Host-directed therapy for intracellular bacterial Infections

Korbee, C.J.

3 | Combined Chemical Genetics and Data-driven Bioinformatics Approach Identifies Receptor Tyrosine Kinase Inhibitors as Host-directed Antimicrobials

Cornelis J. Korbee*, Matthias T. Heemskerk*, Dragi Kocev, Elisabeth van Strijen, Omid Rabiee, Kees L. M. C. Franken, Louis Wilson, Nigel D. L. Savage, Sašo Džeroski, Mariëlle C. Haks**, Tom H. M. Ottenhoff**

Antibiotic-resistance poses rapidly increasing global problems in combating multidrug-resistant (MDR) infectious diseases like MDR tuberculosis, prompting for novel approaches including host-directed therapies (HDT). Intracellular pathogens like *Salmonellae* and *Mycobacterium tuberculosis* (*Mtb*) exploit host pathways to survive. Only very few HDT compounds targeting host pathways are currently known. In a Library Of Pharmacologically Active Compounds (LOPAC) based drug-repurposing screen, we identify multiple compounds, which target Receptor Tyrosine Kinases (RTKs) and inhibit intracellular *Mtb* and *Salmonellae* more potently than currently known HDT-compounds. By developing a data-driven *in silico* model based on confirmed targets from public databases, we successfully predict additional efficacious HDT compounds. These compounds target host RTK signaling and inhibit intracellular (MDR-)*Mtb*. A complementary human kinome siRNA screen independently confirms the role of RTK signaling and kinases (BLK, ABL1 and NTRK1) in host control of *Mtb*. These approaches validate RTK signaling as a drugable host pathway for HDT against intracellular bacteria.

Adapted from:
Korbee, C.J.*, Heemskerk, M.T.*, Kocev, D., van Strijen, E., Rabiee, O., Franken, K.L.M.C., Wilson, L., Savage, N.D.L., Džeroski, S., Haks, M.C.**, Ottenhoff, T.H.M.**, 2018. **Combined chemical genetics and data-driven bioinformatics approach identifies receptor tyrosine kinase inhibitors as host-directed antimicrobials.** *Nat Commun* **9**, 358. doi:10.1038/s41467-017-02777-6**.**

* Contributed equally
** Contributed equally

Introduction

With an estimated 1/4 of the world population carrying a latent *Mycobacterium tuberculosis* (*Mtb*) infection, 10.5 million new cases and 1.8 million deaths annually, tuberculosis (TB) is an increasing global health issue[1-3]. This is further aggravated by the emergence of multi-, extensively- and totally drug-resistant (MDR/XDR/TDR) *Mtb* strains, threatening to render TB untreatable using current antibiotics[4-6]. In 2015 480,000 patients suffered from MDR-TB.

Although novel candidate antibiotics have recently been identified[7], current antibiotics already cover the majority of druggable targets of pathogens, resulting in a continuous decline in the number of new and approved antibiotics[8-13]. Intracellular bacteria such as *Salmonellae* and *Mtb* pose additional challenges by manipulating host signaling pathways to subvert innate and adaptive immunity. This, however, also creates potential for novel treatment strategies like host-directed therapy (HDT), to reprogram the host immune system by pharmacological and chemical-genetic manipulation. Importantly, HDT-driven manipulation of host signaling pathways may be effective also against drug-resistant bacteria, and help to restore host control of infection in metabolically perturbed cells[14,15]. Several recent studies, including our own, have demonstrated the feasibility of HDT approaches to inhibit bacteria both *in vitro* in human and murine cells[16-20] and *in vivo* in mice, rabbits and zebrafish[21-31]. Using reciprocal chemical-genetics targeting the human kinome, we previously identified AKT1 as a central regulator of *Salmonella enterica* serovar Typhimurium (*Stm*), *Mtb*, and MDR-*Mtb* survival. Treatment of infected cells with the kinase inhibitor H-89 significantly decreased intracellular bacterial loads. Despite H-89 being known as a PKA inhibitor, we demonstrated that this compound inhibited intracellular bacteria by targeting AKT1[16]. However, H-89 had a substantially lower impact on intracellular growth of *Mtb* compared to *Stm*, suggesting that *Mtb* modulates additional host signaling pathways to survive. This is in agreement with reports that *Mtb* arrests vesicle maturation at an earlier stage than *Stm*[16,32,33]. Other studies identified additional drugable human kinases that regulate *Mtb* survival, including TGFβRI and CSNK1[18] and imatinib-sensitive kinases ABL1 and ABL2[21]. In addition to kinases and kinase inhibitors, other potential targets and compounds for TB HDT were identified, including two antipsychotics (Haloperidol and Prochlorperazine) and an antidepressant (Nortryptiline)[19], phosphodiesterase inhibitors[22,23], anti-inflammatory agents like Ibuprofen[25], the FDA-approved drug Zileuton[26], the anti-diabetic drug Metformin[34], phenylbutyrate[35,36] and human metabolic targets[37,38]. Nevertheless, the field of TB HDT has not fully progressed towards clinical application and many interactions between host and bacterium remain to be deciphered. Therefore, better compounds are urgently needed as drug candidates for TB HDT as well as for the identification of cellular events occurring at the host-pathogen interface, which may enable rational drug design for HDT.

We used the screening assay described in **Chapter 2** in drug-repurposing screens, and identified compounds with host-directed anti-(myco)bacterial activity

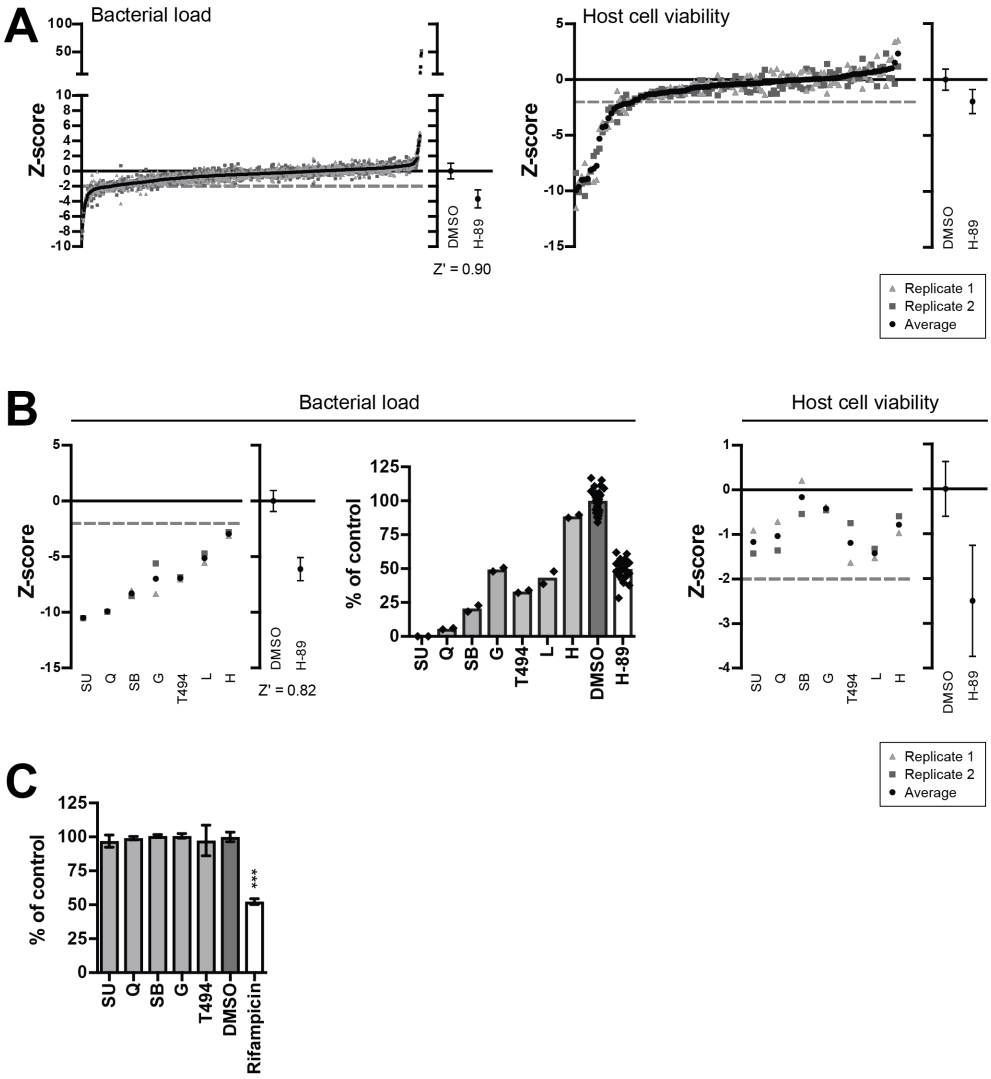
against *Mtb* and *Stm*, outperforming published HDT compounds' activities. Based on these data, together with confirmed target profiles of the screened compounds we next developed a predictive *in silico* model in order to be able to identify additional HDT compounds. This model was applied to predict host-directed compounds amongst all compounds present in the PubChem repository and to identify their key targets with predicted activity against intracellular *Stm* or *Mtb*. Interestingly, both our experimental wet lab screens as well as the novel *in silico* predictive model identified inhibitors of (growth factor) receptor tyrosine kinases (RTKs) and downstream intermediates of RTK signaling as candidate host-directed drugs to control intracellular infection. Moreover, an siRNA screen of the human kinome in *Mtb*-infected human cells independently validated a key role for RTK signaling in host control of *Mtb*. Thus, using two independent chemical genetic experimental approaches as well as a computational method, we find and validate RTK signaling as a novel important host pathway that controls intracellular *Mtb* (including MDR-*Mtb*) survival. This pathway is druggable by compounds and drugs including clinical drugs Dovitinib, AT9283 and ENMD-2076. These findings offer new approaches to combat intracellular infectious diseases in the face of rapidly rising multi drug resistance.

Results

Identification of host-directed antimicrobial compounds

We applied the novel screening assay described in **Chapter 2** for a TB drug-repurposing screen of a library of 1260 pharmacologically active compounds (LOPAC) in order to identify host-directed compounds with stronger activity against intracellular *Mtb* than H-89. The primary screen in the MelJuSo-*Mtb* intracellular infection model identified 110 compounds that significantly reduced and 16 compounds that increased intracellular bacterial loads. Ninety of these did not affect host cell viability (**Figure 1A** and **Table 1**) and were therefore pursued further. Seven compounds decreased *Mtb* bacterial load more potently than H-89 (**Table 2**). A rescreen of these 7 compounds confirmed their activity and 5 of these compounds again surpassed H-89 (SU 6656, Quinacrine, SB 216763, GW 5074 and Tyrphostin AG 494; **Figure 1B** and **Table 2**). **Figure 1B** shows z-score values in the left panel, with the actual percentage inhibition of *Mtb* growth shown in the middle panel, expressed as the % of control value. These latter data confirmed the strong inhibitory effect of these HDT compounds on intracellular *Mtb*. We next confirmed that these compounds exerted their antimicrobial effects via the host by excluding any direct microbicidal activity against extracellular *Mtb* (**Figure 1C**). As a control, the classical *Mtb* antibiotic rifampicin significantly inhibited *Mtb*.

To investigate whether also compounds existed with host-directed activity against *Stm*, and whether their activity was selective for *Mtb*, *Stm* or both, we also screened the same LOPAC library using the very similar HeLa-*Stm* infection model (**Figure 2A**). Twelve compounds were identified that significantly reduced the *Stm* bacterial load and 10 of these did not affect host cell viability (**Table 3**). 173



⬆ Figure 1. Identification of host-directed compounds inhibiting *Mtb*.

A. Results of a screen of 1260 compounds of the LOPAC library at 10 μM in the MelJuSo-*Mtb* infection model using *Mtb* constitutively expressing stable DsRed, expressed as z-scores (left panel). Individual replicates of the screened compounds are shown as grey points and the average z-score for each compound is displayed in black. The average z-score and standard deviation of the controls (DMSO and H-89) are displayed separately and the assay window expressed as a Z'-factor is shown below. Cell viability z-scores of the 110 hit compounds are shown in the right panel. The dashed line depicts a cut-off at a z-score of -2. **B.** A rescreen of the hit compounds that were superior to H-89 without affecting cell viability at 10 μM is shown as in **A**. The bacterial load is expressed as z-score in the utmost left panel and as a percentage of control

value in the middle panel to indicate the extent of bacterial inhibition. Individual screening datapoints are overlayed on the bar graph. Compound abbreviations: SU = SU 6656, Q = Quinacrine, SB = SB 216763, G = GW5074, T494 = Tyrphostin AG 494, L = L-594,881, H = Haloperidol. **C.** 6-Day treatment of an *Mtb* broth culture with the 5 hit compounds of the *Mtb* screen at 10 μM. Rifampicin (20 μg/ml) was used as a positive control. The average bacterial density +/- standard deviation of 4 replicates from a representative experiment (out of 3 experiments) is shown, expressed as a percentage of the DMSO control. Statistically significant difference compared to DMSO was tested using a one-way ANOVA ($F_{(6,25)} = 81.66$; *** = p-value < 0.001).

Table 1. LOPAC MelJuSo-*Mtb* primary screen hits using a bacterial load cut-off at z<-2 and a host cell viability cut-off at z>-2.

Bacterial load z-score	Cell viability z-score	Compound name
-6.02	**-1.36**	**SB 216763**
-5.79	**-0.33**	**SU 6656**
-5.25	**-1.14**	**Quinacrine dihydrochloride**
-4.86	**0.38**	**GW5074**
-3.87	-0.14	3',4'-Dichlorobenzamil hydrochloride
-3.83	**-1.01**	**Tyrphostin AG 494**
-3.77	-1.20	Haloperidol
-3.43	-1.55	Metaproterenol hemisulfate
-3.16	2.35	Serotonin hydrochloride
-3.14	0.96	Hydrocortisone 21-hemisuccinate sodium salt
-2.98	-1.07	Nortriptyline hydrochloride
-2.95	-0.88	LY-294,002 hydrochloride
-2.94	-0.14	Emodin
-2.93	-0.77	NNC 55-0396
-2.92	-0.64	Metrifudil
-2.91	0.76	LY-367,265

Bacterial load z-score	Cell viability z-score	Compound name
-2.78	-1.31	Fluspirilene
-2.75	0.06	nor-Binaltorphimine dihydrochloride
-2.66	-0.38	R-(-)-Fluoxetine hydrochloride
-2.65	-1.30	Loperamide hydrochloride
-2.63	0.08	BU224 hydrochloride
-2.62	-1.23	Nylidrin hydrochloride
-2.57	1.52	Farnesylthiosalicylic acid
-2.50	-0.46	PD 168,077 maleate
-2.48	-1.80	GR 127935 hydrochloride hydrate
-2.48	-0.49	5-Hydroxyindolacetic acid
-2.47	-0.46	Fenoldopam monohydrobromide
-2.47	0.01	S-Nitrosoglutathione
-2.44	-0.04	L-Histidine hydrochloride
-2.43	0.65	L-165,041
-2.41	0.04	4-Hydroxybenzhydrazide
-2.41	-1.83	Forskolin
-2.40	-0.38	AMN082
-2.36	-0.45	NAN-190 hydrobromide
-2.36	-0.68	Labetalol hydrochloride
-2.35	-0.41	Hexahydro-sila-difenidol hydrochloride, p-fluoro analog
-2.35	-0.01	L-Canavanine
-2.33	0.05	BRL 50481
-2.31	-0.11	N-Methyl-beta-carboline-3-carboxamide
-2.31	0.23	B-HT 933 dihydrochloride
-2.30	-1.05	Tyrphostin AG 527
-2.30	-0.52	1,3-Dimethyl-8-phenylxanthine

Bacterial load z-score	Cell viability z-score	Compound name
-2.30	0.10	Dopamine hydrochloride
-2.29	-1.17	A-77636 hydrochloride
-2.28	-1.53	Formoterol fumarate dihydrate
-2.25	-0.22	cis-(Z)-Flupenthixol dihydrochloride
-2.25	0.18	5-hydroxydecanoic acid sodium salt
-2.24	0.64	Isoguvacine hydrochloride
-2.23	-0.38	Nimesulide
-2.23	-0.66	alpha-Lobeline hydrochloride
-2.20	-0.67	Hydroxyurea
-2.19	-1.31	Fenoterol hydrobromide
-2.17	-0.34	L-733,060 hydrochloride
-2.17	-0.05	Minocycline hydrochloride
-2.17	-0.71	3-Nitropropionic acid
-2.17	0.41	LFM-A13
-2.16	-1.01	Nalidixic acid sodium salt
-2.16	-1.07	1,3,5-tris(4-hydroxyphenyl)-4-propyl-1H-pyrazole
-2.16	0.57	CR 2249
-2.15	1.02	p-MPPF dihydrochloride
-2.15	-0.04	Naltrexone hydrochloride
-2.15	0.87	Fluphenazine dihydrochloride
-2.14	-1.08	(-)-Tetramisole hydrochloride
-2.14	-0.31	Hydralazine hydrochloride
-2.14	0.09	(+)-Hydrastine
-2.12	0.54	MHPG sulfate potassium
-2.12	-0.67	6-Hydroxy-DL-DOPA
-2.12	-0.52	Ro 90-7501

Bacterial load z-score	Cell viability z-score	Compound name
-2.12	0.75	Neostigmine bromide
-2.11	-0.20	4-Amino-1,8-naphthalimide
-2.10	0.26	Flunarizine dihydrochloride
-2.09	-0.39	2-Methyl-5-hydroxytryptamine maleate
-2.09	0.01	L-368,899 hydrochloride
-2.09	-0.48	Tyrphostin AG 528
-2.09	-0.50	Lamotrigine
-2.09	-0.38	VER-3323 hemifumarate salt
-2.09	0.51	BU99006
-2.09	0.11	GYKI 52466 hydrochloride
-2.09	0.51	Hexamethonium bromide
-2.09	0.49	Flutamide
-2.07	0.05	Hypotaurine
-2.06	-0.06	NCS-356 sodium salt hydrate
-2.06	-0.07	(±)-7-Hydroxy-DPAT hydrobromide
-2.06	0.88	Hydroxylamine hydrochloride
-2.05	0.14	MDL 26,630 trihydrochloride
-2.04	-0.23	4-Hydroxy-3-methoxyphenylacetic acid
-2.03	-0.35	Fenofibrate
-2.03	-0.02	(±)-8-Hydroxy-DPAT hydrobromide
-2.02	-0.25	5-Hydroxy-L-tryptophan
-2.01	0.78	Methiothepin mesylate

Hit compounds that performed better than H-89 in both the primary screen and the rescreen are displayed in bold.

Compounds increased the *Stm* bacterial load without affecting host cell viability. Four of the hit compounds that decreased the bacterial load (Trimethoprim, Haloperidol, Mibefradil and Ofloxacin) were superior to H-89 in inhibiting

Table 2. Details of validated hit compounds from the *Mtb* and *Stm* LOPAC screens.

Abbr.	Compound name	Alternative name(s)	Primary screen z-score	Rescreen z-score	Activity
Mycobacterium tuberculosis					
SU	SU 6656	2,3-Dihydro-N,N-dimethyl-2-oxo-3-[(4,5,6,7-tetrahydro-1H-indol-2-yl)methylene]-1H-indole-5-sulfonamide	**-5.79**	**-10.51**	Src family kinase inhibitor
Q	Quinacrine dihydrochloride		**-5.25**	**-9.90**	MAO inhibitor
SB	SB 216763	3-(2,4-Dichlorophenyl)-4-(1-methyl-1H-indol-3-yl)-1H-pyrrole-2,5-dione	**-6.02**	**-8.29**	GSK-3 kinase inhibitor
G	GW5074	3-(3, 5-Dibromo-4-hydroxybenzylidine-5-iodo-1,3-dihydro-indol-2-one)	**-4.86**	**-6.98**	Raf1 kinase inhibitor
T494	Tyrphostin AG 494	N-Phenyl-3,4-dihydroxybenzylidenecyanoacetamide	**-3.83**	**-6.93**	EGFR kinase inhibitor
L	3',4'-Dichlorobenzamil hydrochloride	L-594,881	**-3.87**	-5.13	$Na^+/Ca2^+$ exchanger inhibitor
H	Haloperidol		**-3.77**	-2.96	D2/D1 dopamine receptor antagonist
Salmonella Typhimurium					
T	Trimethoprim		**-4.06**	-12.18	Antibiotic; dihydrofolate reductase inhibitor
H	Haloperidol		**-3.90**	-12.09	D2/D1 dopamine receptor antagonist
M	Mibefradil dihydrochloride	Ro 40-5967; (1S,2S)-2-[2[[3-(2-benzimidazolylpropyl]methylamino]ethyl]-6-fluoro-1,2,3,4-tetrahydro-1-isopropyl-2-naphthyl methoxyacetate dihydrochloride	**-3.64**	**-12.76**	$Ca2^+$ channel blocker
O	Ofloxacin	Ofloxacine; DL-8280; HOE-280	**-3.45**	-11.60	Antibiotic; DNA synthesis inhibitor

Z-scores lower than the z-score of H-89 are displayed in bold.

intracellular *Stm* (**Table 2**). Mibefradil again exceeded the inhibitory effect of H-89 in a rescreen (**Figure 2B**), while all four compounds consistently and strongly decreased the *Stm* bacterial load. While **Figure 2A** and the left panel of **Figure 2B** show z score values, the percentage inhibition of intracellular *Stm* growth is shown in the middle panel of **Figure 2B**, expressed as the % of control value, demonstrating the strong inhibitory effect of these HDT compounds on intracellular *Stm*. We next excluded any direct microbicidal activity of these HDT-compounds against extracellular *Stm* (**Figure 2C**). By contrast, Trimethoprim and Ofloxacin (both known antibiotics), which were part of the LOPAC library and therefore tested here as well, directly inhibited extracellular *Stm* as expected. The

fact that these known antibiotics for *Stm* were hits in our screen further confirms the strength and validity of our approach, showing that we can clearly distinguish antibiotics from host-directed compounds. Taken together, Haloperidol (a known HDT inhibitor[19]) and Mibefradil (newly discovered here) were confirmed and identified, respectively, as host-directed inhibitors of *Stm*.

Interestingly, a comparison of the *Mtb* and *Stm* HDT compound screening results revealed a highly limited overlap between hits in the two infection models (**Figure 2D**). This observation agrees well with reports that *Mtb* and *Stm* arrest

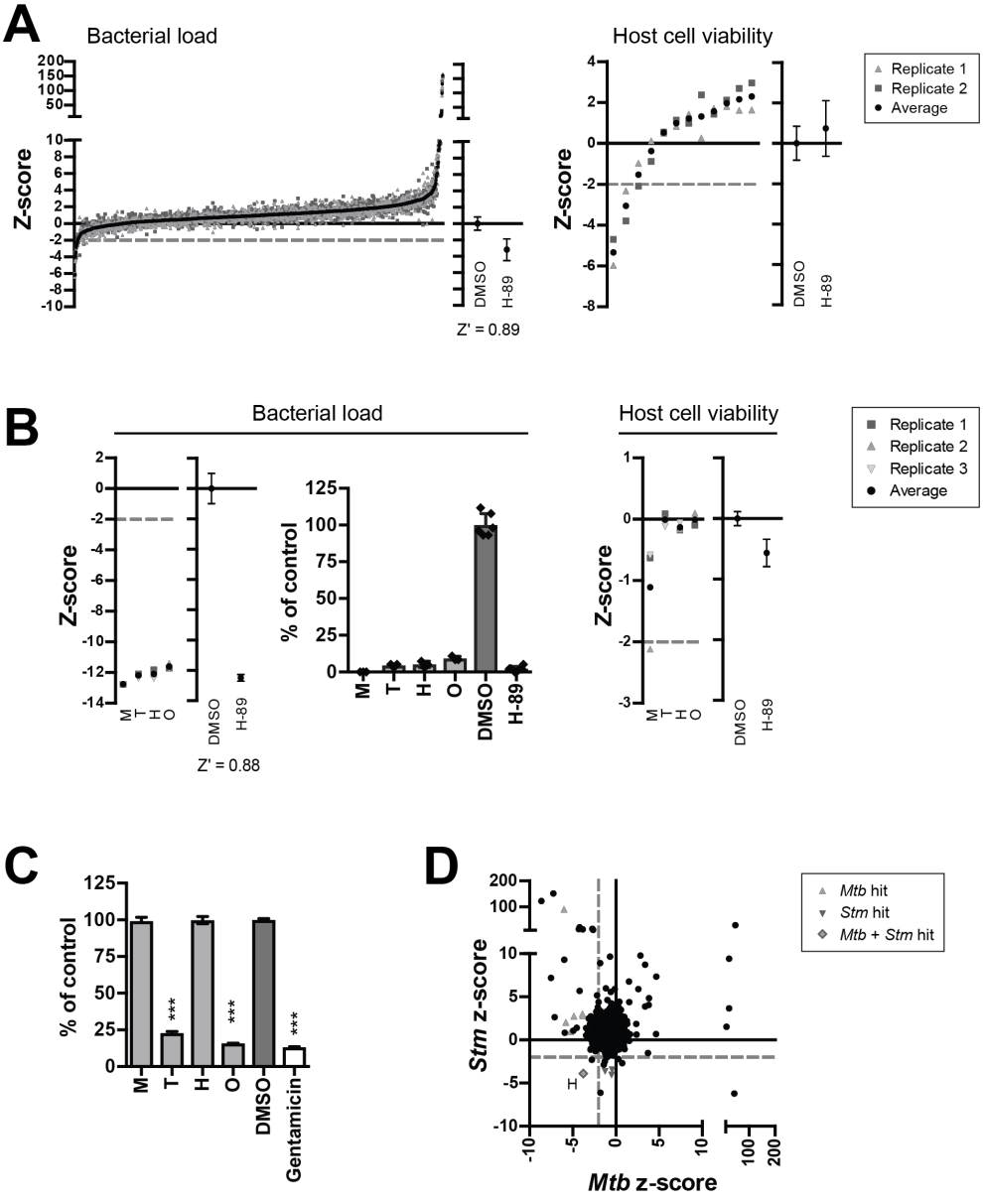


Table 3. LOPAC HeLa-*Stm* primary screen hits using a bacterial load cut-off at z<-2 and a host cell viability cut-off at z>-2.

Bacterial load z-score	Cell viability z-score	Compound name
-4.06	-1.54	Trimethoprim
-3.90	0.98	Haloperidol
-3.64	**1.57**	**Mibefradil dihydrochloride**
-3.45	1.21	Ofloxacin
-2.86	1.96	Demeclocycline hydrochloride
-2.70	-0.38	Doxazosin mesylate
-2.47	2.29	Metergoline
-2.30	1.31	Fluspirilene
-2.20	0.52	8-(3-Chlorostyryl)caffeine
-2.00	2.15	GW2974

Hit compounds that performed better than H-89 in both the primary screen and the rescreen are displayed in bold.

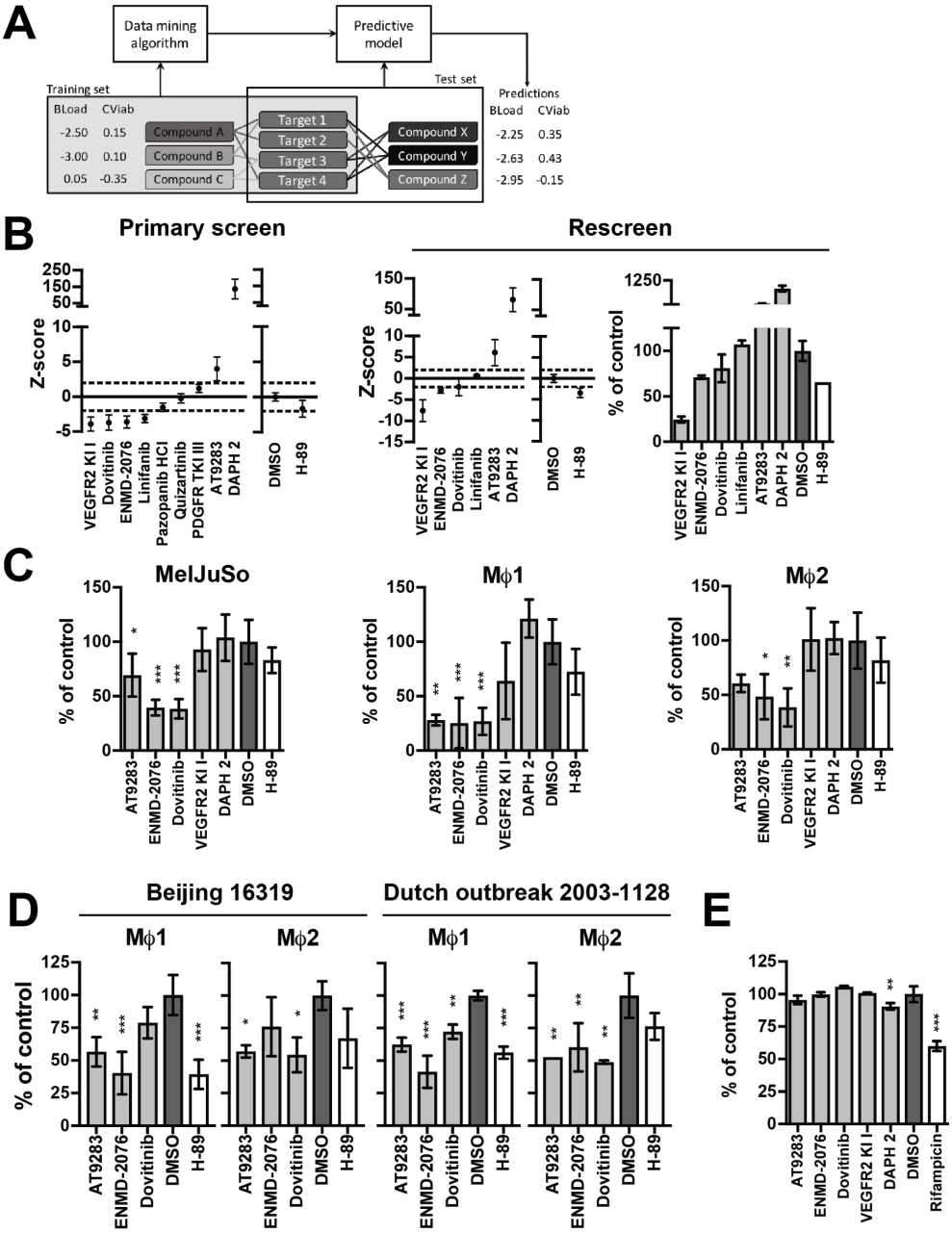
⬅ Figure 2. Identification of host-directed compounds inhibiting *Stm*.

A. Screen of the LOPAC library in the HeLa-*Stm* infection model using *Stm* constitutively expressing stable DsRed, as in **Figure 2A**. **B.** Rescreen of the hit compounds from the HeLa-*Stm* screen that were superior to H-89 without affecting cell viability, displayed as in **Figure 2B**. The bacterial load is expressed as z-score in the utmost left panel and as a percentage of control value +/- standard deviation in the middle panel to indicate the extent of bacterial inhibition. Compound abbreviations: T = Trimethoprim, H = Haloperidol, M = Mibefradil, O = Ofloxacin. **C.** Overnight treatment of a *Stm* broth culture with the hit compounds at 10 µM. Gentamicin (50 µg/ml) was used as a positive control. The average bacterial density +/- standard deviation of 6 replicates from a representative experiment out of 3 experiments is shown. The bacterial load is expressed percentage of the DMSO control value to indicate the extent of bacterial inhibition. Statistically significant difference compared to DMSO was tested using a one-way ANOVA ($F_{(5,30)}$ = 4871; *** = p-value < 0.001). **D.** Comparison of the *Stm* and *Mtb* primary screening data. Compounds that were superior to H-89 and subsequently confirmed in a rescreen are indicated in grey triangles. H = haloperidol.

vesicle maturation at different stages[16,32,33]. Haloperidol was the only compound that inhibited both *Mtb* and *Stm*.

Identification of HDT compounds using an *in silico* model

We next decided to use the above experimental data obtained in our LOPAC screens, and combine these with bioactivity assay based data available for all



1260 LOPAC compounds in PubChem, to develop a novel bioinformatics predictive model using machine learning. The model was constructed to predict new chemical compounds with host-directed activity against intracellular *Stm* or *Mtb*, based on target protein profiles identified by machine learning from our own LOPAC screening data (**Figure 3A**). An extended description of the machine learning methods describing the predictive model is provided as **Supplementary Information** at the end of this chapter. In brief, we first linked all LOPAC

⬅ Figure 3. Screen of *in silico* predicted compounds active against intracellular *Mtb*.

A. Schematic of the predictive model. Abbreviations: BLoad = bacterial load z-score; CViab = cell viability z-score. **B.** Compound primary screen (left panel) and rescreen (middle panel) at 10 µM in the MelJuSo-*Mtb* model using *Mtb* constitutively expressing stable DsRed, expressed as mean z-scores +/- standard deviation. Dashed lines indicate a hit cut-off at a z-score of 2 or -2. Average z-score and standard deviations of controls (DMSO and H-89) are displayed separately. To indicate the extent of bacterial inhibition, rescreen results are expressed both as z-score and as percentage of control value +/- standard deviation in the right panel. **C.** CFU assay of MelJuSo (left panel) and human primary Mϕ1 (middle panel) and Mϕ2 (right panel) *Mtb* infection models treated with the hit compounds from **B** at 10 µM. Mϕ1 and Mϕ2 models have been described by Verreck *et al.*[39]. Shown are representative data out of 3 experiments (MelJuSo) and data from a representative donor (Mϕs) out of 2 (Mϕ1) or 5 (Mϕ2) blood bank donors. To indicate the extent of bacterial inhibition, results are expressed as percentage of control +/- standard deviation. Replicates in the MelJuSo model: AT9283 and ENMD-2076: n=6; Dovitinib, VEGFR KI I and DAPH2: n=5; DMSO and H-89: n=9. Replicates in the Mϕ models: AT9283, ENMD-2076, Dovitinib, VEGFR KI I and DAPH 2: n=3; DMSO and H-89: n=5. Statistically significant difference compared to DMSO was tested by one-way ANOVA (MelJuSo: $F_{(6,39)} = 16.35$; Mϕ1: $F_{(6,18)} = 10.88$; Mϕ2: $F_{(6,18)} = 5.23$; * = p-value < 0.05, ** = p-value <0.01, *** = p-value < 0.001). **D.** CFU assay of the Mϕ1 and Mϕ2 models infected with MDR-*Mtb* (Beijing family China 16319 and Dutch outbreak 2003-1128) and treated with the validated hit compounds from **C** at 10 µM. Shown are data (n=3 technical replicates) from a representative donor out of 4 different blood bank donors, displayed as percentage of the DMSO control +/- standard deviation. Statistically significant differences compared to DMSO were tested by one-way ANOVA (Mϕ1 Beijing: $F_{(4,10)} = 11.43$; Mϕ2 Beijing: $F_{(4,10)} = 3.72$; Mϕ1 Dutch outbreak: $F_{(4,10)} = 29.09$; Mϕ2 Dutch outbreak: $F_{(4,10)} = 8.81$; * = p-value < 0.05, ** = p-value <0.01, *** = p-value < 0.001). **E.** Six-day treatment of *Mtb* cultures with hit compounds at 10 µM. Rifampicin (20 µg/ml) was used as a positive control. Average bacterial density +/- standard deviation of 3 replicates is shown, expressed as a percentage of the DMSO control. Displayed are representative results out of 3 individual experiments. Statistically significant difference compared to DMSO was tested by one-way ANOVA ($F_{(6,25)} = 101.4$; ** = p-value <0.01, *** = p-value < 0.001).

compounds to PubChem, and retrieved bioassay data by using a pre-processing pipeline (**Supplementary Figure 1A**), which identified 1058 confirmed human protein targets for these 1260 compounds. This resulted in a data table comprising all LOPAC compounds annotated with their corresponding impact on intracellular bacterial survival and host cell viability from our screens, expressed as z-scores, combined with their PubChem bioassay activity for each confirmed human target. An example of the table structure is shown in **Supplementary Table 1**. This was then used as a training set to learn ensembles of predictive clustering trees (PCTs; **Supplementary Figure 2**) to predict impact on intracellular bacterial survival and host cell viability. We next employed this *in silico* tool (the learned model) to identify and select candidate compounds from PubChem with predicted host-directed antimicrobial activity. Querying PubChem for compounds that are known to target one or more of the above 1058 confirmed human protein targets yielded 460,580 compounds, which were then annotated with their bioassay data and fed into the predictive model as a testing set. Using the ensembles of PCTs learned from the training data to predict the intracellular bacterial survival and host cell viability z-scores of these 460,580 compounds, we identified 47 candidate compounds predicted to affect intracellular *Mtb* load (**Table 4**) and 30 compounds predicted to affect intracellular *Stm* load (**Table 5**). From these two lists of compounds, commercially available compounds (**Table 6**) were ordered and screened in the MelJuSo-*Mtb* and HeLa-*Stm* infection models. As the PubChem BioAssay data contains compound-target relations based only on IC50 and EC50 values as well as binding constants, the predictive model was able to identify only compound-target interactions rather than the direction of the target effects. Thus, as we were therefore unable to predict whether compounds would actually inhibit or activate their associated targets, predicted negative z-scores might result in experimentally positive z-scores in *in vitro* intracellular bacterial inhibition tests and vice versa. In the *Mtb* screen 6 out of 9 compounds predicted to affect the bacterial load indeed decreased or increased the bacterial load (**Figure 3B, left panel**). A rescreen of the hit compounds confirmed 5 out of 6 hits (VEGFR KI I, ENMD-2076, Dovitinib, AT9283 and DAPH 2; **Figure 3B, middle and right panels**). Results are shown as z-scores as well as the percentage inhibition of *Mtb* growth expressed as the % of control value, to confirm the strong inhibitory effect of these HDT compounds on intracellular *Mtb* (**Figure 3B, utter right panel**).

As compound autofluorescence might result in false positive z-scores in our assay, we further validated all the confirmed hit compounds independently in classical CFU assays, both in cell lines and in primary human Mϕs. The compounds AT9283, ENMD-2076 and Dovitinib significantly decreased *Mtb* CFUs in both MelJuSo cells and human primary Mϕs (**Figure 3C**; results are shown as percentage inhibition of *Mtb* growth expressed as % of control value). Importantly, AT9283, ENMD-2076 and Dovitinib also reduced CFUs in human primary Mϕs infected with two different MDR-*Mtb* strains (Beijing family China 16319 and Dutch outbreak 2003-1128; **Figure 3D**). These data independently confirm and validate the results obtained in our novel screening and prediction pipeline, and -importantly- extend the newly identified HDT-compounds' effects to intracellular multidrug resistant bacteria. Finally, none of the compounds directly affected

Table 4. Complete list of compounds identified as potential hits from the *Mtb* predictive model output.

PubChem ID	Predicted bacterial load z-score	Predicted cell viability z-score	Reliability
6604502	-2.59	-0.59	0.61
46233889	-2.38	-0.84	0.54
46235770	-2.38	-0.84	0.54
56945171	-2.38	-0.84	0.54
56945172	-2.38	-0.84	0.54
56945173	-2.38	-0.84	0.54
56945174	-2.38	-0.84	0.54
56945175	-2.38	-0.84	0.54
56945277	-2.38	-0.84	0.54
24995659	-2.35	-0.96	0.61
10113978	-2.29	-0.91	0.53
11496629	-2.27	-0.95	0.54
10907042	-2.24	-0.87	0.54
59627005	-2.21	-0.93	0.54
16041424	-2.15	-0.89	0.54
9977819	-2.14	-0.93	0.53
6419834	-2.14	-0.93	0.53
67161540	-2.13	-0.94	0.52
11485656	-2.10	-0.87	0.54
16757867	-2.09	-0.73	0.68
6711154	-2.08	-0.93	0.68
10142586	-2.07	-0.99	0.52
657806	-2.07	-0.48	0.66
9532258	-2.05	-0.86	0.71
10209082	-2.01	-0.93	0.66
24889392	-2.00	-0.87	0.65
5782470	-1.99	-0.64	0.75
660914	-1.95	-0.22	0.67
1552034	-1.91	-0.50	0.79
3246585	-1.89	-0.55	0.66
5284352	-1.86	-0.48	0.62
16235522	-1.84	-0.42	0.71
5284416	-1.84	-0.72	0.65
661761	-1.83	-0.23	0.66
6097179	-1.82	-0.43	0.65
1745927	-1.82	-0.34	0.75
3246543	-1.82	-0.63	0.63
6918515	-1.81	0.31	0.75
5765289	-1.80	-0.47	0.78
664864	-1.78	-0.25	0.66
1363897	-1.78	-0.19	0.71
3246495	-1.78	-0.50	0.65
6604530	-1.78	-0.50	0.65
663169	-1.78	-0.50	0.65
456214	-1.78	-0.45	0.63
660368	-1.77	-0.40	0.65
660838	-1.77	-0.52	0.65

Commercially available compounds selected for the study are indicated in bold.

Table 5. Complete list of compounds identified as potential hits from the *Stm* predictive model output.

PubChem ID	Predicted bacterial load z-score	Predicted cell viability z-score	Reliability
5035	-1.88	0.60	0.65
50994498	-1.68	0.52	0.72
202478	-1.58	0.47	0.71
7333	-1.56	0.04	0.66
44474938	-1.55	0.54	0.73
57402462	-1.55	0.54	0.73
11743300	-1.52	0.44	0.73
13998486	-1.52	0.44	0.73
15163141	-1.52	0.44	0.73
185834	-1.52	0.44	0.73
4416	-1.52	0.44	0.73
44303090	-1.52	0.44	0.73
44398003	-1.52	0.44	0.73
44398036	-1.52	0.44	0.73
44398114	-1.52	0.44	0.73
47641	-1.52	0.44	0.73
50266	-1.52	0.44	0.73
5474589	-1.52	0.44	0.73
6437849	-1.52	0.44	0.73
6439331	-1.52	0.44	0.73
65638	-1.52	0.44	0.73
6713949	-1.52	0.44	0.73
72027	-1.52	0.44	0.73
73345319	-1.52	0.44	0.73
93365	-1.52	0.44	0.73
9799239	-1.52	0.44	0.73
9841596	-1.52	0.44	0.73
9951886	-1.52	0.44	0.73
9954083	-1.52	0.44	0.73
9417	-1.51	0.15	0.68

Commercially available compounds selected for the study are indicated in bold.

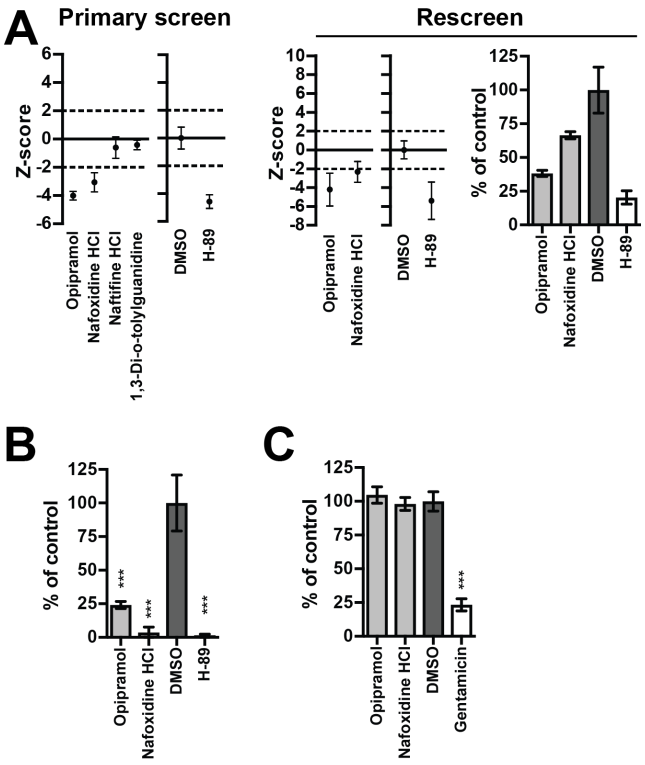
Table 6. Compounds selected from the predictive model output.

PubChem ID	Compound name	Predicted bacterial load z-score	Predicted cell viability z-score	Reliability score	Primary screen z-score	Rescreen z-score	Activity
Mycobacterium tuberculosis							
10113978	Pazopanib·HCl	-2.30	-0.91	0.53	-1.50	N.D.	Receptor Tyrosine Kinase (RTK) inhibitor
11496629	AT9283	-2.27	-0.95	0.54	**4.01**	**6.09**	JAK/Aurora kinase inhibitor
16041424	ENMD-2076	-2.15	-0.89	0.54	**-3.62**	**-2.83**	RTK/Aurora A inhibitor
11485656	Linifanib (ABT-869)	-2.11	-0.87	0.54	**-3.13**	0.66	VEGFR/PDGFR inhibitor
10907042	PDGFR Tyrosine Kinase Inhibitor III	-2.24	-0.88	0.53	1.19	N.D.	PDGFR inhibitor
9977819	Dovitinib (TKI-258, CHIR-258)	-2.14	-0.93	0.53	**-3.70**	**-2.02**	RTK inhibitor
6419834	VEGFR2 Kinase Inhibitor I	-2.14	-0.93	0.53	**-3.91**	**-7.63**	VEGFR2 inhibitor
6711154	DAPH 2	-2.08	-0.93	0.68	**136.21**	**80.77**	PKC inhibitor
24889392	Quizartinib	-2.00	-0.87	0.65	-0.24	N.D.	FLT3 inhibitor
Salmonella Typhimurium							
4416	Nafoxidine hydrochloride	-1.52	0.44	0.73	**-3.06**	**-2.33**	Estrogen receptor modulator
7333	1,3-Di-o-tolylguanidine	-1.56	0.04	0.66	-0.42	N.D.	Sigma 1 receptor agonist
47641	Naftifine hydrochloride	-1.52	0.44	0.73	-0.61	N.D.	Fungal squalene epoxidase inhibitor
9417	Opipramol	-1.51	0.15	0.69	**-4.01**	**-4.21**	Sigma receptor agonist

Z-scores exceeding the cut-off (2 < z-score <-2) are displayed in bold, N.D. = not determined.

extracellular bacterial growth in liquid cultures, while classical antibiotics (rifampicin) did, confirming that the mode of action of the new HDT-compounds is via modulation of host and not direct bacterial mechanisms (**Figure 3E**).

Using this same screening and validation approach for *Stm* in the HeLa-*Stm* infection model, we confirmed that 2 out of 4 compounds predicted to affect *Stm* survival indeed decreased the bacterial load of *Stm*-infected cells in a primary screen (**Figure 4A, left panel**). Both of these hits (Opipramol and Nafoxidine) were subsequently confirmed in a rescreen (**Figure 4A, middle and right panels**; results shown as z-scores and as % inhibition of *Stm* growth expressed as the % of control value). Both hit compounds also reduced the *Stm* bacterial load independently in classical CFU assays (**Figure 4B**), again without directly affecting bacterial growth in a liquid overnight *Stm* culture (**Figure 4C**), confirming their HDT



⬆ Figure 4. Screen of *in silico* predicted compounds active against intracellular *Stm*.

A. Chemical compound primary screen (left panel) and rescreen (middle panel) at 10 µM in the HeLa-*Stm* infection model using *Stm* constitutively expressing stable DsRed, expressed as mean z-scores +/- standard deviation. Horizontal dashed lines indicate a hit cut-off at a z-score of 2 or -2. The average z-scores and standard deviations of the controls (DMSO and H-89) are displayed separately. To indicate the extent of bacterial inhibition, rescreen results are expressed as percentage of control value +/- standard deviation in the right panel. **B.** CFU assay of the HeLa-*Stm* infection model treated with the validated hit compounds from **A** at 10 µM. Shown are representative data out of 3 independent experiments, displayed as a percentage of the DMSO control. The average +/- standard deviation of 3 replicates is shown. Statistically significant differences compared to DMSO were tested using a one-way ANOVA ($F_{(3,8)}$ = 56.31; *** = p-value < 0.001). **C.** Overnight treatment of *Stm* broth cultures with the hit compounds at 10 µM. The *Stm* antibiotic Gentamicin (50 µg/ml) was used as a positive control. The average bacterial density +/- standard deviation of 3 replicates is shown, expressed as a percentage of the DMSO control. Displayed are representative results out of 3 individual experiments. Statistically significant differences compared to DMSO were tested using a one-way ANOVA ($F_{(3,38)}$ = 579.5; *** = p-value < 0.001).

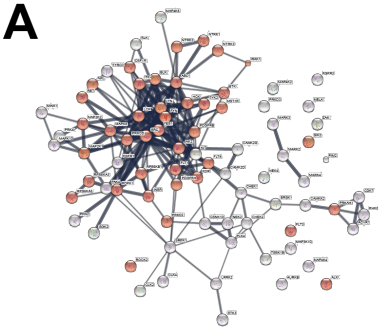
mode of action. These data therefore confirm and validate our novel screening and prediction pipeline not only for *Mtb* but also *Stm*.

Thus, we have successfully developed and used a data-driven novel *in silico* predictive model to identify host-directed compounds with antimicrobial activity against intracellular bacteria. The model significantly enhanced the identification of *de novo* hit compounds (5 out of 9 (55.6%) and 2 out of 4 (50%) for *Mtb* and *Stm*, respectively) compared to random LOPAC library primary screening (126 out of 1260 (10%) for *Mtb* and 185 out of 1260 (14.6%) for *Stm*). In addition, the results were replicated and validated in primary human Mϕs infected with *Mtb*, strongly agreeing with and further validating the MelJuSo-*Mtb* model used in our novel flow cytometry-based screening assay.

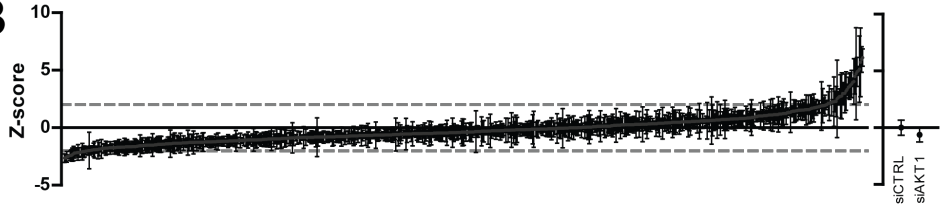
RTK signaling is a novel host pathway controlling *Mtb*

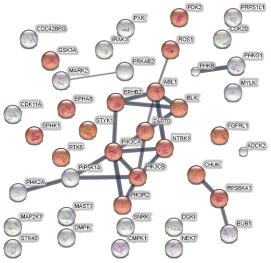
As AT9283, ENMD-2076 and Dovitinib are RTK inhibitors[40-42] we used a chemical genetics approach to confirm a role for RTK signaling in host-mediated *Mtb* control. We first retrieved human protein targets of AT9283, ENMD-2076 and Dovitinib from the Compound Bioactivity section in ChEMBL (http://www.www.ebi.ac.uk/chembl/) and further downselected targets for which the compounds were annotated as 'Active'. Because no targets annotated as 'Active' could be retrieved for AT9283 and ENMD-2076, we first constructed a STRING protein network and performed gene ontology (GO) analysis using the targets of Dovitinib (n=86 proteins; **Figure 5A**). Due to the hierarchical organization of GO-terms, general cellular and molecular functions tend to be highly enriched in GO term enrichment analyses. Therefore, we focused on identifying the highest ranked GO terms that described distinct pathways rather than the overall highest ranked GO terms. As expected from the reported target specificities of Dovitinib[42], 'transmembrane receptor protein tyrosine kinase signaling pathway' (GO:0007169, false discovery rate (FDR) 3.82E-33) was the highest ranking enriched pathway and 40 protein targets participated in this pathway (**Figure 5A** and **Figure 6A**). We next verified that both AT9283 and ENMD-2076 target RTKs by retrieving human protein targets from the Target Summary section in ChEMBL and performed an identical STRING analysis (**Figures 6B** and **6C**). Even though this analysis resulted in small networks due to the limited number of studied targets and the lists of targets from the Target Summary section also include non-confirmed targets, the GO-term 'transmembrane receptor protein tyrosine kinase signaling pathway' (GO:0007169) was again highly enriched in the target networks of AT9283 (FDR 1.11E-12) and ENMD-2076 (FDR 6.47E-5).

To independently confirm RTK signaling as a functional pathway that controls intracellular survival of *Mtb*, we next performed an unbiased siRNA screen of the human kinome in the MelJuSo-*Mtb* infection model (**Figure 5B**), agnostic to the above data. The siRNA screen identified 20 targets that decreased and 21 that increased the *Mtb* bacterial load whilst not affecting host cell viability (**Table 7**). These 41 hit kinases were then used in a STRING protein network and GO analysis. Independently confirming the STRING analysis of the targets of Dovitinib, AT9283 and ENMD-2076, also in this analysis 'transmembrane receptor protein tyrosine kinase signaling pathway' (GO:0007169, FDR 1.32E-13) was the highest-ranking enriched pathway, and 18 hit kinases from the siRNA screen

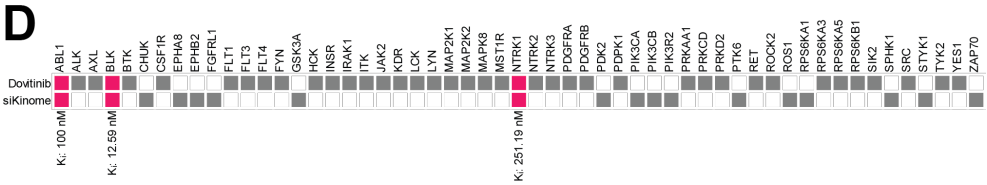


#pathway ID	pathway description	gene count	FDR
GO.0006468	protein phosphorylation	78	1.82E-104
GO.0006796	phosphate-containing compound metabolic process	77	5.02E-75
GO.0046777	protein autophosphorylation	48	5.02E-75
GO.0006464	cellular protein modification process	77	2.12E-62
GO.0018108	peptidyl-tyrosine phosphorylation	37	3.82E-55
GO.0018193	peptidyl-amino acid modification	51	3.73E-48
GO.0035556	intracellular signal transduction	57	3.48E-40
GO.0007169	**transmembrane receptor protein tyrosine kinase signaling pathway**	40	3.82E-35
GO.0001934	positive regulation of protein phosphorylation	40	7.72E-33
GO.0007167	enzyme linked receptor protein signaling pathway	41	7.77E-32





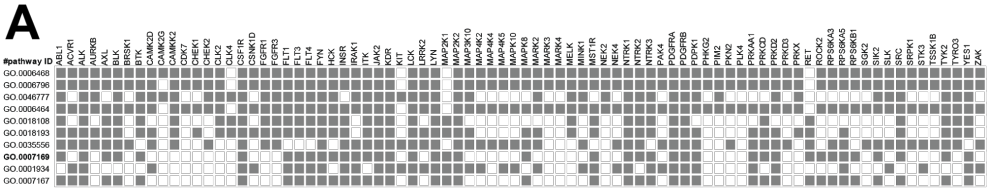
#pathway ID	pathway description	gene count	FDR
GO.0016310	phosphorylation	36	2.07E-38
GO.0006468	protein phosphorylation	29	1.86E-29
GO.0044267	cellular protein metabolic process	30	1.32E-13
GO.0007169	**transmembrane receptor protein tyrosine kinase signaling pathway**	18	1.32E-13
GO.0018108	peptidyl-tyrosine phosphorylation	10	3.76E-10
GO.0035556	intracellular signal transduction	21	6.49E-10
GO.0046777	protein autophosphorylation	10	2.44E-09
GO.0007166	cell surface receptor signaling pathway	20	5.30E-08
GO.0044237	cellular metabolic process	35	7.52E-08
GO.0046834	lipid phosphorylation	6	7.52E-08



participated in this pathway (**Figure 5C** and **Figure 6D**). Three of the kinases (ABL1, BLK and NTRK1) were both hits in the siRNA screen and confirmed targets of Dovitinib (**Figure 5D**). Of these 3 kinases, only ABL1 was present in the potential target networks of AT9283 and ENMD-2076 (**Figures 6B** and **6C**). However, a lower dissociation constant (K_i) is reported in ChEMBL for the interaction between Dovitinib and BLK (K_i: 12.59 nM) than between Dovitinib and ABL1 (K_i: 100 nM), suggesting that BLK is targeted more strongly by Dovitinib. To identify the top enriched RTK signaling pathway targeted by Dovitinib and siRNA, we used the kinases shown in **Figure 5D** in a STRING analysis. This identified the neurotrophin signaling pathway as the top enriched KEGG pathway (**Figure 7**)[43]. Silencing of Neurotrophic Receptor Tyrosine Kinase 1 (NTRK1) resulted in an increased *Mtb* bacterial load (**Table 7**), establishing a functional link between

⬅ Figure 5. Identification of host kinases controlling intracellular *Mtb* survival.
A. STRING network of confirmed targets of Dovitinib retrieved from the ChEMBL repository Compound Bioactivity section (left panel). Individual proteins are displayed as nodes. Lines represent protein-protein interactions and the thickness of the lines indicates confidence. Proteins participating in the 'transmembrane receptor tyrosine kinase signaling pathway' are displayed in red. The top 10 enriched GO terms in the 'Biological Function' category are displayed along with the number of genes/proteins annotated with the indicated GO terms and the false discovery rate (FDR) of the enrichment (right panel). **B.** Results of a siRNA screen of the human kinome in the MelJuSo-*Mtb* infection model using *Mtb* constitutively expressing destabilized DsRed, expressed as z-scores. The average z-score +/- standard deviation for each siRNA pool is displayed. A hit cut off at z=2 or z=-2 is displayed as a dashed line. The average z-score and standard deviation of the controls (siCTRL and siAKT) are displayed separately. SiCTRL: non-targeting siRNA pool. **C.** STRING network of the siRNA screen hits (left panel) is displayed along the top 10 enriched GO terms in the 'Biological Function' category (right panel), as in **A**. **D.** Participation of individual targets of Dovitinib (top row) or hits from the siRNA screen (bottom row) in the 'transmembrane receptor tyrosine kinase signaling pathway' is indicated by filled squares. Proteins that are both targeted by Dovitinib and were a hit in the siRNA screen are shown in magenta. Dissociation constants (K_i) retrieved from ChEMBL are shown below for the interaction between Dovitinib and ABL1, BLK and NTRK1.

➡ Figure 6 (next page). STRING analysis of targets of Dovitinib, AT9283, ENMD-2076 and siRNA screening hits.
A. Association of individual targets of Dovitinib with the top 10 enriched GO terms is indicated by filled squares. **B.** STRING network of potential targets of AT9283 retrieved from the ChEMBL repository Target Summary section (top panel). Individual proteins are displayed as nodes. Lines represent protein-protein interactions and the thickness of the lines indicates confidence. Proteins participating in the 'transmembrane receptor tyrosine kinase signaling pathway' are displayed in red. The top 10 enriched GO terms in the 'Biological Function' category are displayed along with the number of genes/proteins annotated with the indicated GO terms and the false discovery rate (FDR) of the enrichment (bottom panel). **C.** STRING network of potential targets of ENMD-2076 retrieved from the ChEMBL repository Target Summary section (top panel) and the top 18 enriched GO terms in the 'Biological Function' category (bottom panel) are displayed as in **B**. **D.** Association of individual siRNA hit kinases with the top 10 enriched GO terms is indicated by filled squares.



neurotrophin signaling and *Mtb* survival.

Thus, using independent chemical genetic, functional and computational approaches, we find and validate that (1) RTK signaling is a novel host pathway that controls intracellular (MDR)-*Mtb* survival and that (2) repurposable drugs such as Dovitinib, AT9283 and ENMD-2076 that target RTK signaling are new candidates for HDT in treating TB, including MDR-*Mtb*.

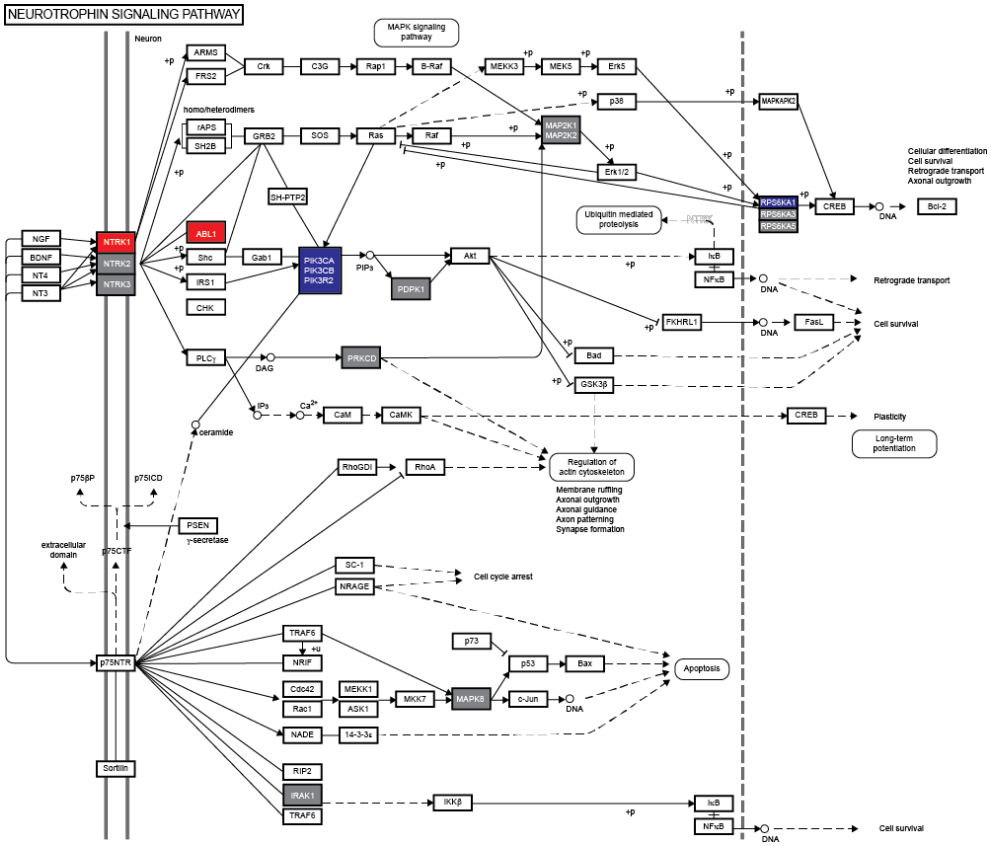
Table 7. siKinome screen hits in the MelJuSo-*Mtb* infection model.

GenBank Accession	Gene Symbol	Z-score
NM_006213	PHKG1	-2,65
NM_133494	NEK7	-2,65
NM_018425	PI4KII	-2,50
NM_032017	MGC4796	-2,47
NM_019884	GSK3A	-2,44
NM_014975	SAST	-2,42
NM_006219	PIK3CB	-2,34
NM_001079	ZAP70	-2,25
NM_005157	ABL1	-2,19
NM_012119	CCRK	-2,18
NM_017525	HSMDPKIN	-2,17
NM_007199	IRAK3	-2,16
NM_001715	BLK	-2,15
NM_001278	CHUK	-2,14
NM_000293	PHKB	-2,11
NM_002611	PDK2	-2,09
NM_017771	PXK	-2,08
NM_005399	PRKAB2	-2,02
NM_021923	FGFRL1	-2,01
NM_004717	DGKI	-2,00
NM_005027	PIK3R2	2,02
NM_021972	SPHK1	2,12
NM_001100594	SNRK	2,20
NM_006218	PIK3CA	2,26
NM_175886	PRPS1L1	2,26

GenBank Accession	Gene Symbol	Z-score
NM_005975	PTK6	2,27
NM_033487	CDC2L1	2,51
NM_001007792	NTRK1	2,62
NM_001081562	DMPK	2,69
NM_001006665	RPS6KA1	2,72
NM_016308	UMP-CMPK	2,77
NM_002944	ROS1	3,31
NM_004336	BUB1	3,50
NM_182493	MLCK	3,58
NM_001006943	EPHA8	3,70
NM_052853	ADCK2	3,80
NM_018423	STYK1	4,02
NM_001039468	MARK2	4,46
NM_145185	MAP2K7	4,85
NM_003557	PIP5K1A	6,10
NM_017449	EPHB2	7,68

➡ **Figure 7. STRING analysis of the siKinome screening data.**
STRING analysis to identify enriched KEGG pathways using the kinases from **Figure 5D**. Displayed are the top 10 enriched KEGG pathways along with the number of genes/proteins annotated with the indicated GO terms and the false discovery rate (FDR) of the enrichment (left panel). Involvement of individual proteins is overlaid on the 'neurotrophin signaling pathway' KEGG pathway retrieved from the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/). Proteins in grey are targeted by Dovitinib only, blue proteins were siRNA screening hits and proteins in red are both targeted by Dovitinib and silencing of these genes affected the *Mtb* bacterial load.

#pathway ID	pathway description	gene count	FDR
4722	Neurotrophin signaling pathway	15	1.95e-19
4014	Ras signaling pathway	15	2.01e-15
4151	PI3K-Akt signaling pathway	16	3.13e-14
4664	Fc epsilon RI signaling pathway	10	1.12e-13
4015	Rap1 signaling pathway	13	3.15e-13
4510	Focal adhesion	13	3.15e-13
4660	T cell receptor signaling pathway	10	3.99e-12
4380	Osteoclast differentiation	10	3.56e-11
4062	Chemokine signaling pathway	11	4.34e-11
5200	Pathways in cancer	13	6.08e-11



Discussion

Employing chemical genetic screens complemented with newly developed computational approaches, we have identified host-directed therapy (HDT) compounds and drugs (Dovitinib, AT9283 and ENMD-2076) that target human RTK signaling to control intracellular *Mtb* survival, including MDR-*Mtb*. Perhaps more importantly, our findings pave the way towards identifying additional compounds targeting human RTK signaling to improve control of intracellular *Mtb* infection since all compounds were confirmed to be effective in primary human Mϕ infection models.

Current efforts to develop HDT are a topic of interest for infectious diseases and cancer (reviewed recently[38]). In order to be able to screen larger HDT-compound libraries for novel leads with activity against intracellular *Mtb* and *Stm*, we have developed a new robust and rapid fluorescence-based intracellular screening assay (**Chapter 2**). This assay allowed us to identify host-directed *Mtb*-inhibiting compounds (SU 6656, Quinacrine, SB 216763, GW5074 and Tyrphostin AG 494) and host-directed *Stm*-inhibiting compounds (Mibefradil), which performed significantly better than our best reference compound H-89, in a LOPAC library drug-repurposing screening effort. We were also able to confirm the activity of previously published HDT compounds in our screening approach (Imatinib, D4476, LY-364947, Haloperidol), lending strong plausibility and validity to our strategy.

We next developed a novel *in silico* model which was data-driven and based on known and confirmed targets from public databases, by which we could successfully predict and verify additional compounds with host-directed activity against *Mtb* (Dovitinib, AT9283 and ENMD-2076) and *Stm* (Nafoxidine and Opipramol). Using STRING network analysis we uncovered RTK signaling as a novel host pathway controlling *Mtb* intracellular survival, which is targeted by compounds identified in this study. Finally we performed an independent unbiased siRNA screen of the human kinome, which confirmed a role for RTK signaling in control of intracellular *Mtb* survival. Collectively, our results uncover new host signaling pathways as well as corresponding active chemical compounds targeting these to control intracellular bacterial infections, including MDR-TB and *Stm*.

Our LOPAC screen provides important and general proof-of-principle for drug repurposing, since we successfully identified several candidate compounds that displayed host-directed antimicrobial activity while their known targets have not previously been associated with infectious diseases. Strikingly, 4 of the 5 hit compounds that consistently outperformed H-89 in controlling *Mtb* infection are known to affect (growth factor) RTK signaling. Tyrphostin AG 494, SU 6656, SB 216763 and GW5074 are inhibitors of EGFR, SRC Family Kinases (SFKs), GSK-3 and RAF1, respectively, which are all kinases participating in RTK pathways[44-48]. In addition to compounds affecting RTK signaling we identified 3 other host-directed *Mtb*-inhibiting compounds with vastly different target specificities. Firstly, Quinacrine was originally developed as an antimalarial drug but has displayed

activity in a myriad of diseases via a wide range of targets[49]. Interestingly, reported targets of Quinacrine include AKT1 and NF-κB as well as phospholipase A2[50]. The latter is a central enzyme in the eicosanoid pathway, which was recently shown to be involved in *Mtb* control by balancing the type I interferon response[26]. Secondly, Haloperidol is an antipsychotic drug targeting dopamine receptors[51]. Importantly, Haloperidol was recently shown to affect survival of intracellular mycobacteria in a host-directed fashion[19], providing important additional and independent validation of our screening strategy and models. Finally, 3',4'-Dichlorobenzamil is an amiloride-analogue Na+/Ca2+ exchanger inhibitor[52]. This compound may act by inhibiting Ca2+ transport in the cell, as activation of calcineurin by increased Ca2+ levels has previously been proposed as a mechanism for inhibition of phagosome maturation in *Mtb*-infected cells[53].

A similar LOPAC library screen in the HeLa-*Stm* infection model resulted in 4 compounds that more strongly reduced the bacterial load than our reference compound H-89, and Mibefradil was further confirmed to surpass H-89's activity in a rescreen. However, H-89 is already a highly potent host-directed inhibitor of *Stm* and all 4 compounds consistently and significantly reduced the *Stm* bacterial load. Two of the 4 hit compounds from the primary screen were known antibiotics (Trimethoprim and Ofloxacin) but these were tested nevertheless in our screen because they were part of the LOPAC. Of the remaining 2 HDT compounds, Haloperidol, which was already found in a previous HDT screen study in TB, was confirmed as a HDT compound with activity against *Mtb*, but we extend these results here to *Stm* as well. These data again show the validity of our screening and prediction approach since we are able to consistently and faithfully confirm already available knowledge. The data on Haloperidol also suggest that this compound may be applicable for HDT in a broader spectrum of intracellular bacterial infections. The second hit compound, Mibefradil, is a Ca2+ channel blocker[54]. The majority of screening hits in the HeLa-*Stm* infection model exacerbated bacterial loads and even though these compounds can therefore not be used for drug repurposing, all of the identified compounds may be important starting points for gaining deeper mechanistic insight into *Stm*-host interactions. The limited overlap between the hit compounds from *Mtb* and *Stm* screens likely reflects the vastly different intracellular 'lifestyles' of these pathogens. Notwithstanding this, several compounds display consistent intracellular antimicrobial activity in both *Mtb* and *Stm* infection models, such as Haloperidol. These compounds are therefore promising candidate drugs with wider application against (antibiotic resistant) intracellular bacterial infections.

Selecting hits for follow-up analysis in large (chemical) screens poses substantial challenges. Here, we employed two complementary strategies for screening follow-up. Firstly, as we aimed to identify compounds with superior host-directed antimicrobial activity, we focused on compounds performing better than the reference compound H-89, resulting in a strictly data-driven hit cut-off. Using this strategy we identified SU 6656, Quinacrine, SB 216763, GW5074 and Tyrphostin AG 494 as the most promising candidate compounds for TB and Mibefradil for salmonellosis, as well as confirmed Haloperidol as an attractive drug for HDT against both *Mtb* and *Stm*. Secondly, as screening outcome may be strongly influenced by compound properties such as solubility, hydrophobicity,

concentration, IC50 and target selectivity, using a strict cut-off may mask valuable data hidden in the large dataset and will be lost to follow-up. We therefore used a complementary follow-up approach by developing an innovative *in silico* compound predictive model to uncover relevant chemical compound classes and target profiles in screening data. Focusing on confirmed target profiles by automated extraction of bioassay data from PubChem we were able to both discern compound targets and predict novel active compounds. As the target profiles were ranked without using a hit cut-off, this approach enabled unbiased validation and follow-up of the primary chemical compound screen. The use of simple numerical values as predictive parameters renders this prediction model highly adaptable and easily applicable to other chemical screens. The model significantly enhanced the identification of de novo hit compounds (55.6% and 50% for *Mtb* and *Stm*, respectively) compared to random LOPAC library primary screening (10% for *Mtb* and 14.6% for *Stm*). Remarkably, the predicted *Mtb* hit compounds AT9283, ENMD-2076 and Dovitinib were all (growth factor) RTK inhibitors[40-42].

As inhibitors of RTK signaling molecules were already observed to be over-represented in the hits from our drug-repurposing screen, our predictive model successfully provided an unbiased validation of this observation and prompted us to further focus our screening endeavor on RTK inhibitors. RTK inhibitors are widely studied in cancer research for their anti-neoplastic properties[55]. Phase II clinical trials have been performed with both AT9283 and ENMD-2076 and Dovitinib has already passed phase III clinical trials[56-61] (http://www.clinicaltrials.gov), enabling swift future drug repurposing as host-directed antimicrobials. Our unbiased siRNA screen of the human kinome independently identified and validated RTK signaling as a host pathway regulating *Mtb* survival, identifying BLK, ABL1 and NTRK1 as host kinases controlling intracellular *Mtb* and possible drugable targets. BLK is an SFK involved in B-cell receptor signaling and the insulin response to glucose uptake in pancreatic islet cells[62,63]. The non-receptor tyrosine kinase ABL1 was previously linked to mycobacterial infection and its commonly used inhibitor Imatinib was shown to exert host-directed *Mtb* inhibiting activity *in vivo*[15,21], providing independent validation of our siRNA screening. Finally, NTRK1 is an RTK involved in peripheral nervous system development and synaptic function and plasticity[64]. Various cells of the hematopoietic lineage have been shown to produce the NTRK ligand nerve growth factor during inflammation and autoimmunity[65] and expression of NTRKs in monocytes has been previously reported[66]. Next to the confirmation of these compound targets by genetic silencing as described here, there were other siRNA hits involved in RTK signaling which might represent as yet unknown molecular targets for these or other hit compounds. Conversely, confirmed compound targets that were not identified in our siRNA screen may still contribute to *Mtb* control due to redundancy and possible incomplete genetic knockdown inherent to siRNA screens.

A role for growth factors in mycobacterial infection has been previously reported. The growth factor VEGF was linked to mycobacterial infection in a zebrafish Mycobacterium marinum (*Mm*) infection study[28] as well as in a rabbit *Mtb* infection model[27]. However, in both studies the reported effect of VEGF was

primarily systemic rather than (sub)cellular, inducing enhanced angiogenesis in granulomas. Our data strongly suggest that an intracellular response to growth factor receptor signaling via RTKs may be another important determinant for mycobacterial infection outcome. Interestingly, Oehlers *et al*. used Pazopanib, one of the compounds identified by our predictive model to show an inhibitory effect of VEGF receptor (VEGFR) inhibition on vascularization around nascent granulomas in their model. Though not meeting our strict hit selection criteria, Pazopanib statistically significantly (z-score -1.50) decreased *Mtb* loads in our screen (and thus in the absence of a vascular system), suggesting that cellular *Mtb* inhibition by Pazopanib might precede or complement the vascularization effects observed *in vivo* by Oehlers *et al*. Additionally, epidermal growth factor receptor (EGFR) signaling has previously been linked to mycobacterial infection through a chemical screen identifying EGFR inhibitor Gefitinib as a compound that restricts *Mtb* growth[31]. Our study significantly expands this knowledge by introducing additional RTK-targeting compounds that can be used for drug repurposing, including compounds targeting VEGFR (Dovitinib) and EGFR (Tyrphostin AG 494) signaling.

Our *in silico* predictive model successfully identified two compounds (Nafoxidine, an estrogen receptor modulator and Opipramol, a Sigma receptor agonist) with host-directed *Stm*-inhibiting activity. Interestingly, Haloperidol (a hit in both the *Mtb* and *Stm* LOPAC screens) was previously reported to interact with Sigma receptors with high affinity[67], suggesting mechanistic involvement of Sigma receptors in host control of intracellular bacteria.

In conclusion, the results from our chemical genetic and novel bioinformatics approach provide an important proof-of-concept of HDT for intracellular infections, such as (MDR) TB and salmonellosis. Moreover, our results identify human RTK signaling as a signaling pathway targetable by novel repurposable drugs, providing a new and promising therapeutic starting point for drug development against *Mtb*, including MDR-*Mtb*.

Experimental procedures

Reagents

H-89 dihydrochloride, DAPH 2, Nafoxidine hydrochloride, 1,3-Di-o-tolylguanidine, Naftifine hydrochloride, Opipramol, Rifampicin, Kanamycin and the Library of Pharmacologically Active Compounds (LOPAC) were purchased from Sigma-Aldrich, Zwijndrecht, The Netherlands. Hygromycin B was acquired from Life Technologies-Invitrogen, Bleiswijk, The Netherlands. VEGFR2 Kinase Inhibitor I and Ampicillin were purchased from Calbiochem Merck-Millipore, Darmstadt, Germany. Pazopanib HCl, AT9283 and Linifanib (ABT-869) were acquired from Selleck Chemicals, Munich, Germany. Quizartinib was purchased from MedChemExpress, Stockholm, Sweden. Santa Cruz BioTechnology, Heidelberg, Germany was the supplier of PDGFR Tyrosine Kinase Inhibitor III. Dovitinib

(TKI-258, CHIR-258) was from APExBIO, Houston, TX, USA. The siKinome library was acquired from Thermo Fisher Dharmacon, Waltham Massachusetts, USA.

Cell culture

HeLa cells and the MelJuSo human melanoma cell line were maintained at 37°C and 5% CO_2 in Gibco Iscove's Modified Dulbecco's Medium (IMDM; Life Technologies-Invitrogen) with 10% fetal bovine serum (FBS, Greiner Bio-One, Alphen a/d Rijn, The Netherlands), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Life Technologies-Invitrogen). Pro-inflammatory Mϕ1s and anti-inflammatory Mϕ2s were generated from monocytes isolated from whole blood of healthy donors by FICOLL separation and CD14 MACS sorting (Miltenyi Biotec, Teterow, Germany) followed by 6 days differentiation with 5 ng/ml granulocyte macrophage-colony stimulating factor (GM-CSF; BioSource Life Technologies-Invitrogen) or 50 ng/ml macrophage-colony stimulating factor (M-CSF; R&D Systems, Abingdon, United Kingdom) respectively, as previously reported[68]. Mϕs were cultured in Gibco Roswell Park Memorial Institute (RPMI) 1640 medium (Life Technologies-Invitrogen) with 10% FBS and 2 mM L-Alanyl-L-Glutamine (PAA, Linz, Austria).

Bacterial culture

Bacterial strains used are displayed in **Table 8**. Mycobacteria were cultured in Difco Middlebrook 7H9 broth (Becton Dickinson, Breda, The Netherlands) supplemented with 10% ADC (Becton Dickinson), 0.5% Tween-80 (Sigma-Aldrich) and appropriate antibiotics. *Stm* was cultured on Difco Luria-Bertani (LB) agar (Becton Dickinson) or in Difco LB broth (Becton Dickinson) supplemented with appropriate antibiotics.

Table 8. Bacterial strains, plasmids used for fluorescent protein expression and their respective antibiotic selection markers.

Base strain	Plasmid	Antibiotic resistance (source, concentration)
Stm SL1344.	pMW211[C.10E/DsRed] (Constitutive promoter).	Ampicillin (plasmid, 100 µg/ml).
Mtb H37Rv.	pSMT3[Phsp60/DsRed].	Hygromycin (plasmid, 50 µg/ml).
Mtb H37Rv.	pSMT3[Phsp60/ destabilized DsRed].	Hygromycin (plasmid, 50 µg/ml).
MDR *Mtb* Beijing family China (Kremer 43) 16319	None.	Rifampicin, Isoniazid, Ethambutol, Pyrazinamide (intrinsic, n/a).
MDR *Mtb* Dutch outbreak 2003-1128.	None.	Rifampicin, Isoniazid, Streptomycin, Claritromycin (intrinsic, n/a).

Stm and *Mtb* infections

One day before infection, mycobacterial cultures were diluted to a density corresponding with early log phase growth (optical density at 600 nm (OD_{600}) of 0.4). *Stm* was grown either in LB broth or on LB agar with appropriate antibiotics. After overnight incubation *Stm* liquid cultures were diluted 1:33 and cultured for an additional 3-4 hours while plate grown *Stm* was suspended in PBS by rinsing the agar plates. Bacterial density was determined by measuring the OD_{600} and the bacterial suspension was diluted in cell culture medium without antibiotics to reach a multiplicity of infection (MOI) of 10 (unless indicated otherwise). Accuracy of bacterial density measurements was verified by a standard colony forming unit (CFU) assay. Cell cultures (HeLa for *Stm* infections and MelJuSo for *Mtb* infections), seeded in 96-well flat-bottom plates as described below, were inoculated with 100 µl of the bacterial suspension, centrifuged for 3 minutes at 800 rpm and incubated at 37°C/5% CO_2 for 20 minutes if infected with *Stm* or 60 minutes if infected with *Mtb*. Plates were then washed with culture medium containing 30 µg/ml gentamicin sulfate (Lonza BioWhittaker, Basel, Switzerland) and incubated at 37°C and 5% CO_2 in medium containing 5 µg/ml gentamicin and indicated chemical compounds until readout by flow cytometry or CFU, as indicated.

Chemical compound treatment

10,000 HeLa or MelJuSo cells were seeded per well in 96-well flat-bottom plates or 300,000 primary Mϕs were seeded per well in 24-well plates in appropriate culture medium without antibiotics one day prior to infection with *Mtb* or broth-grown *Stm*. Infected cells were treated overnight with chemical compounds at 10 µM (unless indicated otherwise) or DMSO at equal v/v in medium containing 5 µg/ml gentamicin.

siRNA transfections

3,000 HeLa or MelJuSo cells were reverse-transfected with ON-TARGETplus siRNA pools (Thermo Fisher Dharmacon, Waltham Massachusetts, USA) at a 50 nM concentration using 0.2 µl Dharmafect1 (Thermo Fisher Dharmacon) per well in a flat-bottom 96-well plate in appropriate culture medium without antibiotics. Cells transfected with siRNA were infected with *Mtb* at MOI 1000 24 hours post transfection and incubated for an additional 48 hours and infections with agar-grown *Stm* were carried out at MOI 500 72 hours post transfection and incubated overnight, unless indicated otherwise.

Colony forming unit assay

CFU assays were performed using the track dilution method described previously[69]. In short, bacterial suspensions were serially diluted and 10 µl drops were plated on square agar plates, which were subsequently placed at an angle to allow the drops to spread over a larger surface area.

Bacterial growth assay

100 µl *Stm* or *Mtb* culture (OD_{600} of 0.1) was plated in a flat-bottom 96-well plate containing 100 µl of indicated chemical compounds at 20 µM in LB (*Stm*) or 7H9 (*Mtb*) broth. The plate was incubated at 37°C overnight for *Stm* or during a period of 15 days for *Mtb* and absorbance was measured at a 550 nm wavelength on a Mithras LB 940 plate reader (Berthold Technologies, Bad Wildbad, Germany).

Compound identification within the PubChem repository and retrieval of BioAssay data

Structure-data format (SDF) data supplied with the LOPAC library was converted to InChIKey using the OpenBabel toolbox (http://www.openbabel.org). InChIKeys were subsequently mapped to PubChem IDs and correct identification was checked manually. Compounds were manually linked to PubChem IDs if InChIKey information was insufficient for automated identification. For each of the identified compounds, BioAssay data was retrieved from the PubChem repository (as of July 25th, 2014). Human protein targets for which compounds were confirmed to be active were then extracted from the BioAssay data. Compounds were subsequently described with their confirmed protein targets, as well as z-scores for bacterial load and cell viability from the primary screening data. All remaining compounds in the PubChem repository that were not included in the LOPAC library were described with their confirmed protein targets as above. Compounds that were not confirmed to target any of the protein targets identified for the LOPAC compounds were excluded from analysis and the remaining compounds were used as a testing set for the predictive model.

Predictive model

Using LOPAC compounds as a training set, BioAssay data obtained from PubChem (descriptive variables) were related to the z-scores for bacterial load and cell viability from the primary screening data (target variables) using the predictive modelling approach of multi-target regression to simultaneously predict both target variables. Predictive models were constructed within the predictive clustering framework[35,36], using predictive clustering trees (PCTs) as predictive models for multi-target regression. Ensembles of predictive clustering trees were generated[37,38] using the Bagging ensemble learning method[70,71] as implemented in the data mining tool CLUS (http://clus.sourceforge.net). Multiple predictive models were constructed using different bootstrap samples of the training dataset and their predictions were averaged to obtain an overall prediction. The variance of the predictions for the two target variables across the models in the ensemble was calculated for each target variable separately, averaged between the two targets and then used as a reliability estimation score[72].

STRING analysis

Protein interaction networks were generated using STRING version 10 (http://string-db.org/)[73] using experiments and databases as data sources and a minimal confidence score of 0.4.

Statistics

Student's T-test, one-way ANOVA and linear regression were performed using GraphPad Prism version 6.0 for Mac OS X (GraphPad Software, San Diego California, USA; www.graphpad.com). Z' factors were calculated using the formula $Z' = \frac{\left(AVG_{DMSO} - \frac{3SD_{DMSO}}{\sqrt{n}}\right) - \left(AVG_{H\text{-}89} + \frac{3SD_{H\text{-}89}}{\sqrt{n}}\right)}{AVG_{DMSO} - AVG_{H\text{-}89}}$, where AVG is the average percentage of DsRed positive events measured after DMSO or H-89 treatment, SD is the standard deviation of these measurements and n is the number of replicates (as in **Chapter 2**). Z-scores were calculated using the formula $z = \frac{x - AVG_{DMSO}}{STDEV_{DMSO}}$, where the difference between the percentage of DsRed positive events (bacterial load) or the total event count (cell viability) of a single replicate of an experimental condition (x) and the average percentage of DsRed positive events or the total event count of the DMSO control (AVG_{DMSO}) is divided by the standard deviation of the DMSO control ($STDEV_{DMSO}$). Z-scores for the primary screens were calculated using a similar formula, where the average percentage of fluorescent events and the standard deviation of all samples on each plate (instead of the DMSO control) were used (to provide plate normalization). The average DMSO z-score was then subtracted from each sample. An average z-score ≤ -2 or ≥ 2 was used as a hit cut-off, unless otherwise indicated.

Data availability

The data that support the findings of this study are available from the corresponding authors upon request.

Code availability

The code of the machine learning software CLUS that was used to build the in-silico models for predicting compound activity is available for download from the SourceForge repository (at https://sourceforge.net/projects/clus/).

Supplementary Information: Development and use of an *in silico* model for predicting compound activity.

Machine learning in a nutshell.

Machine learning studies computer programs/algorithms that have the ability to learn (improve with experience) where the experience is given in the form of data examples (instances). The input to a typical machine learning algorithm is a single flat table comprising a number of records (rows) and attributes (columns). In general, each row represents an object and columns represent properties of objects{Dzeroski:2001di}. An excerpt of the data table that we used to learn a predictive model is given in **Supplementary Table 1**. Here, rows correspond to individual compounds and columns contain different properties of these compounds, including bioactivity profiles retrieved from PubChem, intracellular

Supplementary Table 1. Excerpt from the data table for the *Mtb* screen used to learn the predictive models.

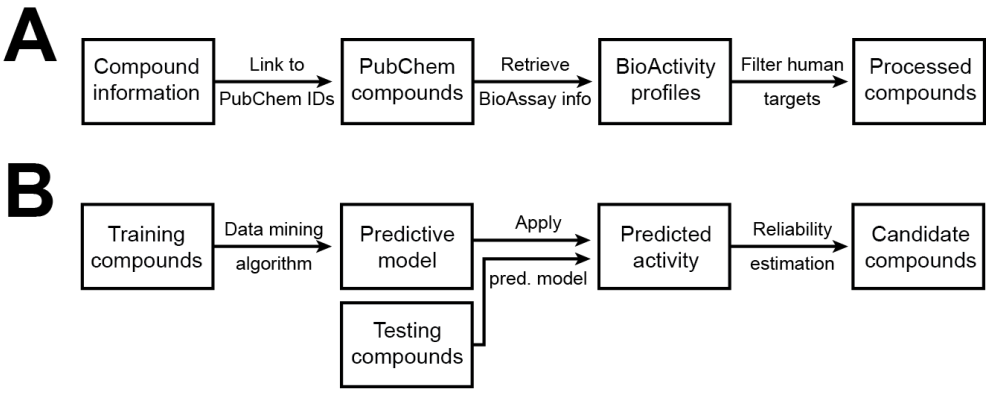
	Descriptive/Input space (PubChem BioAssay accession)					Target/Output space	
PubChem ID	gi: 10864009	gi: 10880131	gi: 10937869	gi: 112938	...	Bacterial load z-score	Cell viability z-score
ID1	1[a]	0	0	1	...	-2.61	0.29
ID2	0	0	0	1	...	-1.57	-0.43
ID3	0	0	0	0	...	0.47	0.22
ID4	0	0	1	0	...	-0.83	-0.13
ID5	1	1	0	0	...	-2.58	-0.53
ID6	0	1	0	0	...	1.78	0.97
...	...	...	...	...	...	...	...

[a] **'1' indicates that the compound has the corresponding protein as a confirmed target in a PubChem BioAssay.**

bacterial survival z-scores and host cell viability z-scores. The task formulated here is to predict the intracellular bacterial survival and the host cell viability z-scores for a novel compound using the information from its PubChem bioactivity profile. In machine learning terminology, this translates into a predictive modelling task (or supervised learning) where the two z-scores are called target (or output or dependent) variables/attributes and the variables describing the bioactivity profile are called descriptive (or input or independent) variables/attributes. Furthermore, considering that there are two numeric target variables, the task at hand is called multi-target regression[74]. This is illustrated in the data excerpt in **Supplementary Table 1**. The output of a data mining algorithm is typically a predictive model (or a set of predictive models) valid for the given data. The dataset used to learn the models is usually called training dataset. The model can then be applied to a different set of data, usually called testing dataset.

Data pre-processing

In this study, the training set of compounds consisted of our reference compound H-89 and the LOPAC library compounds that were screened in our HeLa-*Stm* and MelJuSo-*Mtb* infection models, while the testing set consisted of all other compounds available in the PubChem public repository. We performed separate analyses on the *Mtb* and *Stm* datasets, but the pre-processing of the data and the data analysis were performed following identical procedures. A schematic overview of the complete pre-processing pipeline is displayed in **Supplementary Figure 1A**.



⬆ Supplementary Figure 1. Data pre-processing pipeline.

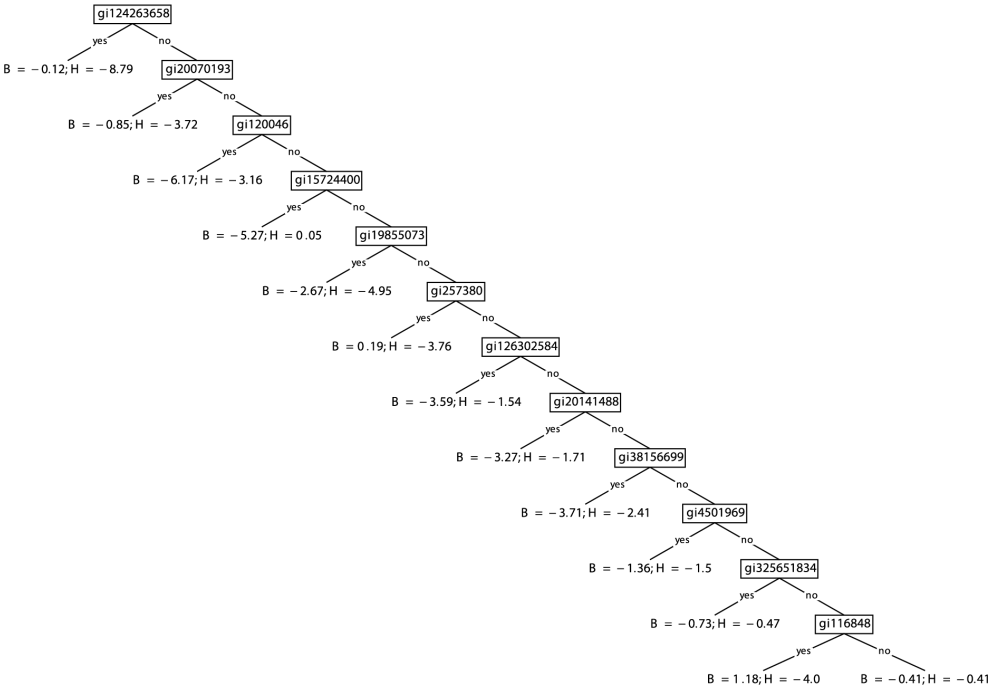
A. Pre-processing pipeline used to link compounds described by structured-data files to compounds in the PubChem database of compounds. **B.** Data analysis pipeline from the pre-processed compounds to the new candidate compounds for wet-lab experiments.

The first step of the data pre-processing was to uniquely identify the LOPAC compounds by linking them to their corresponding PubChem IDs. Based on the structure-data format (SDF) information provided by the compound supplier, we linked the LOPAC compounds to compounds from PubChem. To this end, SDF information of the compounds was first converted into InChIKey using the OpenBabel toolbox (http://www.openbabel.org) and then mapped to PubChem IDs. Next, we manually checked whether the mapping was correct and provided manual mapping where the InChIKey information was not sufficient, obtaining a list of PubChem compounds that were used in our study. Next, biological activity information was retrieved for the LOPAC compounds from each compounds' 'bioassays' section in PubChem. From the bioassays, only human protein targets for which compounds were confirmed to be active were extracted, yielding a total of 1058 protein targets. This resulted in the columns on the left-hand side of **Supplementary Table 1** (the descriptive variables). At the end of the pre-processing pipeline, each compound is described with both its protein targets (as descriptive attributes for machine learning) and experimental measurements of activity and viability (as target attributes for the machine learning). These compound descriptions comprise our training set. Finally, we considered all of the remaining compounds from PubChem as potential candidates for drug repurposing (**Supplementary Figure 1A**). We applied the pre-processing pipeline on each of these compounds as described above. Only compounds confirmed to target at least one of the 1058 human target proteins were included, thus obtaining a testing set of 460,580 compounds. Note that the compounds from the testing set have information only for the bioactivity profiles (the descriptive attributes), while the intracellular bacterial survival and host cell viability z-scores are not known but the goal is to predict these. We obtained these predictions by applying the predictive model (predictive clustering tree) learned from the training

data to each of the compounds from the testing set, as described in more detail below.

Predictive clustering trees

To analyze the data and learn a predictive model, we used the machine learning tool CLUS (available at http://clus.sourceforge.net). Specifically, we used predictive clustering trees (PCTs) for multi-target regression as models[74,75]. PCTs are a generalization of regression trees, a machine learning approach commonly used for regression. An example PCT is shown in **Supplementary Figure 2**. Similar to regression trees, PCTs are tree-like structures that have internal nodes and leaves. The internal nodes contain tests on the descriptive variables (i.e. asking whether a given protein is targeted or not), while leaves give predictions for the target variables (the predicted z-scores for intracellular bacterial survival and host cell viability). We opted to use PCTs because they are able to implicitly exploit



⬆ Supplementary Figure 2. Example predictive clustering tree.

Example predictive clustering tree (PCT) obtained from the screening data for *Mtb*. The internal nodes of the tree refer to the descriptive variables and check whether or not a compound targets a given protein. The leaves then give the predictions for the intracellular bacterial survival and the host cell viability z-scores. For example, compounds that target *gi15724400*, but not *gi14263638*, *gi20070193* or *gi120046*, are predicted to drastically reduce bacterial load (z-score of -5.27) and not affect cell viability (z-score of 0.05).

the relation between the target variables during model construction. Furthermore, PCTs are easily interpretable. A PCT can be viewed as a hierarchy of clusters with each node corresponding to a cluster. The top-node of a PCT corresponds to one cluster (group) containing all data points. This cluster is recursively partitioned into smaller clusters while moving down the tree. The leaves represent the clusters at the lowest level of the hierarchy and each leaf is labeled with its cluster's centroid/ prototype (the averages of the target variables are the prediction made by the leaf).

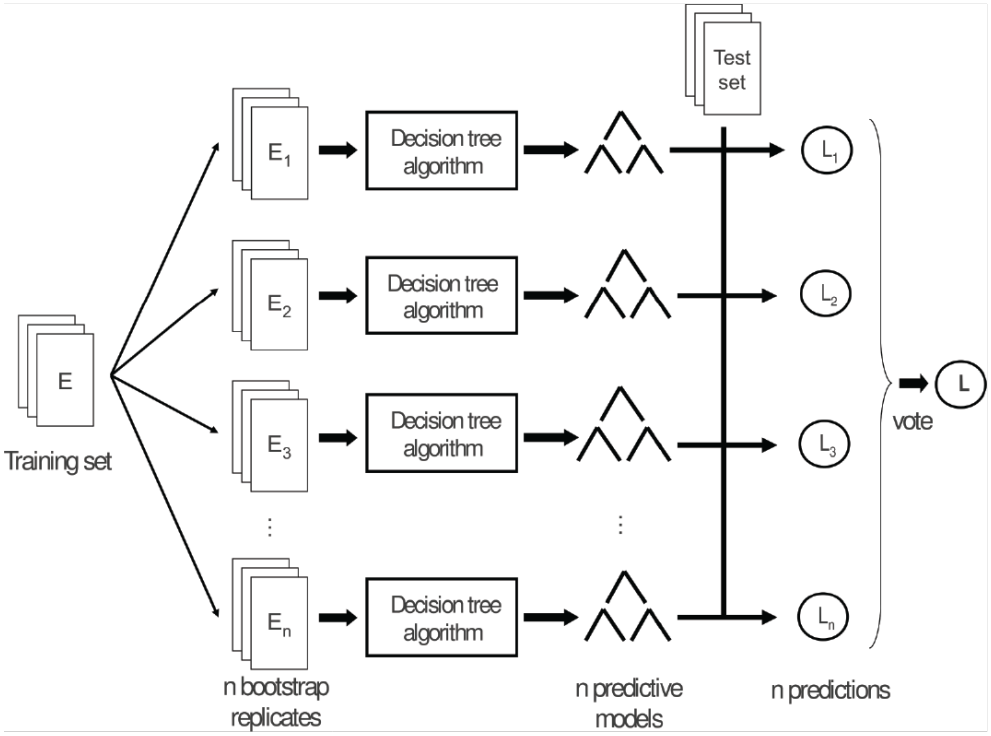
PCTs are built with a greedy recursive top-down induction algorithm. This learning algorithm starts by selecting a test for the root node by using a heuristic function computed on the training examples. The goal of the heuristic is to guide the algorithm towards small trees with good predictive performance. Based on the selected test, the training set is partitioned into subsets according to the test outcome. This is recursively repeated to construct the subtrees. The partitioning process stops when a stopping criterion is satisfied (i.e. the minimal number of examples per leaf is reached or the heuristic score no longer changes). In that case, the prototype (the prediction) is calculated as the averages of the target variables and stored in a leaf.

Ensembles of PCTs

An ensemble is a set of predictive models (called base models). The prediction of an ensemble for a new example is obtained by combining the predictions of all base models from the ensemble. These predictions can be combined by averaging them. The ensemble learning procedure is illustrated in **Supplementary Figure 3**. Here, we consider ensembles of PCTs for multi-target regression[74]. For constructing the base models, we used the Bagging method[76]. Bagging is an ensemble method that constructs the base models in the ensemble by making bootstrap samples (E_i) of the training set (also called bootstrap replicates) and using each of these replicates to construct a predictive model. Each bootstrap sample is obtained by randomly sampling training instances, with replacement, from the original training set, until an equal number of instances as in the training set is obtained.

Reliability scores

A very important aspect of using a predictive model is the ability to estimate the reliability of the predictions it makes. This reliability indicates how confident the model is about its prediction. Ensembles offer a natural way of estimating the reliability of their predictions by exploiting their voting mechanism[77]. When a prediction is made for an unlabeled example (these are examples that do not have z-score values for intracellular bacterial survival and host cell viability) by an ensemble, we consider it reliable if the predictions of the individual models in the ensemble are coherent, i.e., if the variance of the predictions is low. Here, we get the reliability score for a prediction of two targets by averaging the variances of the predictions for each of the two targets (the variances of the predicted z-scores for intracellular bacterial survival and host cell viability).



⬆ Supplementary Figure 3. Illustration of the ensemble learning method of bagging.
From the training set of examples E, n bootstrap samples are created (E_1, E_2, ..., E_n). Predictive models are then constructed (using a tree construction algorithm) on each of the n replicates. The predictions of the base predictive models (L_1, L_2... L_n) are combined by a voting (averaging) scheme into the final prediction (L) of the ensemble.

Data analysis workflow

To identify candidate compounds in the set of testing compounds to screen in our MelJuSo-*Mtb* or HeLa-*Stm* infection models, we followed the data analysis workflow outlined in **Supplementary Figure 1B**. First, we used the training dataset to construct a predictive model (a PCT) using a data-mining algorithm (the PCT algorithm). Next, the predictive model was applied to the testing set to obtain the predictions for the activity of the compounds, expressed as z-scores. Finally, we calculated a reliability score for each prediction for a test compound.

This data analysis workflow resulted in a small set of selected candidate compounds from all of the 460,580 compounds in the testing set. Predicted *Mtb* hits were defined as compounds with a predicted intracellular bacterial survival z-score below -2 and a host cell viability z-score between -1 and 1 with a prediction reliability greater than 0.5, or an intracellular bacterial survival z-score below -1.75,

a host cell viability z-score between -0.75 and 0.75 and a prediction reliability higher than 0.75. This yielded a total of 47 candidate compounds (**Table 4**). Predicted *Stm* hits were defined as compounds with a predicted intracellular bacterial survival z-score below -2, a host cell viability z-score between -1 and 1 and a prediction reliability greater than 0.5 or an intracellular bacterial survival z-score below -1.5, a host cell viability z-score between -0.75 and 0.75 and a prediction reliability higher than 0.5. This yielded a total of 30 candidate compounds (**Table 5**). From the resulting lists of predicted hits, compounds were then selected for further experiments based on their commercial availability.

References

1. Diedrich, C. R. & Flynn, J. L. HIV-1/mycobacterium tuberculosis coinfection immunology: how does HIV-1 exacerbate tuberculosis? *Infection and Immunity* **79,** 1407–1417 (2011).
2. Ottenhoff, T. H. M. The knowns and unknowns of the immunopathogenesis of tuberculosis. *Int. J. Tuberc. Lung Dis.* **16,** 1424–1432 (2012).
3. World Health Organization. *Global tuberculosis report 2015*. (2015).
4. Ottenhoff, T. H. M. Overcoming the global crisis: 'yes, we can', but also for TB ... ? *Eur. J. Immunol.* **39,** 2014–2020 (2009).
5. Jassal, M. S. & Bishai, W. R. Epidemiology and challenges to the elimination of global tuberculosis. *CLIN INFECT DIS* **50 Suppl 3,** S156–64 (2010).
6. Ottenhoff, T. H. M. New pathways of protective and pathological host defense to mycobacteria. *Trends Microbiol.* **20,** 419–428 (2012).
7. Barry, C. E. & Blanchard, J. S. The chemical biology of new drugs in the development for tuberculosis. *Curr Opin Chem Biol* **14,** 456–466 (2010).
8. Norrby, S. R., Nord, C. E., Finch, R.European Society of Clinical Microbiology and Infectious Diseases. Lack of development of new antimicrobial drugs: a potential serious threat to public health. *Lancet Infect Dis* **5,** 115–119 (2005).
9. Becker, D. *et al.* Robust Salmonella metabolism limits possibilities for new antimicrobials. **440,** 303–307 (2006).
10. Makarov, V. *et al.* Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis. *Science* **324,** 801–804 (2009).
11. Christophe, T. *et al.* High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. *PLoS Pathog* **5,** e1000645 (2009).
12. Willand, N. *et al.* Synthetic EthR inhibitors boost antituberculous activity of ethionamide. *Nat. Med.* **15,** 537–544 (2009).
13. Lawn, S. D. & Zumla, A. I. Tuberculosis. *Lancet* **378,** 57–72 (2011).
14. Guler, R. & Brombacher, F. Host-directed drug therapy for tuberculosis. *Nature Chemical Biology* **11,** 748–751 (2015).

15. Hawn, T. R., Shah, J. A. & Kalman, D. New tricks for old dogs: countering antibiotic resistance in tuberculosis with host-directed therapeutics. *Immunol. Rev.* **264,** 344–362 (2015).
16. Kuijl, C. *et al.* Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1. **450,** 725–730 (2007).
17. Kumar, D. *et al.* Genome-wide analysis of the host intracellular network that regulates survival of Mycobacterium tuberculosis. *Cell* **140,** 731–743 (2010).
18. Jayaswal, S. *et al.* Identification of host-dependent survival factors for intracellular Mycobacterium tuberculosis through an siRNA screen. *PLoS Pathog* **6,** e1000839 (2010).
19. Sundaramurthy, V. *et al.* Integration of chemical and RNAi multiparametric profiles identifies triggers of intracellular mycobacterial killing. *Cell Host and Microbe* **13,** 129–142 (2013).
20. Machado, D. *et al.* Ion Channel Blockers as Antimicrobial Agents, Efflux Inhibitors, and Enhancers of Macrophage Killing Activity against Drug Resistant Mycobacterium tuberculosis. *PLoS ONE* **11,** e0149326 (2016).
21. Napier, R. J. *et al.* Imatinib-sensitive tyrosine kinases regulate mycobacterial pathogenesis and represent therapeutic targets against tuberculosis. *Cell Host and Microbe* **10,** 475–485 (2011).
22. Subbian, S. *et al.* Phosphodiesterase-4 inhibition alters gene expression and improves isoniazid-mediated clearance of Mycobacterium tuberculosis in rabbit lungs. *PLoS Pathog* **7,** e1002262 (2011).
23. Subbian, S. *et al.* Phosphodiesterase-4 inhibition combined with isoniazid treatment of rabbits with pulmonary tuberculosis reduces macrophage activation and lung pathology. *Am. J. Pathol.* **179,** 289–301 (2011).
24. Koo, M.-S. *et al.* Phosphodiesterase 4 inhibition reduces innate immunity and improves isoniazid clearance of Mycobacterium tuberculosis in the lungs of infected mice. *PLoS ONE* **6,** e17091 (2011).
25. Vilaplana, C. *et al.* Ibuprofen therapy resulted in significantly decreased tissue bacillary loads and increased survival in a new murine experimental model of active tuberculosis. *Journal of Infectious Diseases* **208,** 199–202 (2013).
26. Mayer-Barber, K. D. *et al.* Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. **511,** 99–103 (2014).
27. Datta, M. *et al.* Anti-vascular endothelial growth factor treatment normalizes tuberculosis granuloma vasculature and improves small molecule delivery. *Proc Natl Acad Sci USA* **112,** 1827–1832 (2015).
28. Oehlers, S. H. *et al.* Interception of host angiogenic signalling limits mycobacterial growth. **517,** 612–615 (2015).
29. Schiebler, M. *et al.* Functional drug screening reveals anticonvulsants as enhancers of mTOR-independent autophagic killing of Mycobacterium tuberculosis through inositol depletion. *EMBO Molecular Medicine* **7,** 127–139 (2015).
30. Skerry, C. *et al.* Simvastatin increases the in vivo activity of the first-line tuberculosis regimen. *J. Antimicrob. Chemother.* **69,** 2453–2457 (2014).

31. Stanley, S. A. *et al.* Identification of host-targeted small molecules that restrict intracellular Mycobacterium tuberculosis growth. *PLoS Pathog* **10,** e1003946 (2014).
32. Vergne, I., Chua, J., Singh, S. B. & Deretic, V. Cell biology of mycobacterium tuberculosis phagosome. *Annu. Rev. Cell Dev. Biol.* **20,** 367–394 (2004).
33. Brumell, J. H. & Grinstein, S. Salmonella redirects phagosomal maturation. *Current Opinion in Microbiology* **7,** 78–84 (2004).
34. Singhal, A. *et al.* Metformin as adjunct antituberculosis therapy. *Science Translational Medicine* **6,** 263ra159–263ra159 (2014).
35. Coussens, A. K., Wilkinson, R. J. & Martineau, A. R. Phenylbutyrate Is Bacteriostatic against Mycobacterium tuberculosis and Regulates the Macrophage Response to Infection, Synergistically with 25-Hydroxy-Vitamin D3. *PLoS Pathog* **11,** e1005007 (2015).
36. Mily, A. *et al.* Significant Effects of Oral Phenylbutyrate and Vitamin D3 Adjunctive Therapy in Pulmonary Tuberculosis: A Randomized Controlled Trial. *PLoS ONE* **10,** e0138340 (2015).
37. Mehrotra, P. *et al.* Pathogenicity of Mycobacterium tuberculosis is expressed by regulating metabolic thresholds of the host macrophage. *PLoS Pathog* **10,** e1004265 (2014).
38. Kaufmann, S. H. E., Dorhoi, A., Hotchkiss, R. S. & Bartenschlager, R. Host-directed therapies for bacterial and viral infections. *Nat Rev Drug Discov* (2017). doi:10.1038/nrd.2017.162
39. Verreck, F. A. W. *et al.* Human IL-23-producing type 1 macrophages promote but IL-10-producing type 2 macrophages subvert immunity to (myco)bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 4560–4565 (2004).
40. Howard, S. *et al.* Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J. Med. Chem.* **52,** 379–388 (2009).
41. Tentler, J. J. *et al.* Assessment of the in vivo antitumor effects of ENMD-2076, a novel multitargeted kinase inhibitor, against primary and cell line-derived human colorectal cancer xenograft models. *Clinical Cancer Research* **16,** 2989–2998 (2010).
42. Trudel, S. *et al.* CHIR-258, a novel, multitargeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma. *Blood* **105,** 2941–2948 (2005).
43. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45,** D353–D361 (2017).
44. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–1134 (2010).
45. Levitzki, A. & Gazit, A. Tyrosine kinase inhibition: an approach to drug development. *Science* **267,** 1782–1788 (1995).
46. Varga, E. V. *et al.* Involvement of Raf-1 in chronic delta-opioid receptor agonist-mediated adenylyl cyclase superactivation. *Eur. J. Pharmacol.* **451,** 101–102 (2002).

47. Cicha, I., Zitzmann, R. & Goppelt-Struebe, M. Dual inhibition of Src family kinases and Aurora kinases by SU6656 modulates CTGF (connective tissue growth factor) expression in an ERK-dependent manner. *Int. J. Biochem. Cell Biol.* **46,** 39–48 (2014).
48. Coghlan, M. P. *et al.* Selective small molecule inhibitors of glycogen synthase kinase-3 modulate glycogen metabolism and gene transcription. *Chem. Biol.* **7,** 793–803 (2000).
49. Al-Bari, M. A. A. Chloroquine analogues in drug discovery: new directions of uses, mechanisms of actions and toxic manifestations from malaria to multifarious diseases. *J. Antimicrob. Chemother.* **70,** 1608–1621 (2015).
50. Ehsanian, R., Van Waes, C. & Feller, S. M. Beyond DNA binding - a review of the potential mechanisms mediating quinacrine's therapeutic activities in parasitic infections, inflammation, and cancers. *Cell Commun. Signal* **9,** 13 (2011).
51. Lako, I. M., van den Heuvel, E. R., Knegtering, H., Bruggeman, R. & Taxis, K. Estimating dopamine D_2 receptor occupancy for doses of 8 antipsychotics: a meta-analysis. *J Clin Psychopharmacol* **33,** 675–681 (2013).
52. Blaustein, M. P. & Lederer, W. J. Sodium/calcium exchange: its physiological implications. *Physiol. Rev.* **79,** 763–854 (1999).
53. Jayachandran, R. *et al.* Survival of mycobacteria in macrophages is mediated by coronin 1-dependent activation of calcineurin. *Cell* **130,** 37–50 (2007).
54. Osterrieder, W. & Holck, M. In vitro pharmacologic profile of Ro 40-5967, a novel Ca2+ channel blocker with potent vasodilator but weak inotropic action. *J. Cardiovasc. Pharmacol.* **13,** 754–759 (1989).
55. Gaumann, A. K. A. *et al.* Receptor tyrosine kinase inhibitors: Are they real tumor killers? *Int. J. Cancer* **138,** 540–554 (2016).
56. Hay, A. E. *et al.* A phase II study of AT9283, an aurora kinase inhibitor, in patients with relapsed or refractory multiple myeloma: NCIC clinical trials group IND.191. *Leuk. Lymphoma* **57,** 1463–1466 (2016).
57. Moreno, L. *et al.* A phase I trial of AT9283 (a selective inhibitor of aurora kinases) in children and adolescents with solid tumors: a Cancer Research UK study. *Clinical Cancer Research* **21,** 267–273 (2015).
58. Schäfer, N. *et al.* Phase I trial of dovitinib (TKI258) in recurrent glioblastoma. *J. Cancer Res. Clin. Oncol.* **142,** 1581–1589 (2016).
59. Cheng, A.-L. *et al.* Randomized, Open-Label Phase 2 Study Comparing Frontline Dovitinib vs Sorafenib in Patients With Advanced Hepatocellular Carcinoma. *Hepatology* (2016). doi:10.1002/hep.28600
60. Lim, S. H. *et al.* Efficacy and safety of dovitinib in pretreated patients with advanced squamous non-small cell lung cancer with FGFR1 amplification: A single-arm, phase 2 study. *Cancer* (2016). doi:10.1002/cncr.30135
61. Yee, K. W. L. *et al.* A phase I trial of the aurora kinase inhibitor, ENMD-2076, in patients with relapsed or refractory acute myeloid leukemia or chronic myelomonocytic leukemia. *Invest New Drugs* (2016). doi: 10.1007/s10637-016-0375-2

62. Gauld, S. B. & Cambier, J. C. Src-family kinases in B-cell development and signaling. *Oncogene* **23,** 8001–8006 (2004).
63. Borowiec, M. *et al.* Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc Natl Acad Sci USA* **106,** 14460–14465 (2009).
64. Skaper, S. D. The biology of neurotrophins, signalling pathways, and functional peptide mimetics of neurotrophins and their receptors. *CNS Neurol Disord Drug Targets* **7,** 46–62 (2008).
65. Aloe, L., Rocco, M. L., Bianchi, P. & Manni, L. Nerve growth factor: from the early discoveries to the potential clinical use. *J Transl Med* **10,** 239 (2012).
66. Ehrhard, P. B., Ganter, U., Stalder, A., Bauer, J. & Otten, U. Expression of functional trk protooncogene in human monocytes. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 5423–5427 (1993).
67. Maurice, T. & Su, T.-P. The pharmacology of sigma-1 receptors. *Pharmacol. Ther.* **124,** 195–206 (2009).
68. Verreck, F. A. W., de Boer, T., Langenberg, D. M. L., van der Zanden, L. & Ottenhoff, T. H. M. Phenotypic and functional profiling of human proinflammatory type-1 and anti-inflammatory type-2 macrophages in response to microbial antigens and IFN-gamma- and CD40L-mediated costimulation. *Journal of Leukocyte Biology* **79,** 285–293 (2006).
69. Jett, B. D., Hatter, K. L., Huycke, M. M. & Gilmore, M. S. Simplified agar plate method for quantifying viable bacteria. *BioTechniques* **23,** 648–650 (1997).
70. Liu, W. S. & Heckman, C. A. The sevenfold way of PKC regulation. *Cell. Signal.* **10,** 529–542 (1998).
71. Wu-zhang, A. X. & Newton, A. C. Protein kinase C pharmacology: refining the toolbox. *Biochem. J.* **452,** 195–209 (2013).
72. Le Poole, I. C. *et al.* Phagocytosis by normal human melanocytes in vitro. *Exp. Cell Res.* **205,** 388–395 (1993).
73. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–52 (2015).
74. Kocev, D., Vens, C., Struyf, J. & Džeroski, S. Tree ensembles for predicting structured outputs. *Pattern Recognition* (2013).
75. Blockeel, H., De Raedt, L. & Ramon, J. Top-down induction of clustering trees. *ICML Proceedings of the Fifteenth International Conference on Machine Learning* 55–63 (1998).
76. Breiman, L. Bagging predictors. *Machine learning* (1996).
77. Levatić, J., Ceci, M., Kocev, D. & Džeroski, S. Self-training for multi-target regression with tree ensembles. *Knowledge-Based Systems* (2017). doi: 10.1016/j.knosys.2017.02.014