC(N)=O)ccc2)=O	
49	COCc1nc(cc(cc2)NC(C=C)=O)c2o1	
50	CN(c(cc(cc1)NC(C=C)=O)c1C=C1)C1=O	
51	COCCN(CCN(C1)C(C=C)=O)C1=O	
52	C=CC(Nc1cn2nccc2nc1)=O	
53	C=CC(Nc1noc(C2CC2)n1)=O	

54	CCN(Cc1cn(C)nn1)C(C=C)=O	
55	C=CC(Nc(cc1)cc(N2)c1SCC2=O)=O	
56	C=CC(N(CC1)C[C@@H](C(N)=O)N1C1CC1)=O	
57	C=CC(N(CCc1ccc2)Cc1c2Cl)=O	
58	C=CC(NC(CC1)CCC1(C1)C1(F)F)=O	
59	C=CC(NCc(cc1)cnc1C(N)=O)=O	
60	C=CC(NCCc(cc1)ncc1F)=O	
61	C=CC(NC(CC1)CCC1Oc(cc1)ncc1C#N)=O	
62	Cn(c(OC(F)F)c1)nc1NC(C=C)=O	
63	C=CC(NCc1nc(CCC2)c2cn1)=O	
64	Cc(c1c2)noc1ncc2NC(C=C)=O	
65	C=CC(NCCN(CCN1)CC1=O)=O	
66	C=CC(NCc1nc(C(N)=O)co1)=O	
67	Cc(s1)cn2c1nc(C)c2CN(C)C(C=C)=O	

68	C=CC(N(CCO)c(cc1)ccc1C(O)=O)=O	
69	C=CC(Nc(cc1)ccc1NC(c1cscc1)=O)=O	
70	C=CC(N[C@@H](Cc1cc(F)c2)COc1c2F)=O	
71	C=CC(N(CC1)CCC1N(CCCN1)C1=O)=O	
72	C=CC(Nc1c2OCCc2cc(Cl)c1)=O	
73	C=CC(Nc1cnc(C(F)F)cc1)=O	
74	CN(C[C@@H](CNC(C=C)=O)O1)C1=O	
75	CCc1nocc1CN(C)C(C=C)=O	
76	C=CC(N1CCN(Cc2nccn2C(F)F)CC1)=O	
77	C=CC(Nc1cc(-c2c3OCCOc3ccc2)no1)=O	
78	O=S(C=C)(CC1=CC=CC=C1)=N	
79	N=S(/C=C/C1=CC=C(C#C)C=C1)(C(C)(C)C)=O	
80	N=S(/C=C/C1=NC=CS1)(C2=CC=C(OC)C=C2)=O	

81	C=CS(N1CCC(C2=CC=CC=C2)CC1)(=O)=O	
82	O=S(/C=C/C1=CC=C(C#C)C=C1)(C2=CC=CC=C2)=NC3=CC=CC=C3	 racemate
83	O=S(/C=C/C1=CC=C(C#C)C=C1)(C2=CC=CC=C2)=NS(C)(=O)=O	 racemate
84	CN(S(C=C)(=N)=O)CC/C=C1C2=CC=CC=C2CCC3=CC=CC=C\13	 racemate

8.4. Dark-Cycling
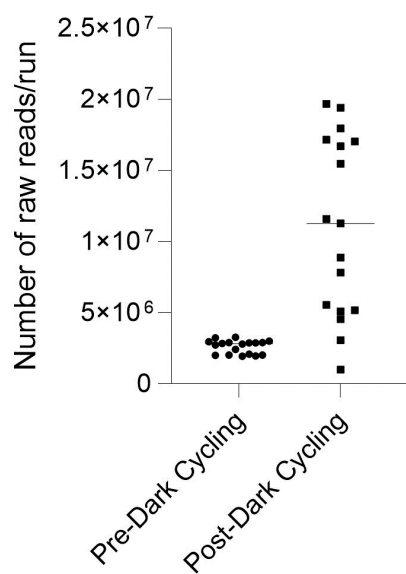
The number of reads obtained in both pre- and post-dark cycling runs. Means were determined to be statistically significant by a one-sample t-test.