



## University of Dundee

### **New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites**

Peña, Imanol; Pilar Manzano, M.; Cantizani, Juan; Kessler, Albane; Alonso-Padilla, Julio; Bardera, Ana I.; Alvarez, Emilio; Colmenarejo, Gonzalo; Cotillo, Ignacio; Roquero, Irene; de Dios-Anton, Francisco; Barroso, Vanessa; Rodriguez, Ana; Gray, David W.; Navarro, Miguel; Kumar, Vinod; Sherstnev, Alexander; Drewry, David H.; Brown, James R.; Fiandor, Jose M.; Julio Martin, J.

*Published in:*  
Scientific Reports

*DOI:*  
[10.1038/srep08771](https://doi.org/10.1038/srep08771)

*Publication date:*  
2015

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

#### *Citation for published version (APA):*

Peña, I., Pilar Manzano, M., Cantizani, J., Kessler, A., Alonso-Padilla, J., Bardera, A. I., ... Julio Martin, J. (2015). New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource. *Scientific Reports*, 5, [8771]. [10.1038/srep08771](https://doi.org/10.1038/srep08771)

#### **General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.



OPEN

SUBJECT AREAS:  
HIGH-THROUGHPUT  
SCREENING  
PARASITOLOGYReceived  
3 December 2014Accepted  
28 January 2015Published  
5 March 2015Correspondence and  
requests for materials  
should be addressed to  
J.J.M. (julio.j.martin@  
gsk.com)

# New Compound Sets Identified from High Throughput Phenotypic Screening Against Three Kinetoplastid Parasites: An Open Resource

Imanol Peña<sup>1</sup>, M. Pilar Manzano<sup>2</sup>, Juan Cantizani<sup>2</sup>, Albane Kessler<sup>2</sup>, Julio Alonso-Padilla<sup>3</sup>, Ana I. Bardera<sup>1</sup>, Emilio Alvarez<sup>1</sup>, Gonzalo Colmenarejo<sup>1</sup>, Ignacio Cotillo<sup>2</sup>, Irene Roquero<sup>1</sup>, Francisco de Dios-Anton<sup>1</sup>, Vanessa Barroso<sup>1</sup>, Ana Rodriguez<sup>3</sup>, David W. Gray<sup>4</sup>, Miguel Navarro<sup>5</sup>, Vinod Kumar<sup>6</sup>, Alexander Sherstnev<sup>7</sup>, David H. Drewry<sup>8</sup>, James R. Brown<sup>6</sup>, Jose M. Fiandor<sup>2</sup> & J. Julio Martin<sup>1</sup>

<sup>1</sup>Molecular Discovery Research, Tres Cantos Medicines Development Campus, GlaxoSmithKline, Tres Cantos, Spain, <sup>2</sup>Diseases of the Developing World (DDW), Tres Cantos Medicines Development Campus, GlaxoSmithKline, Tres Cantos, Spain, <sup>3</sup>Department of Microbiology, Division of Parasitology, New York University School of Medicine, New York, NY, USA, <sup>4</sup>Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dundee, UK, <sup>5</sup>Instituto de Parasitología y Biomedicina "López-Neyra" Consejo Superior de Investigaciones Científicas, Granada, Spain, <sup>6</sup>Computational Biology, Quantitative Sciences, GlaxoSmithKline, Collegeville, PA, USA, <sup>7</sup>Computational Biology, Quantitative Sciences, GlaxoSmithKline, Medicines Research Center, Stevenage, Hertfordshire, UK, <sup>8</sup>Chemical Sciences, Molecular Discovery Research, GlaxoSmithKline, Research Triangle Park, NC, USA.

Using whole-cell phenotypic assays, the GlaxoSmithKline high-throughput screening (HTS) diversity set of 1.8 million compounds was screened against the three kinetoplastids most relevant to human disease, i.e. *Leishmania donovani*, *Trypanosoma cruzi* and *Trypanosoma brucei*. Secondary confirmatory and orthogonal intracellular anti-parasitocidal assays were conducted, and the potential for non-specific cytotoxicity determined. Hit compounds were chemically clustered and triaged for desirable physicochemical properties. The hypothetical biological target space covered by these diversity sets was investigated through bioinformatics methodologies. Consequently, three anti-kinetoplastid chemical boxes of ~200 compounds each were assembled. Functional analyses of these compounds suggest a wide array of potential modes of action against kinetoplastid kinases, proteases and cytochromes as well as potential host-pathogen targets. This is the first published parallel high throughput screening of a pharma compound collection against kinetoplastids. The compound sets are provided as an open resource for future lead discovery programs, and to address important research questions.

Half a billion people are at risk of contracting diseases caused by kinetoplastid parasites of the genera *Leishmania* and *Trypanosoma*<sup>1</sup>. This group of flagellated protozoans causes African trypanosomiasis (*T. brucei*), Chagas disease (*T. cruzi*) and various clinical manifestations of leishmaniasis (*Leishmania* spp.). In particular, *L. donovani* and *L. infantum* are responsible for life-threatening visceral leishmaniasis. An estimated 20 million individuals are infected with kinetoplastid pathogens, resulting in extensive suffering and around 95,000 deaths per year<sup>1</sup>. Despite this considerable disease burden, effective drug treatments for these parasites are lacking or inadequate and new therapies are required<sup>2,3</sup>.

Although these protozoans are transmitted by different insects and the human diseases they cause are clinically distinctive, much of their molecular and cellular biology is similar<sup>1</sup>. They are defined by the presence of a DNA-containing region, the 'kinetoplast', in their single large mitochondrion; have similar genomic organization and cellular structures (e.g. a single flagellum for motion, and glycosomes); and undergo morphological changes during their life cycles in the insect and vertebrate hosts. The genome of each parasite exceeds 8,000 genes, more than 6,000 being common orthologs<sup>4</sup>. Analysis of the genomes of *L. major*, *T. cruzi* and *T. brucei* (known as TriTryp) has revealed many common core metabolic functions as well as pathways that might reflect specific adaptations to environments within their insect and vertebrate hosts<sup>4</sup>. Strikingly, approximately 50% of these genomes encode for hypothetical proteins that do not resemble orthologs in the human genome<sup>1</sup>. In addition,



several human protein classes are not represented in kinetoplastid genomes, e.g. no orthologs have been found for tyrosine kinases<sup>5</sup>. These differences suggest that there may be essential proteins that can be exploited as selective targets for chemotherapy.

This paper reports the application of whole-cell phenotypic assays against *L. donovani*, *T. cruzi* and *T. brucei* to screen the GlaxoSmithKline HTS diversity set of 1.8 million compounds. This is the first parallel HTS program which has been disclosed for any pharma compound set against the three kinetoplastids most relevant to human disease. Three kinetoplastid chemical boxes have been assembled and all data are publically available to encourage research and drug discovery efforts in combating these devastating infections.

## Results

### High throughput screening (HTS) campaigns and hit identification.

The 1.8 million GlaxoSmithKline HTS screening collection was tested against *L. donovani*, *T. cruzi* and *T. brucei* between October 2012–May 2014 using a primary whole-cell phenotypic screen as described in the Methods. All compounds were tested at a final assay concentration of 5  $\mu$ M in the *L. donovani* and *T. cruzi* assays and at 4.2  $\mu$ M in the *T. brucei* assays. Primary hits were identified using algorithms developed in-house<sup>6</sup>. For each of the three phenotypic (primary) screens, one corresponding orthogonal assay was conducted to prove genuine activity and help to rule out false activity caused by assay interference. HTS campaigns are described in Supplementary Table 1, HTS results are summarized in Supplementary Figure 1 and the HTS progression cascade in Supplementary Figure 2.

*Leishmania donovani*. Growth inhibition of free-living *L. donovani* amastigotes in axenic cultures was determined in an assay adapted from de Ryker, *et al.*<sup>7</sup>. At an average statistical cut-off of >30%, 67,400 primary hits were identified; an overall hit rate of 4%. Using the same assay, confirmatory activity above the cut-off in at least one replicate was displayed for 32,200 compounds. These remaining compounds were tested in an intracellular assay of *L. donovani*-infected THP1-derived macrophages<sup>8</sup>. The cut-off obtained using the amastigotes per macrophage output was >34% and the corresponding hit confirmation rate was 31%, resulting in ca. 5,500 active compounds selective for *L. donovani* over macrophage cells. Using physicochemical parameters<sup>9</sup>, such as molecular weight >500 Da, calculated Property Forecast Index (cPFI) < 8, and <5 aromatic rings, the number of hits was reduced to 4,700. Compound potency (pIC<sub>50</sub>) was determined in a dose–response experiment and acute cytotoxicity of the compounds was assessed using the HepG2 assay (see Methods). Consequently, 351 non-cytotoxic anti-*L. donovani* compounds were identified.

*Trypanosoma cruzi*. Growth inhibition was determined using NIH-3T3 fibroblasts infected with a recombinant *T. cruzi* strain expressing beta-galactosidase as an intracellular reporter, as adapted from Bettiol, *et al.*<sup>10</sup>. The average statistical cutoff was >27% inhibition and the obtained overall hit rate was 7.7%, i.e. 130,678 active compounds. This list was reduced to 45,100 compounds by increasing the cut-off for biological response to >45%, excluding compounds with historical annotation in GSK databases for cytotoxicity *versus* the HepG2 cell line (pIC<sub>50</sub> > 6), and including only those compounds with sub- $\mu$ M IC<sub>50</sub> values, cPFI < 8, and aromatic rings < 4. Duplicate confirmation experiments were performed plus an interference assay against the host cell (i.e. NIH-3T3 fibroblasts). Based on >70% inhibition in the *T. cruzi* assay and <25% in the NIH-3T3 interference assay, 3,985 compounds were selected for dose–response and tested in parallel in the primary *T. cruzi* assay, the 3T3 host cell interference assay and for cytotoxicity against HepG2. A total of 2,310 compounds were identified with a *T. cruzi* pIC<sub>50</sub> > 5 and a selectivity index > 10. These compounds were tested in an intracellular imaging assay in H9c2 cells (rat cardiomyocytes)<sup>11</sup>. Also,

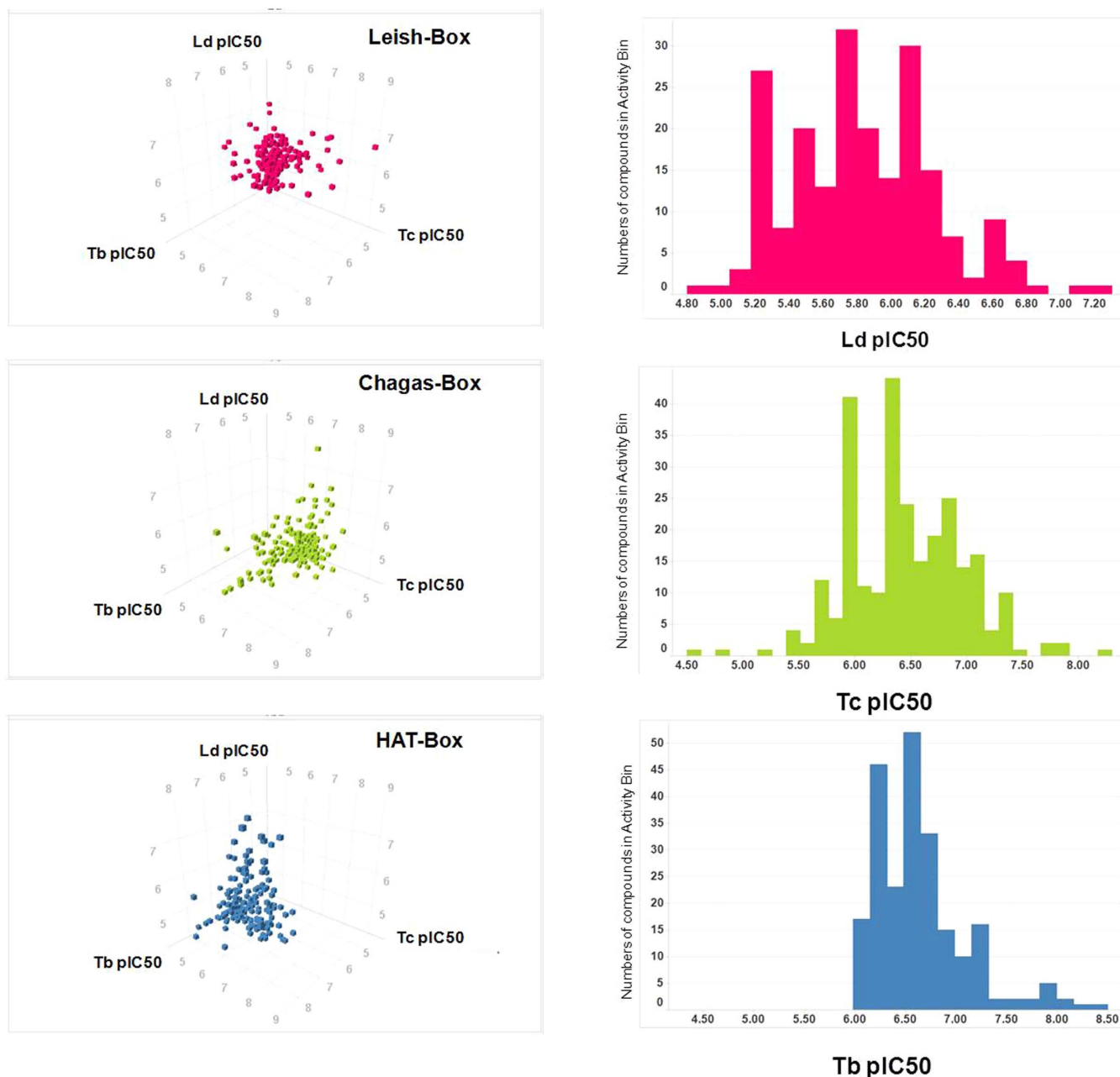
because sterol 14 $\alpha$ -demethylase (CYP51) inhibitors have known activity against *T. cruzi*<sup>12</sup>, hits were triaged based on *T. cruzi* CYP51 inhibition data obtained using a recently developed assay<sup>13</sup>. Compounds without CYP51 activity or a selectivity index > 10 for *T. cruzi* *versus* CYP51 were first selected. CYP51 inhibitors with a lower index were selected only if highly potent against *T. cruzi* (pIC<sub>50</sub> > 6). Overall, these investigations identified 500 non-cytotoxic anti-*T. cruzi* compounds.

*Trypanosoma brucei*. A resazurin fluorescent *T. brucei* whole-cell viability assay was used, derived from Sykes & Avery<sup>14</sup>. An average cut-off value of >40% resulted in 27,600 hits; a 1.5% overall hit rate. Confirmatory screening identified 15,200 compounds displaying a response above cut-off in at least one duplicate. A total of 4,200 compounds were selected for dose–response studies based on >80% growth inhibition in the confirmation step, cPFI < 8, aromatic rings < 5, molecular weight < 500 Da, and *in silico* determination of potential CNS penetration<sup>15</sup>. As well as the primary assay, these compounds were screened against *T. brucei* using an orthogonal ATP-based luminescence assay<sup>16</sup>, and acute cytotoxicity was evaluated against HepG2. Based on a pIC<sub>50</sub> > 6 and selectivity > 10, there were 700 non-cytotoxic anti-*T. brucei* compounds identified.

**Anti-kinetoplastid chemical boxes.** Selection of representative chemical boxes for the three kinetoplastids started from the most potent, specific, and non-cytotoxic compounds in the dose–response outputs of each screen after having filtered for lead-like properties as described in the Methods. In order to generate representative boxes with high chemical diversity and potency, compounds were clustered initially by similarity using a complete-linkage algorithm<sup>17</sup>, and a threshold of 0.55. Secondly, they were sorted by decreasing potency (i.e. pIC<sub>50</sub>). Compounds were then selected so that all the clusters were represented in the final ranked boxes by no more than two members with the highest potency. The final boxes contained 592 compound entries; 192 were active against *L. donovani* (Leish-Box), 222 against *T. cruzi* (Chagas-Box) and 192 against *T. brucei* (HAT-Box) (Figure 1, Supplementary Table 2 and available at <http://ebi.ac.uk/chemblntd>). The three anti-kinetoplastid chemical boxes showed little overlap, pointing to specific mechanisms of growth inhibition or structural divergence across molecular targets in each parasite. Three compounds were in both the Leish-Box and Chagas-Box, nine in both the Chagas-Box and HAT-Box and one compound was present in all three chemical boxes.

**Biological profiling and chemical characterization.** Database searches in ChEMBL showed that 88% of compounds in the three chemical boxes were not previously published. The literature annotation for the remaining 12% of compounds was unrelated to activity against *Leishmania* or *Trypanosoma*. Furthermore, these kinetoplastid chemical boxes do not contain analogs (Tanimoto coefficient > 0.7) to drugs currently used in the clinic for leishmaniasis, Chagas disease or sleeping sickness. Thus, they represent highly novel chemical diversity for potential starting points in drug discovery for kinetoplastid anti-parasitic agents. Only five compounds were natural products; a similar proportion to the screening collection.

There was no correlation between compound potency as determined by the axenic *versus* intracellular *L. donovani* assays (Supplementary Figure 3a). One explanation is that, in the intracellular assay, the compound must pass through two host cell membranes (i.e. plasma membrane and phagolysosome) plus the parasite cell membrane<sup>7</sup>, whereas in the axenic assay only the parasite membrane must be crossed. Also, there are significant pH changes from the extracellular media (acidic in the axenic assay, neutral in the intra-macrophage assay) to cytosol (neutral) to the parasitophorous vacuole (acidic pH 5.4). A reduced metabolism within the host cell



**Figure 1 | Distribution of potency by chemical box (*Leishmania donovani* in pink, *Trypanosoma brucei* in blue and *T. cruzi* in green) and parasite. Left panels: 3D scatter plot of pIC<sub>50</sub> values for each compound against *L. donovani* (Ld pIC<sub>50</sub>, as inhibition of amastigotes per macrophage readout from imaging assay), *T. cruzi* (Tc pIC<sub>50</sub>, as inhibition of amastigotes per cell readout from imaging assay) and *T. brucei* (Tb pIC<sub>50</sub>, as inhibition of FLINT from fluorescent assay). Right panels: Distribution of potency for all compounds of each of the three chemical boxes (~200 compounds distributed into 40 potency bins).**

may also impair drug activity against intra-macrophage *L. donovani* amastigotes, and these are slow-replicating, compared to axenic amastigotes<sup>7</sup>. In contrast, for *T. cruzi* and *T. brucei* there was good correlation for compound potency between the primary and orthogonal assays (Supplementary Figure 3b, 3c). This is as expected because the biological system for the assay differs only in the readout (i.e. fluorescent reporter *versus* imaging in *T. cruzi* and fluorescent-resazurin *versus* luminescent-ATP in *T. brucei*) or the host cell (i.e. NIH-3T3 fibroblasts *versus* H9c2 cardiomyocytes in *T. cruzi*).

As mentioned previously, the preclinical and clinical pipeline for Chagas disease is already well populated with CYP51 inhibitors<sup>12</sup>. In the Chagas-box, 42% of compounds represented a sub-set that was clearly differentiated from the overall compound population, exhibiting

a pIC<sub>50</sub> > 6 in both biochemical CYP51 and cellular anti-*T. cruzi* imaging assays (Supplementary Figure 4). These compounds may be flagged as putative CYP51-targeting hits. Conversely, 44% of compounds were CYP51-inactive; the remaining 14% poorly active.

The two imaging assays employed to determine compound activity against intracellular *L. donovani* and *T. cruzi* allowed simultaneous measurement of the number of amastigotes and infected cells as a dual assay readout. Interestingly, compounds in the Leish-Box and Chagas-Box showed good correlations between both anti-parasitic readouts, indicating parasite eradication (Supplementary Figure 5).

The screening methods and selection processes produced relatively small shifts in the distributions of clogP and molecular weight towards higher values, in many cases not significant (Supplementary





Figure 6a, b) (Student t-test for means, data not shown). For the number of aromatic rings, the distributions were centered at 3 rings; none of the compounds had >4 aromatic rings (Supplementary Figure 6c). As expected from the guided triage process, most compounds lie in appropriate lead-like space<sup>18</sup>.

### Discussion: modes of action hypotheses

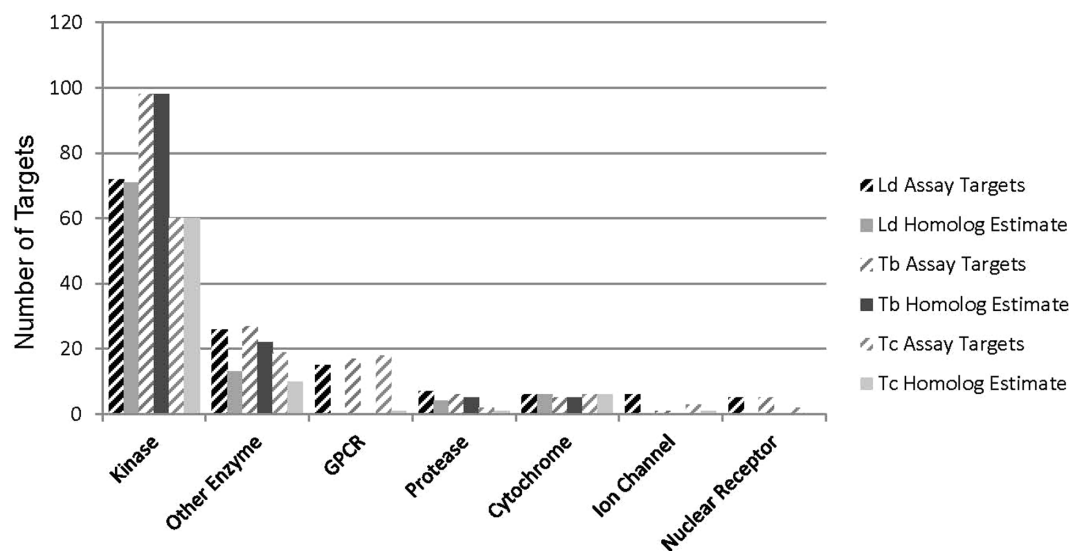
To assist target identification efforts, we used historical GSK screening data to generate modes of action hypotheses for the compounds in the three kinetoplastid chemical boxes. Based on conservative thresholds (i.e. overall  $pXC_{50} \geq 5.0$ , see Methods) for agonist or antagonist assays, compounds were associated with human protein targets which were used as BLASTP queries against the respective parasite genome<sup>19</sup>. The number of compounds meeting target specificity criteria was similar across *L. donovani* (80 compounds) and *T. brucei* (82 compounds) but lower for *T. cruzi* (46 compounds) (Figure 2).

'Kinase' was the most frequent target class (Table 1). A range of human kinase inhibitor chemotypes were present in the three boxes, including common hinge binding motifs such as pyrimidines, pyrrolopyrimidines, azaindoles, and benzimidazoles. Kinase inhibitors are a potentially important novel class of anti-kinetoplastid therapeutics because of high chemotype diversity in pharmaceutical compound collections and their potency in disrupting parasite survival<sup>20</sup>. Compared to the human genome which encodes over 450 kinases<sup>21</sup>, the kinomes of *Leishmania* and *Trypanosoma* sp. are less diverse, having <200 protein kinases<sup>5</sup>. Kinome-wide RNAi screens of *T. brucei* suggest that at least 43 kinases have essential roles in cell cycle and differentiation by *in vitro* culture analysis<sup>22</sup>, and others might prove to be essential only when tested *in vivo*. Several putative *T. brucei* kinase targets have had cell death or arrest phenotypes reported in kinome-wide RNAi knock-down experiments<sup>22</sup>. Although kinome-wide RNAi studies have not been published for *L. donovani* and *T. cruzi*, kinase inhibitors have activity against *L. donovani* *in vitro*<sup>23</sup>. The present compound set might be interacting with 34–40 serine/threonine kinases, depending upon the parasite.

A phylogenetic tree of the entire complement of human and *T. brucei* protein kinases was constructed (Figure 3), with additional annotations for those kinases with reported growth defects<sup>22</sup>. Several

*T. brucei* kinases with cell death or arrest phenotypes cluster with human kinases which also met our historical assay specificity criteria, including aurora kinases (AURK), CDK, CSNK1/2, MAP3K, PIM/CHEK, and PRKAA. Some kinases are potential drug targets, such as AURK which are known to be involved in cell cycle progression, exist as single copy genes in kinetoplastids<sup>24</sup> and have small molecule inhibitors with *in vitro* activity against *T. brucei*, and *L. major*<sup>25</sup>. The phylogenetic tree maps inhibitors of related human kinases against a single kinetoplastid isoform, illustrating the ancestral relationship of *T. brucei* kinases to multiple human homologues. These human–kinetoplastid kinase clusters provide starting points for further modes of action validation, with the caveat that many kinase inhibitors inhibit multiple, distantly related enzymes<sup>26</sup>. In addition to serine/threonine inhibitors kinases, several lipid kinases (phosphatidylinositol-kinases) were identified (Table 1). Lipid kinases also have a much lower complement in kinetoplastids compared to humans<sup>27</sup>, and could be important targets for either direct-acting anti-parasite drugs or disrupters of host–pathogen interactions<sup>28</sup>.

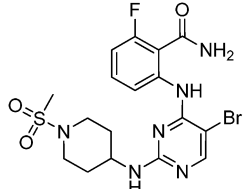
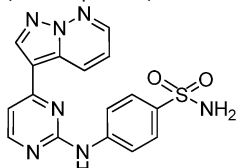
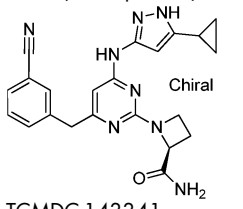
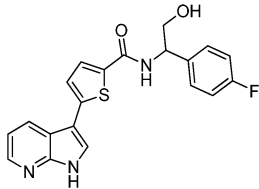
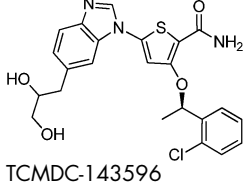
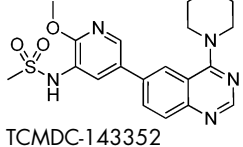
Potential modes of action hypotheses for the chemical box compounds were suggested against other kinetoplastid proteins, mainly phosphodiesterases (PDEs), cysteine peptidases and cytochromes (CYPs) (Table 2). All these proteins have been previously proposed as new kinetoplastid drug targets and some are under active investigation. Phosphodiesterase C is conserved across *Leishmania* sp. and *Trypanosoma* sp. where it is essential for disposal of cyclic nucleotides<sup>29</sup>. Several PDE inhibitors have been tested against kinetoplastids, in particular *T. cruzi*<sup>30</sup> and *T. brucei*<sup>31</sup>. Kinetoplastid cysteine peptidases are homologous to mammalian cathepsins B and L peptidases and perform essential roles. *T. brucei* requires active cysteine protease to deactivate the protective action of human apolipoprotein LI<sup>32,33</sup>. Kinetoplastid cytochromes are also a growing area of new drug development. CYP51 inhibition clears Chagas disease in mouse models<sup>34</sup>, and ferric reductase, which has cytochrome P450 sequence homology, is the primary mechanism of host iron uptake by *T. brucei*<sup>35</sup>. Several recent papers highlight the large diversity of structures that can bind to CYP51<sup>36–38</sup>, suggesting that a wide range of heterocyclic moieties can serve as a key pharmacophore binding to the cytochrome metal group. The large number of potential metal-binding pharmacophores in the kinetoplastid chemical boxes is striking.



**Figure 2** | Histogram of assignment of positive compounds based on historical assay data and homology of human protein targets to kinetoplastid genomes. Histograms are clustered by assays for *Leishmania donovani* (Ld), *Trypanosoma brucei* (Tb) and *T. cruzi* (Tc). The threshold above which compound efficacy against specific human targets was considered significant was defined as  $pIC_{50} \geq 5$  for inhibition or antagonist assays, and  $pEC_{50} \geq 5$  for agonist, activation or modulator assays (i.e. overall  $pXC_{50} \geq 5$ ). Kinetoplastid homologs to human proteins were assigned based on BLASTP E-values  $\leq 1.0e-05$ . Some compounds met the criteria for multiple human proteins (cross-hatched bars) while some kinetoplastid proteins had homology to multiple human proteins (solid bars).



**Table 1 |** Examples of proposed kinetoplastid targets for kinases based on homology to mammalian proteins used in historical assay with some examples compounds

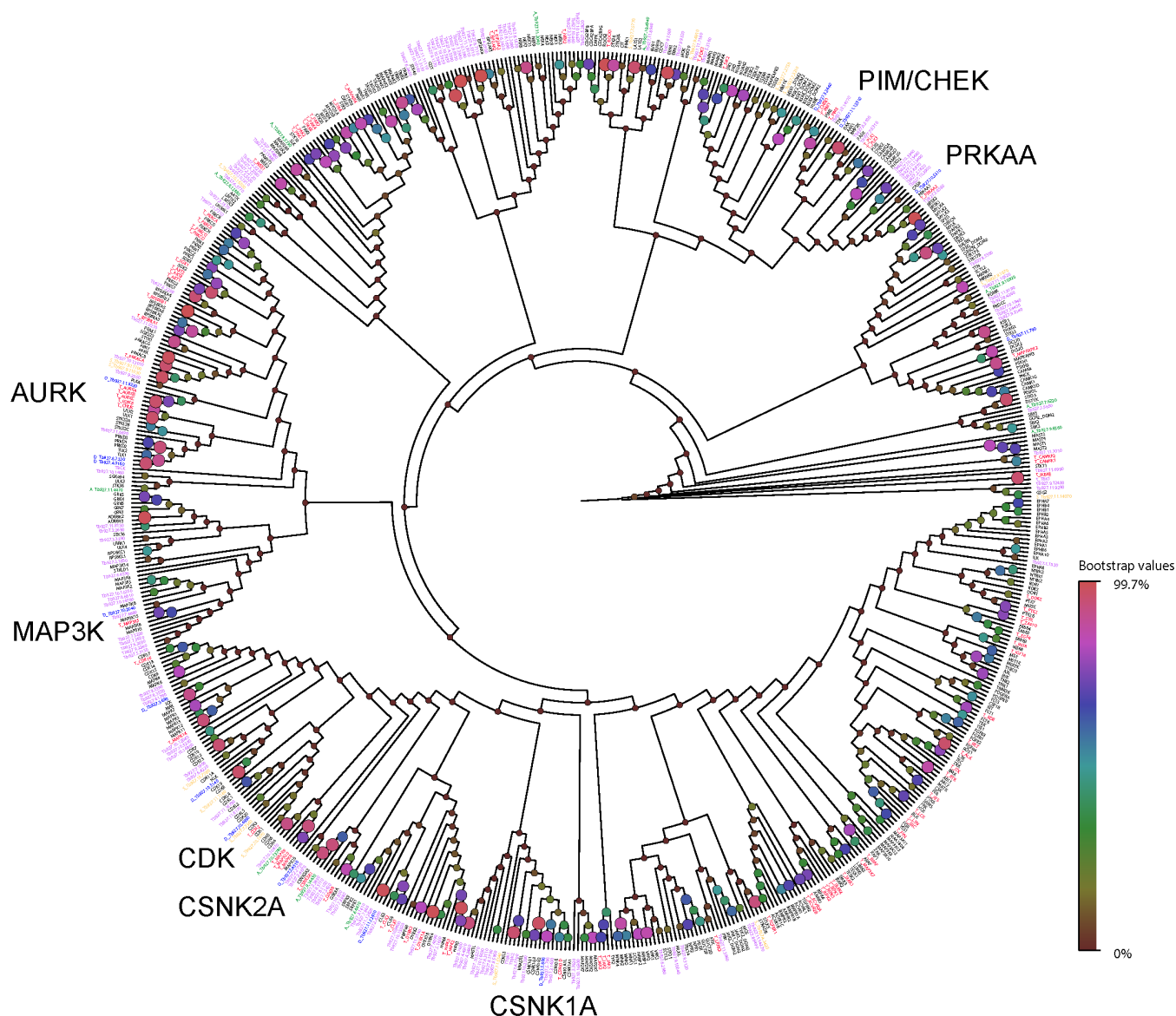
Target class	Potential kinetoplastid target <sup>a</sup>			No. of chemotypes/compounds	Example structures
	<i>L. donovani</i>	<i>T. brucei</i> (RNAi phenotype) <sup>b</sup>	<i>T. cruzi</i>		
Ser/Thr kinases	LDBPK_050390	Tb927.9.4910	Tc00.1047053398235.10	29/78	<p>2,4-diamino-pyrimidine series (16 compounds)</p>  <p>TCMDC-143365</p> <p>2-amino-4-aryl pyrimidine series (11 compounds)</p>  <p>TCMDC-143634</p> <p>Amino-heterocycle-pyrimidine series (7 compounds)</p>  <p>TCMDC-143341</p> <p>Azaindole series (5 compounds)</p>  <p>TCMDC-143234</p> <p>Thiophene benzimidazole series (4 compounds)</p>  <p>TCMDC-143596</p> <p>TCMDC-143352</p>
	LDBPK_071030	Tb927.9.11030	Tc00.1047053506007.40		
	LDBPK_081250	Tb927.9.14430	Tc00.1047053504181.40		
	LDBPK_100540	Tb927.10.460	Tc00.1047053504001.6		
	LDBPK_130280	Tb927.10.1070	Tc00.1047053508041.10		
	LDBPK_150180	Tb927.10.1910	Tc00.1047053503953.30		
	LDBPK_151630	Tb927.10.4990	Tc00.1047053508917.10		
	LDBPK_160310	Tb927.10.5140	Tc00.1047053510409.10		
	LDBPK_170440	Tb927.10.5310	Tc00.1047053511299.70		
	LDBPK_170770	Tb927.10.8420	Tc00.1047053510733.10		
	LDBPK_180270	Tb927.10.10350	Tc00.1047053511269.50		
	LDBPK_181090	Tb927.10.12040	Tc00.1047053511633.70		
	LDBPK_191480	Tb927.10.13010	Tc00.1047053506477.60		
	LDBPK_200780	Tb927.10.13490	Tc00.1047053510407.9		
	LDBPK_211320	Tb927.10.13780	Tc00.1047053509607.70		
	LDBPK_211870	Tb927.10.14420	Tc00.1047053509065.180		
	LDBPK_220770	Tb927.10.14780	Tc00.1047053509941.120		
	LDBPK_242410	Tb927.10.16030	Tc00.1047053508817.80		
	LDBPK_261710	Tb927.11.3140	Tc00.1047053508965.60		
	LDBPK_272420	Tb927.11.4470	Tc00.1047053503823.169		
	LDBPK_280550	Tb927.11.8220	Tc00.1047053510055.160		
	LDBPK_292140	Tb927.3.3190	Tc00.1047053509167.190		
	LDBPK_292830	Tb927.4.800	Tc00.1047053510759.40		
	LDBPK_300370	Tb927.4.2500	Tc00.1047053509007.60		
	LDBPK_300620	Tb927.4.3420 (S)	Tc00.1047053509937.130		
	LDBPK_303090	Tb927.4.3770	Tc00.1047053510329.210		
	LDBPK_320270	Tb927.5.790 (D)	Tc00.1047053511127.320		
	LDBPK_331930	Tb927.5.1650	Tc00.1047053508461.280		
	LDBPK_332420	Tb927.5.2820	Tc00.1047053511277.140		
	LDBPK_340030	Tb927.6.2250	Tc00.1047053508153.400		
LDBPK_343750	Tb927.7.3580	Tc00.1047053506825.180			
LDBPK_344160	Tb927.7.3650	Tc00.1047053507063.130			
LDBPK_360600	Tb927.7.6220 (A)	Tc00.1047053506679.80			
LDBPK_360970	Tb927.7.6310	Tc00.1047053508909.330			
	Tb927.8.870	Tc00.1047053507993.80			
	Tb927.8.3550	Tc00.1047053511283.280			
	Tb927.8.5730 (A)				
	Tb927.8.5950				
	Tb927.11.14070				
	Tb927.1.1000				
Phosphatidylinositol-kinases	LDBPK_291550	Tb927.3.4020	Tc00.1047053510003.30	2/5	<p>TCMDC-143352</p> 
	LDBPK_242090	Tb927.4.1140	Tc00.1047053510167.10		
	LDBPK_140020	Tb927.8.6210			
	LDBPK_044320				

<sup>a</sup>Gene loci identifiers given in descending order for *Leishmania donovani* (LDBPK), *Trypanosoma brucei* (Tb) and *T. cruzi* (Tc).

<sup>b</sup>RNAi phenotypes scored for Arrest (A), Slow (S) and Death (D) in *Trypanosoma brucei* according to Table 1 in Jones *et al.*<sup>22</sup>.

ing, i.e. 39 meta-substituted pyridyl compounds, 20 *para*-substituted pyridyl compounds, and 20 imidazole-containing compounds. There are also an unusually high number of methylene dioxy con-

taining compounds (28 of them). Although we could find no record of this moiety being a key feature of parasite CYP51 inhibitors, and indeed none of these were found to be *T. cruzi* CYP51 inhibitors, it



**Figure 3 | Neighbor-joining phylogenetic tree of combined human and *Trypanosoma brucei* str. 927 kinomes.** Human kinase targets with putative compounds are colored and prefixed (T\_red); all human kinases are colored black. *T. brucei* kinases are colored and prefixed for the RNAi phenotypes death (D\_blue), arrest (A\_dark green) or slow (S\_orange) according to Table 1 in Jones *et al.*<sup>22</sup>. All other *T. brucei* kinases are colored violet. A few key clusters of human target and essential *T. brucei* kinases are labeled. The size and color of circles on the nodes represent support in 1000 bootstrap replicates.

does interact with human cytochrome P450s. Although it is unlikely that all compounds with these potential metal interacting moieties are working through a parasite cytochrome, it is an area that should be explored in more depth.

Another functional group that appears at an unusually high rate (62 compounds) is the nitro-substituted aryl group, including nitro-pyrazoles, nitro-triazoles, nitro-furans, nitro-thiophenes, and nitro-benzenes. This is perhaps expected as nifurtimox and benznidazole are nitro-substituted aromatics used clinically for Chagas disease. These two drugs are substrates for a parasite type I nitroreductase and are converted into metabolites toxic to the parasite<sup>39</sup>. There is no mammalian homologue to this target, suggesting a potentially useful therapeutic window. There has been increased interest in nitro drugs to treat these parasitic diseases<sup>40</sup>, and the wide range of nitro-aromatic compounds present in these sets facilitates further substrate selectivity studies.

For those compounds putatively associated with protein targets, we developed a preliminary systems biology network view of tar-

geted pathways based on all known KEGG pathways for the kinetoplastid species<sup>41,42</sup>. The analysis included compounds with both human and kinetoplastid putative protein targets, i.e. 31, 48 and 62 compounds were analyzed for *T. cruzi*, *T. brucei* and *L. donovani*, respectively. These compounds are associated to 33 proteins in these parasites. Description of these proteins was taken from the NCBI Gene database (<http://www.ncbi.nlm.nih.gov/gene>). Pathway interactions were mapped for the individual kinetoplastid species and as a combined three species map (Figure 4).

Some additional predictions of homologous protein targets were included in this map, for which there were fewer compounds. A key caveat for these analyses is that the kinetoplastid KEGG reference pathways are incomplete<sup>41</sup>. For example, there are only 91 known pathways for *T. brucei*, versus 288 for humans. Thus, not all kinetoplastid proteins were associated with pathways. However, the total network still included 42 compounds for *T. brucei* (68% of all compounds analyzed), 16 (52%) for *T. cruzi*, and 28 (58%) for *L. donovani*. Other compounds target proteins with no known pathways.



**Table 2 |** Some examples of proposed kinetoplastid targets for non-kinases based on homology to mammalian proteins used in historical assay with some examples compounds

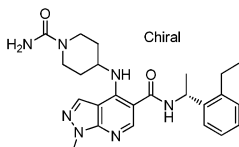
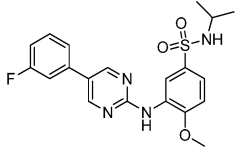
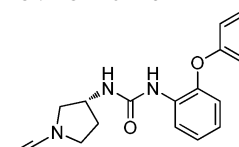
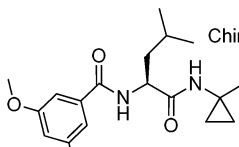
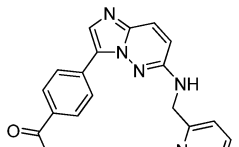
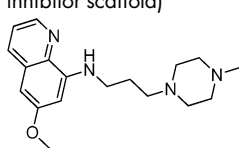
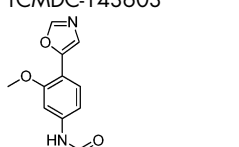
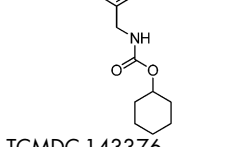
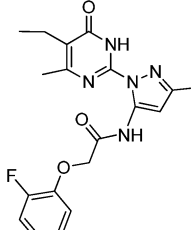
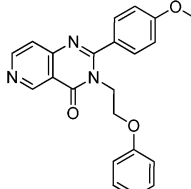
Target class	Potential kinetoplastid target		No. of chemotypes/ compounds	Example structures
	Protein	Gene loci <sup>a</sup>		
Other enzyme	Phosphodiesterase	LDBPK_181100 Tc00.1047053511269.40 Tc00.1047053509805.20 Tb10.389.0510	3/4	 <p>Chiral</p> <p>TCMDC-143547</p>  <p>TCMDC-143248</p>
Proteases	Cysteine peptidases A (CBA) or C (CPC)	LDBPK_290860 (CPC) LDBPK_191460 (CBA) Tb927.6.560 (CPC) Tb927.6.960 (precursor) Tb927.6.1020 (precursor) Tc00.1047053510535. 100 (CPC)	3/6	 <p>TCMDC-143238</p>  <p>Chiral</p> <p>TCMDC-143390</p>
Cytochromes	Ferric reductase (cytochrome homology)	LDBPK_301630 Tb11.02.1990 Tc00.1047053507665.30	5/5	 <p>TCMDC-143560 (also a kinase inhibitor scaffold)</p>  <p>TCMDC-143603</p>
	Cytochrome P450 and lanosterol 14-alpha-demethylase	LDBPK_270090 LDBPK_343110 Tb03.27F10.920 Tb11.02.4080 Tc00.1047053509231.10 Tc00.1047053509719.40 Tc00.1047053510101.50	24/44	 <p>TCMDC-143376</p>  <p>TCMDC-143433</p>





Table 2 | Continued

Target class	Potential kinetoplastid target		No. of chemotypes/ compounds	Example structures
	Protein	Gene loci <sup>a</sup>		
				 TCMDC-142497
				 TCMDC-143182

<sup>a</sup>Gene loci identifiers given in descending order for *Leishmania donovani* (LDBPK), *Trypanosoma brucei* (Tb) and *T. cruzi* (Tc).

Although there were some species-specific pathways for *T. brucei* (n = 4), *T. cruzi* (n = 3) and *L. donovani* (n = 1), the majority of hypothesized targets (and compounds) could be assigned pathways common to all three parasites, including ‘biosynthesis of secondary metabolites’, ‘metabolic pathways’, ‘phosphatidylinositol signaling system (and metabolism)’, ‘regulation of autophagy’, ‘nicotinate and nicotinamide metabolism’ and ‘steroid biosynthesis’. Interestingly, there were some proteins, e.g. silent information regulator 2 (XP 816094, XP 845793 and CBZ34909 for *T. cruzi*, *T. brucei* and *L. donovani*, respectively), with the same biological description and pathway in all three species. The compounds targeting these proteins may, therefore, disrupt this pathway in the same manner.

There were also several compounds associated with human proteins with no known homologs in kinetoplastids, such as G-coupled protein receptors (GPCRs), nuclear receptors, ion channels and transporters. These compounds potentially act on evolutionarily unrelated targets or essential proteins that, although structurally and functionally similar to the human targets, have no significant primary amino acid homology. Importantly, two assays in this study used parasites cultured in human host cells (*L. donovani* in macrophages and *T. cruzi* infecting fibroblasts). Thus, modulation of host–pathogen interactions could be an indirect mode of action for some compounds (with either homologous or non-homologous human targets). For example, *L. donovani* potentially modulates the cAMP signaling pathways mediated by the GPCR EP2 or PTGER2<sup>43</sup>. Disruptors of host–parasite interactions could lead to novel anti-kinetoplastid therapeutic strategies, less susceptible to the selection pressures which drive drug resistance in the parasite.

This paper reports for the first time the outcome from a parallel high throughput screening of a pharma compound collection (1.8 million compounds) against the three most relevant *Leishmania* and *Trypanosoma* parasites causative of human disease. Three corresponding kinetoplastid chemical boxes were identified, encompassing a wide variety of potential targets, such as kinases, proteases, and cytochromes as well as potential host–pathogen interaction targets. The outcomes of this research are publically disclosed in full, and physical samples of the three kinetoplastid chemical boxes are available to collaborators upon request. We expect this information will enable the scientific community to address relevant research questions

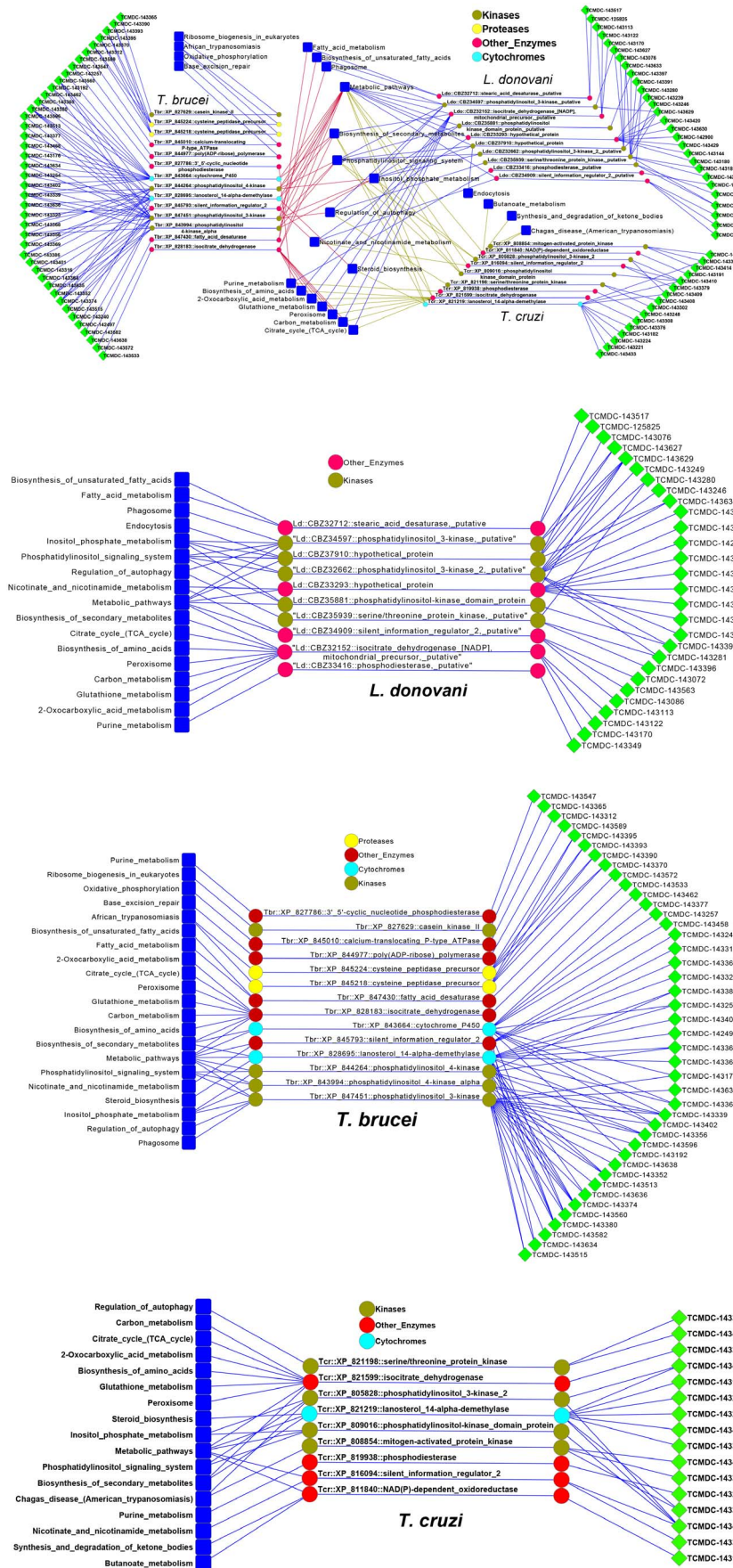
and seed lead discovery programs that eventually deliver innovative treatments for these important, but neglected diseases.

## Methods

**Parasites and mammalian cell cultures.** LLC-MK2 cells (green monkey kidney epithelial cells) were purchased from the European Cell Cultures Collection (ECACC reference 85062804) and used to expand the *T. cruzi* parasite population. NIH-3T3 cells (mouse endothelial fibroblasts) were made available by GSK-Biological Reagents and Assay Development Department (BRAD, Stevenage, UK). H9c2 (rat cardiomyocytes) were purchased at the European Cell Cultures Collection (ECACC, Salisbury, UK). *T. cruzi* parasites from the Tulahuén strain expressing β-galactosidase were kindly provided by Dr. Buckner (University of Washington, Seattle, USA; Buckner et al., 1996). Parasites were maintained in culture by weekly infection of LLC-MK2 cells. Trypomastigote forms were obtained from the supernatants of LLC-MK2 infected cultures harvested between days 5 and 9 of infection, as described elsewhere<sup>10</sup>. THP-1 cells (human monocytic leukemia) were made available by GSK-Biological Reagents and Assay Development Department (BRAD, Stevenage, UK). LdBOB axenic amastigotes expressing eGFP were kindly provided by Manu de Rycker from Dundee University. *Trypanosoma brucei brucei* bloodstream form Lister 427, was provided by Miguel Navarro from Instituto Lopez Neyra (Consejo Superior de Investigaciones Científicas, Granada, Spain). HepG2 (human hepatoma) was provided by GSK-Biological Reagents and Assay Development Department (BRAD, Stevenage, UK).

***Leishmania donovani*.** Primary growth inhibition assay: Assay plates (1536-well) were prepared by adding 30 nL per well of compound. For single shot assay, the final compound concentration was 5 μM. For potency determinations, eleven-point one in three dilution curves were generated with a top concentration of 50 μM. The assay was adapted from de Rycker et al.<sup>7</sup>. Briefly, eGFP LdBOB axenic amastigotes were harvested, fixed and counted in a CASY cell counter (Roche-Applied Science). A final concentration of 1.5 × 10<sup>5</sup> cells per well were prepared in amastigote media containing 500 U/mL penicillin/streptomycin (Invitrogen™) and 6 μL was dispensed to each experimental well using a Multidrop Combi dispenser (Thermo Scientific). 6 μL of assay media was dispensed to control columns used as reference for 100% compound response. Following incubation for 72 h at 37°C, 5% CO<sub>2</sub>, 2 μL of resazurin was added to each well at a final concentration of 0.5 mM in phosphate buffered saline (PBS) with IGEPAL (Sigma-Aldrich), 0.05% (v/v). The plates were incubated for 4 h at room temperature and resorufin fluorescence was detected using a PerkinElmer EnVision plate reader with excitation at 528 nm and emission at 590 nm.

**Intra-macrophage *L. donovani* assay:** Assay plates (384-well) were prepared by adding 250 nL of compound to each well. For single shot assay, the final compound concentration was 5 μM. For potency determinations, eleven-point one in three dilution curves were generated with a top concentration of 50 μM. The assay was adapted from de Rycker et al.<sup>7</sup>. Briefly, THP-1 cells were grown previously in CELLMASTER roller bottles (Greiner cat. # 680048) at a concentration of 9 × 10<sup>5</sup> cells/mL for 72 h. The cells were monitored using an optical microscope and counted using a CASY cell counter. Then they were differentiated in 225 cm<sup>2</sup> T-



**Figure 4 |** Pathway network analysis of combined and individual kinetoplast target hypotheses. Pathways were assigned using KEGG and visualized using Cytoscape<sup>42</sup>. Proteins of three species are grouped and colored according to general functional groups (kinases, proteases, other enzymes and cytochromes).



FLASK in the presence of 30 nM of PMA (phorbol 12-myristate 13-acetate) (Sigma-Aldrich) at a final concentration of  $6 \times 10^5$  cells/mL. Following incubation for 24 h at 37°C, 5% CO<sub>2</sub>, cell differentiation were visually confirmed using an optical microscope. Cells were washed twice using cell culture media. Cells were infected using eGFP LdBOB axenic amastigotes at a multiplicity of infection of 10 (i.e.  $6 \times 10^6$  parasites/mL) in the T-FLASK containing differentiated THP-1 cells. Each T-FLASK was incubated overnight and the remaining parasite was removed washing three times with sterile DPBS. The infected cells were harvested by treatment with 0.05% (w/v) trypsin plus 0.48 mM EDTA for 5 min and an aliquot was fixed and counted in a CASY cell counter. A cell preparation with a final concentration of  $1.6 \times 10^5$  cells/mL was prepared in assay media consisting of RPMI (Invitrogen™), 2% FBS (Gibco) and 25 mM sodium bicarbonate (Invitrogen™) containing 30 nM of PMA. Infected cells were plated onto assay plates containing compounds (3000 cells/well, 50 µL) using a Multidrop Combi dispenser. Amphotericin B (SigmaAldrich) was used as positive control of 100% compound response and was prepared by diluting the compound in an aliquot of infected cells for a final concentration of 2 µM. Treated cells were dispensed into the control column using a Multidrop Combi dispenser. Plates were incubated for 96 h at 37°C, 5% CO<sub>2</sub> and then fixed with 4% (v/v) formaldehyde-PBS for 30 min at room temperature. After fixation, the wells were washed twice with 100 µL PBS using an EL406 multi well plate washer (BioTek), stained with 10 µg/mL DAPI in PBS plus 0.1% (v/v) Triton X-100 for 30 min at room temperature, and washed twice with 50 µL PBS. Finally, 50 µL of PBS was added to each well and the plates were sealed and read on a high-content microscope (Opera QEHS) using a 20x objective, 3 fields per well. Two exposure images were taken for each well using 405 nm and 488 nm laser excitation. Automated image analysis was performed with a script developed on Acapella® High Content Imaging and Analysis Software (PerkinElmer). THP-1 cell count, average number of amastigotes per macrophage and percent of infected cells were reported for each well.

*Trypanosoma cruzi*. Primary growth inhibition assay: The assay developed was adapted from the one previously described by Bettio, *et al.*<sup>10</sup>. Active compounds in this assay format will target intracellular *T. cruzi* amastigotes growing in NIH-3T3 murine fibroblasts, though to a smaller extent they may also target free swimming trypomastigotes and/or the host-parasite interactions required for the parasite invasion. The assay was performed in tissue culture surface treated 1536 well plates. This assay was run at single shot at 5 µM compound concentration per well. The biological reagents consisted of NIH-3T3 host cells and trypomastigotes parasitic cells; both cell types were set to a  $3.3 \times 10^5$  cells/mL dilution in assay DMEM by means of a CASY cell counter device using a 60 µm capillary. 6 µL of the mixture were dispensed per well (about  $1 \times 10^3$  host-cells and  $1 \times 10^3$  parasites) with a Multidrop Combi dispenser. Previously, a solution of  $1.67 \times 10^5$  trypomastigotes per mL in assay DMEM had been prepared and 6 µL per well dispensed in the allotted control columns that defined 100% inhibition. The 0% inhibition of parasitic growth control was obtained by leaving the assay mixture untreated in its corresponding plate columns. In all wells, the percentage of DMSO never exceeded 0.5%. Plates incubated for four days at 37°C, 5% CO<sub>2</sub>. The substrate used for the assay fluorescence intensity (FLINT) readout was resorufin-β-D-galactopyranoside (Sigma-Aldrich) at 5 µM per well. It was diluted in PBS supplemented with the soft detergent Igepal (Fluka) as described elsewhere<sup>44</sup>. Upon addition to the plates, the substrate solution was incubated for 4 h at room temperature and the signal read with the EnVision microtiter plate reader (Perkin-Elmer) using the corresponding set of filters (excitation/emission at 531/595 nm).

*Trypanosoma cruzi* intracellular imaging assay: This intracellular assay in H9c2 (rat cardiomyocyte) cells was as per Alonso-Padilla, *et al.*<sup>11</sup>. Briefly, H9c2 cells were seeded in T-225 flasks (225 cm<sup>2</sup> culture surface; Corning Inc., NY, USA) in DMEM-10% FBS for 4 h to allow attachment. Cells were then washed once with PBS before infection. *T. cruzi* trypomastigotes, collected at days 5 to 8 after infection from LLC-MK2 parasite infected cultures, were allowed to swim out for 4 h at 37°C from a centrifuged pellet (2,500 rpm/10 min/room temperature). Trypomastigotes were then collected and counted in a CASY Cell Counter. Trypomastigotes, in supplemented DMEM, were added to H9c2 cultures in a multiplicity of infection of 1 and incubated for 18 h. Cells were washed once with PBS before incubation of the infected H9c2 monolayer with trypsin (Life-Technologies) to detach cells from the flask. Cells were counted in a CASY Cell Counter using a 150 µm capillary and their density set at  $5 \times 10^4$  cells per mL in supplemented assay DMEM. Infected H9c2 were dispensed into 384-well poly-lysine coated assay plates at 50 µL per well using a Multidrop Combi dispenser. After seeding them, the plates were incubated at 37°C, 5% CO<sub>2</sub> for 72 h. Cultures were then fixed and stained by addition of 50 µL of a solution containing 8% formaldehyde and 4 µM DRAQ5 DNA dye (BioStatus, UK) per well. Plates were kept light-protected and imaged one hour later in a Perkin-Elmer Opera microscope using a 20x air objective (NA 0.4) and the following acquisition set: a 635 nm laser excitation line and a 690/50 emission detection filter for DRAQ5 detection. Five images were collected per well for reliable statistical analysis. Automated image analysis was performed with a script developed on Acapella® High Content Imaging and Analysis Software (PerkinElmer). Three outputs were provided for each sample well: (1) number of host cells nuclei to determine drug-related cytotoxicity; (2) number of amastigotes per cell as infection level measurement; and (3) percentage of infected cells per well as a second infection marker.

*Trypanosoma brucei*. Primary growth inhibition assay: This resazurin (Alamar Blue) fluorescent whole cell viability assay was based on that of Sykes & Avery<sup>14</sup>. Assay plates (1536 well) were prepared by dispensing 25 nL of compound solution per well

using an Echo acoustic dispenser (Labcyte). For single shot assays the final compound concentration was 4.2 µM. For potency determinations, eleven-point one in three dilution curves were generated and the top concentration was 42 µM. For control samples, DMSO was pre-dispensed. *T. brucei* Lister strain 427 was cultured, harvested and counted in a CASY cell counter. A parasite suspension was prepared at a final concentration of  $5 \times 10^4$  cells/mL in complete HMI9 medium and 6 µL dispensed to all wells containing compounds using a Multidrop Combi dispenser. 6 µL of HMI9 media supplemented with 0.4% DMSO was dispensed to C2 wells. After 48 h of incubation at 37°C, 5% CO<sub>2</sub>, plates were tempered for 10 min at room temperature before addition of 2 µL of resazurin to each well at a final concentration of 50 µM in PBS. Plates were incubated for 2 h at room temperature and resorufin fluorescence was determined using a PerkinElmer ViewLux plate reader (excitation at 528 nm and emission at 590 nm).

*Trypanosoma brucei* growth inhibition luminescent assay: An ATP-based luminescence assay was run as confirmatory probe for the selected compounds in dose-response format, as per Sykes & Avery<sup>16</sup>. Pre-dispensed plates (1536 well) were prepared as described above for the fluorescence assay. *T. brucei* cultures were harvested and counted in a CASY cell counter. A parasite suspension was prepared at a final concentration of  $5 \times 10^4$  cells/mL in complete HMI9 medium and 6 µL was dispensed to all wells containing compounds using a Multidrop Combi dispenser. 6 µL of HMI9 media with 0.4% DMSO were dispensed to control columns. Plates were incubated for 24 h at 37°C, 5% CO<sub>2</sub> then tempered for 10 min at room temperature. ATP content was quantified by adding 4 µL of CellTiter-Glo® Reagent (Promega) according to the manufacturer's instructions. Luminescence was measured using a ViewLux.

**HepG2 cytotoxicity assay.** Actively growing HepG2 cells were removed from a T-175 TC flask using 5 mL Eagle's MEM (containing 10% FBS, 1% NEAA, 1% penicillin/streptomycin) and dispersed in the medium by repeated pipetting. Seeding density was checked to ensure that new mono-layers were not more than 50% confluent at the time of harvesting. Cell suspension was added to 500 mL of the same medium at a final density of  $1.2 \times 10^5$  cells/mL. This cell suspension was dispensed (25 µL, 3000 cells per well) into 384-well clear-bottom plates using a Multidrop Combi dispenser. Prior to addition of the cell suspension, the screening compounds (250 nL) were pre-dispensed into the plates with an Echo® liquid handler. Plates were incubated for 48 h at 37°C, 5% CO<sub>2</sub>. After incubation, plates equilibrated at room temperature for 30 min before proceeding to develop the luminescent signal. The signal developer, CellTiter-Glo® Reagent, was allowed to equilibrate at room temperature for 30 min and added to the plates (25 µL per well) using a Multidrop Combi dispenser. Plates were left for 10 min at room temperature for stabilization and then read using a ViewLux.

**NIH-3T3 cytotoxicity.** Since *T. cruzi* is an obligatory intracellular parasite, those compounds that affect the host-cells viability would also be reported as anti-parasite hits by the previously described phenotypic assay. Therefore, a complementary assay was developed to determine the hits for NIH-3T3 cytotoxicity and/or their anti-*T. cruzi* specificity. Tissue culture surface treated assay plates (1536 well) were used. 5 µL of a  $2 \times 10^5$  cells per mL solution were dispensed per well (~10<sup>3</sup> host cells), into plates already containing compounds which had been pre-dispensed. The 100% cells survival control was determined exposing the host-cells to a 0.5% DMSO concentration, whereas the 100% toxicity control was achieved by seeding the cells in presence of Amphotericin B at 50 µM. Plates were incubated for four days at 37°C, 5% CO<sub>2</sub>. The assay readout was developed by addition of 5 µL of CellTiter-Glo® Reagent following the manufacturer's instructions. Luminescence was measured using a ViewLux.

**Data analysis.** Data were normalized to percent of biological response by using positive (i.e. highest response achieved by a chemical tool compound, R<sub>Ctr12</sub>) or negative (i.e. lowest response achieved in the absence of any testing compound, R<sub>Ctr11</sub>) controls by using the following equation (Equation 1):

$$\%Response = \frac{|R_{Ctrl1} - R_x|}{||R_{Ctrl1} - R_{Ctrl2}||} \cdot 100$$

where R<sub>x</sub> is the assay response measured for the compound X. R<sub>Ctr11</sub> and R<sub>Ctr12</sub> are calculated as the average of replicates in the same microtiter plate where the compound X is tested.

Assay performance statistics, such as signal to background ratio, Z' and robust 3SD activity cutoff were calculated using templates in ActivityBase XE (IDBS, Guilford, Surrey, UK). Hit population analysis and visualization were conducted using Spotfire DecisionSite (Spotfire, Inc., Somerville, Massachusetts). pIC<sub>50</sub> values were obtained using the ActivityBase XE nonlinear regression function in the full curve analysis bundle.

**Selection criteria for the anti-kinetoplastid chemical boxes.** In the case of *L. donovani*, this corresponded to a pIC<sub>50</sub> > 5 in the amastigote per macrophage ratio (AM/MAC) or in the percentage of infected cells (INF), a ratio of IC<sub>50AM/MAC</sub>/IC<sub>50MAC</sub> > 5, and a ratio of IC<sub>50AM/MAC</sub>/IC<sub>50HepG2</sub> > 10. In the case of *T. cruzi*, this comprised a pIC<sub>50</sub> > 5 in the *T. cruzi* primary assay, a pIC<sub>50</sub> < 5 in the HepG2 assay, a pIC<sub>50</sub> < 5 in the NIH-3T3 host cell assay, a ratio of IC<sub>50TC</sub>/IC<sub>50HepG2</sub> > 10, and a ratio of IC<sub>50TC</sub>/IC<sub>50T3</sub> > 10. In the case of *T. brucei* this corresponded to a pIC<sub>50</sub> > 6 in the fluorescent (primary) and luminescent (secondary) *T. brucei* whole cell assays and





selectivity indexes  $> 10$  for HepG2. In addition, all the dose–response curves were inspected and compounds were disregarded in case of aberrant fit or bizarre shape (i.e. low slope or low maximum asymptote).

**Bioinformatics analysis.** In order to develop modes of action hypotheses, GSK proprietary compound screening databases were queried for historical assay data associated with active compounds from each of the three kinetoplastid species screens. The majority of these screens were against human protein targets. Thresholds for significant compound efficacy against specific human targets were defined as  $pIC_{50} \geq 5$  for inhibition or antagonist assays, and  $pEC_{50} \geq 5$  for agonist, activation or modulator assays (i.e. overall  $pXC_{50} \geq 5$ ). Human target proteins were binned by screen according to their compound hit against *L. donovani* (139 human proteins), *T. brucei* (161 human proteins) and *T. cruzi* (110 human proteins). The human target protein sequences were used as BLAST<sup>19</sup> queries to search for homologs in the respective translated reference genomes for *L. donovani* strain BPK282A1<sup>45</sup>, *T. brucei* strain 927/4 GUTat10.1<sup>46</sup> and *T. cruzi* strain CL Brener<sup>47</sup>, with a cut-off E-value  $\leq 1.0e-05$  then subsequently confirmed by reciprocal searches against the human genome. A combined kinome phylogenetic tree was reconstructed using all annotated human kinases as well as *T. cruzi* strain 927 kinases annotated by Interpro kinase queries of TriTrypDB<sup>4</sup> or reported by Urbaniak *et al.*<sup>20</sup>. RNAi kinase phenotypes in *T. brucei* strain 927 were annotated as death, slow or arrest according to Table 1 in Jones *et al.*<sup>22</sup>. Initial multiple sequence alignments were performed using the program CLUSTALW v1.7<sup>48</sup> with default settings and subsequently, refined manually using the program JALVIEW<sup>49</sup>. Residues that could not be unambiguously aligned or that contained insertions or deletions were removed from the multiple sequence alignment prior to phylogenetic analysis. The final kinase multiple sequence alignment was 114 amino acids in length and included a combined total of 657 human and *T. brucei* kinases. We reconstructed phylogenetic trees using distance neighbor-joining (NJ) using the JTT option as implemented in the software MEGA6<sup>50</sup>. All trees were visualized using the program FigTree v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree>). Network analysis was done using KEGG reference pathways<sup>41</sup> for individual parasites and general description of all analyzed genes were taken from the NCBI Gene database (<http://www.ncbi.nlm.nih.gov/gene>). Cytoscape was used for visualization of the compound-protein-pathway networks<sup>42</sup>.

**Biosafety.** Experimental work with live *L. donovani* or *T. cruzi* cells were carried out following standard operating procedures in compliance with biosafety level 3 regulations (BSL3). Work with *T. brucei* was developed under biosafety level 2 procedures. HepG2 and THP-1 cells were treated according to GSK policies for management of human biological samples

- Stuart, K. *et al.* Kinetoplastids: related protozoan pathogens, different diseases. *J. Clin. Inv.* **118**, 1301–1310 (2008).
- Pedrique, B. *et al.* The drug and vaccine landscape for neglected diseases (2000–2011): a systematic assessment. *Lancet Glob. Health* **1**, e371–e379 (2013).
- World Health Organization. Working to overcome the global impact of neglected tropical diseases – first WHO report on neglected tropical diseases. [http://whqlibdoc.who.int/publications/2010/9789241564090\\_eng.pdf](http://whqlibdoc.who.int/publications/2010/9789241564090_eng.pdf) (2010) (Date of access: 20<sup>th</sup> November 2014).
- Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457–462 (2010).
- Parsons, M., Worthey, E. A., Ward, P. N. & Mottram, J. C. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **6**, 127 (2005).
- Coma, I., Herranz, J. & Martin, J. Statistics and decision making in high-throughput screening. In *High Throughput Screening Methods and Protocols, Second Edition* (Humana Press, New York), 121–124 (2005).
- De Rycker, M. *et al.* Comparison of a high-throughput high-content intracellular *Leishmania donovani* assay with an axenic amastigote assay. *Antimicrob. Agents Chemother.* **57**, 2913–2922 (2013).
- Aulner, N. *et al.* High content analysis of primary macrophages hosting proliferating leishmania amastigotes: application to anti-leishmanial drug discovery. *PLoS Negl. Trop. Dis.* **7**, e2154 (2013).
- Young, Y., Green, D., Luscombe, C. & Hill, A. Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discovery Today* **16**, 822–830 (2011).
- Bettiol, E. *et al.* Identification of three classes of heteroaromatic compounds with activity against intracellular *Trypanosoma cruzi* by chemical library screening. *PLoS Negl. Trop. Dis.* **3**, e384 (2009).
- Alonso-Padilla, J. *et al.* Automated image-based screening assay for *Trypanosoma cruzi* infection of mammalian cells. *PLoS Negl. Trop. Dis.* **9**, e0003493 (2015).
- Lepesheva, G. I., Villalta, F. & Waterman, M. R. Targeting *Trypanosoma cruzi* sterol 14 $\alpha$ -demethylase (CYP51). *Adv Parasitol* **75**, 65–87 (2011).
- Riley, J. *et al.* (2014) Development of a fluorescence-based *Trypanosoma cruzi* CYP51 inhibition assay for effective compound triaging in drug discovery programmes for Chagas disease. *PLoS Negl. Trop. Dis.* manuscript submitted for publication (2014).
- Sykes, M. L. & Avery, V. M. Development of an Alamar Blue viability assay in 384-well format for high throughput whole cell screening of *Trypanosoma brucei* bloodstream form strain 427. *Am. J. Trop. Med. Hyg.* **81**, 665–674 (2009).
- Wager, T. T., Hou, X., Verhoest, P. R. & Villalobos, A. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* **1**, 435–449 (2010).
- Sykes, M. L. & Avery, V. M. A luciferase based viability assay for ATP detection in 384-well format for high throughput whole cell screening of *Trypanosoma brucei* bloodstream form strain 427. *Parasit. Vectors* **12**, 54 (2009).
- Leach, A. R. & Gillet, V. J. *An Introduction to Chemoinformatics* (Springer, New York), 121–124 (2007).
- Hann, M. M. & Oprea, T. I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr. Opin. Chem. Biol.* **8**, 255–263 (2014).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Urbaniak, M. D. *et al.* Chemical proteomic analysis reveals the drugability of the kinome of *Trypanosoma brucei*. *ACS Chem. Biol.* **7**, 1858–1865 (2012).
- Manning, G. *et al.* The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Jones, N. G. *et al.* Regulators of *Trypanosoma brucei* cell cycle progression and differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog.* **10**, e1003886 (2014).
- Sanderson, L., Yardley, V. & Croft, S. L. Activity of anti-cancer protein kinase inhibitors against *Leishmania* spp. *J. Antimicrob. Chemother.* **69**, 1888–1891 (2014).
- Brown, J. R. *et al.* Evolutionary relationships of Aurora kinases: implications for model organism studies and the development of anti-cancer drugs. *BMC Evol. Biol.* **4**, 39 (2004).
- Patel, G. *et al.* Repurposing human Aurora kinase inhibitors as leads for anti-protozoan drug discovery. *Medchemcomm.* **5**, 655–658 (2014).
- Karaman, M. W. *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **26**, 127–132 (2008).
- Brown, J. R. & Auger, K. R. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. *BMC Evol. Biol.* **11**, 4 (2011).
- Cummings, H. E. *et al.* Critical role for phosphoinositide 3-kinase gamma in parasite invasion and disease progression of cutaneous leishmaniasis. *Proc. Natl. Acad. Sci. USA* **109**, 1251–1256 (2012).
- Seebeck, T., Sterk, G. J. & Ke, H. Phosphodiesterase inhibitors as a new generation of antiprotozoan drugs: exploiting the benefit of enzymes that are highly conserved between host and parasite. *Future Med. Chem.* **3**, 1289–1306 (2011).
- Wang, H. *et al.* Biological and structural characterization of *Trypanosoma cruzi* phosphodiesterase C and Implications for design of parasite selective inhibitors. *J. Biol. Chem.* **287**, 11788–11797 (2012).
- Gould, M. K. *et al.* Cyclic AMP effectors in African trypanosomes revealed by genome-scale RNA interference library screening for resistance to the phosphodiesterase inhibitor CpdA. *Antimicrob. Agents Chemother.* **57**, 4882–4893 (2013).
- Caffrey, C. R. & Steverding, D. Kinetoplastid papain-like cysteine peptidases. *Mol. Biochem. Parasitol.* **167**, 12–19 (2009).
- Uzureau, P. *et al.* Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. *Nature* **501**, 430–434 (2013).
- Doyle, P. S. *et al.* A nonazole CYP51 inhibitor cures Chagas' disease in a mouse model of acute infection. *Antimicrob. Agents Chemother.* **54**, 2480–2488 (2010).
- Mach, J., Tachezy, J. & Sutak, R. Efficient iron uptake via a reductive mechanism in procyclic *Trypanosoma brucei*. *J. Parasitol.* **99**, 363–364 (2013).
- Gunatilleke, S. S. *et al.* Diverse inhibitor chemotypes targeting *Trypanosoma cruzi* CYP51. *PLoS Negl. Trop. Dis.* **6**, e1736 (2012).
- Andriani, G. *et al.* Antitrypanosomal lead discovery: identification of a ligand-efficient inhibitor of *Trypanosoma cruzi* CYP51 and parasite growth. *J. Med. Chem.* **56**, 2556–67 (2013).
- Hargrove, T. Y. *et al.* Complexes of *Trypanosoma cruzi* sterol 14 $\alpha$ -demethylase (CYP51) with two pyridine-based drug candidates for Chagas disease: structural basis for pathogen selectivity. *J. Biol. Chem.* **288**, 31602–31615 (2013).
- Hall, B. S., Bot, C. & Wilkinson, S. R. Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites. *J. Biol. Chem.* **286**, 13088–13095 (2011).
- Patterson, S. & Wyllie, S. Nitro drugs for the treatment of trypanosomatid diseases: past, present, and future prospects. *Trends Parasitol.* **30**, 289–98 (2014).
- Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–205 (2014).
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2013).
- Saha, A. *et al.* Prostaglandin E2 negatively regulates the production of inflammatory cytokines/chemokines and IL-17 in visceral leishmaniasis. *J. Immunol.* **193**, 2330–2339 (2014).
- Buckner, F. S., Verlinde, C. L., La Flamme, A. C. & Van Voorhis, W. C. Efficient technique for screening drugs for activity against *Trypanosoma cruzi* using parasites expressing beta-galactosidase. *Antimicrob. Agents Chemother.* **40**, 2592–2597 (1996).
- Downing, T. *et al.* Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **21**, 2143–2156 (2011).



46. Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–22 (2005).
47. El-Sayed, N. M. *et al.* Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404–409 (2005).
48. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
49. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
50. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

## Acknowledgments

From GSK, we thank Dolores Jimenez-Alfaro and SMTech Department for supplying compounds pre-dispensed in microtiter plates and managing compound logistics. We thank Carmen Magan, Diana Tegazzini, Juliana Cogo, Livia Bandeira, Ivan Caballero, David Calvo, Rosario Diaz and M Eugenia Cruz for the technical assistance in the screening and biological profiling of the compounds. Likewise, Jesus Presa for developing algorithms for automated image analysis. We also thank Nick Cammack, Emilio Diez, Lluís Balle, Francisco J Gamó, Mike Strange, Darren Green, Stephen Pickett and Ricardo Macarrón for critically commenting on the manuscript. Joanna Betts and Pankaj Agarwal facilitated computational biology support. The support and funding of Tres Cantos Open Lab Foundation is gratefully acknowledged. We are also indebted to Manu de Rycker and scientists from Dundee Drug Discovery Unit for the support, advice and training received on kinetoplastids biology. We thank DNDi and their team for their support and scientific discussions. Naomi Richardson of Magenta Communications Ltd received funding from GlaxoSmithKline for editorial assistance in the preparation of this manuscript.

## Author contributions

J.J.M., I.P., A.R. and J.M.F. conceptualized, planned and designed the work. J.J.M. and I.P. supervised all the experimental work. I.P., J.C., I.C., J.A.-P., E.A., V.B., F.deD.-A., I.R. and A.I.B. performed the screening assays and contributed to data analysis. I.P. and M.P.M. managed the project of *Leishmania*. A.I.B. and A.K. managed the project of *T. cruzi*. I.P., M.P.M., E.A. and I.R. managed the project of *T. brucei*. D.W.G., A.R. and M.N. transferred the initial assay protocols and strains for *Leishmania*, *T. cruzi* and *T. brucei* screening campaigns, respectively. G.C. performed the cheminformatic analysis and contribute to data analysis, including the selection of the three chemical boxes. V.K., A.S. and J.R.B. performed the bioinformatic analysis. D.H.D. contributed to the chemical biology analysis of the target space. A.K. and M.P.M. led the final selection of compounds included in the three chemical boxes. J.J.M., I.P., J.C., E.A., A.I.B., D.H.D., J.R.B., G.C., V.K., A.S., M.P.M. and A.K. wrote different sections of the manuscript and prepared figures. I.P., J.C., J.J.M. and J.R.B. contributed critical discussions of all sections. J.C. compiled the final table containing the biological profile of compounds in the chemical boxes, and contributed to formatting of the manuscript. J.J.M. integrated individual contributions and issued the final manuscript. All authors reviewed the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Peña, I. *et al.* New Compound Sets Identified from High Throughput Phenotypic Screening against Three Kinetoplastid Parasites: An Open Resource. *Sci. Rep.* **5**, 8771; DOI:10.1038/srep08771 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>