

# Intracellular Drug Targets In *Mycobacterium Tuberculosis* Revealed By A Chemo-Genetic Approach

**Clement K.M. Tsui**

University of British Columbia

**Flavia Sorrentino**

University of British Columbia

**Narula Gagandeep**

University of British Columbia

**Alfonso Mendoza Losanna**

Global Health R&D, GSK, Tres Cantos

**Ruben Gonzalez Rio**

Global Health R&D, GSK, Tres Cantos

**Esther Pérez Herrán**

Global Health R&D, GSK, Tres Cantos

**Abraham Lopez**

University of British Columbia

**Adama Bojang**

University of British Columbia

**Xingji Zheng**

University of British Columbia

**Modesto Jesus Remuinan-Blanco**

Global Health R&D, GSK, Tres Cantos

**Yossef Av-Gay (✉ [yossi@mail.ubc.ca](mailto:yossi@mail.ubc.ca))**

University of British Columbia

---

## Research Article

**Keywords:** compounds, sensitivity, antimicrobial resistance, selection, chemical genetics, whole genome sequencing, bioinformatics, drug discovery

**Posted Date:** December 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-946186/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

*Mycobacterium tuberculosis* (Mtb), the etiological agent of tuberculosis, is one of the most devastating infectious agents in the world. It causes chronic lung diseases to one third of the world's population. Chemo-genetic characterization through *in vitro* evolution combined with whole genome sequencing analysis can identify novel drug targets and drug resistance genes in Mtb. We performed a genome analysis of 53 Mtb mutants resistant to 15 different hit compounds. We found nonsynonymous mutations/indels in 30 genes that may be associated with drug resistance acquisitions. Beyond confirming previously identified drug resistance mechanisms such as *rpoB* and lead targets reported in novel anti-tuberculosis drug screenings such as *mmpL3*, *ethA*, *mbtA*, we discovered several unrecognized candidate drug targets including *prfB* and *TB18.5*. The exploration of the *M. tuberculosis* chemical mutant genomes could help novel drug discovery and structural biology of compounds and associated mechanisms of action relevant to tuberculosis treatment.

## Introduction

*Mycobacterium tuberculosis* (Mtb), the etiological agent of tuberculosis, is one of the most devastating infectious agents in the world <sup>1</sup>. One third of the world's population is exposed to Mtb, and nearly two million people were killed annually. In 2018, about 1.5 million people died from the disease, and 10 million people developed the illness (<http://www.who.int/en/news-room/fact-sheets/detail/tuberculosis>). TB transmission is airborne, where droplets containing Mtb enter the lungs and circulating alveolar macrophages engulf the bacilli (<http://www.who.int/mediacentre/factsheets/fs104/en/>). Macrophages are key components of the human innate immune system that destroy invading microorganisms. Yet, Mtb is able to survive and persist from the macrophage's killing machinery and even replicate inside the macrophage in a specified organelle termed the phagosome. Mtb can evade the host immune system, and are protected from many antibiotics that fail to reach the phagosome <sup>2</sup>.

There are several first- and second-line anti TB drugs, and the treatment involves a regime of four drugs, isoniazid, rifampin, pyrazinamide, and ethambutol taken daily for 6 to 9 months, a far longer treatment than for most bacterial infections. With the increasing prevalence of multi- and extremely- drug-resistant tuberculosis, treatment of patients often involves the use of more expensive second-line drugs and requires over 24 months. A few candidate drugs and hit compounds have been discovered in the last 2 decades but only two drugs, bedaquiline and pretomanid, have been FDA approved in the past forty years <sup>3</sup>, plus there is an urgent need to combat these new Mtb resistant strains by refueling drug development pipeline with novel drug discovery approaches.

One of the challenges in TB drug discovery is the lack of successful transfer from compounds with *in vitro* activity to efficacy in the clinical settings. For instance, compounds may be selected that are only active *in vitro* conditions; the targets which are inhibited and identified in *in vitro* liquid culture during cell-based phenotypic screens may not be essential *in vivo* <sup>4</sup>. This can be achieved through a more global understanding of the host-Mtb interaction using a chemical-genetic approach. We have developed an

advanced intracellular drug-screening assay to screen compounds in infected macrophages<sup>5,6</sup>. Using our approach, we have further screened two libraries and identified a set of diverse chemical entities that are highly effective against Mtb within the human macrophage with marked intracellular selectivity. Yet, the mode of action (MOA) in these hit compounds and the potential drug targets/inhibitors have not been fully elucidated.

Whole-genome sequencing (WGS) technology followed by bioinformatics analysis has been effective to investigate the epidemiology and transmission of Mtb in outbreak investigation and for infection control<sup>1,7-9</sup>. The technology and variant calling pipeline are also useful in characterizing the genetic polymorphisms and mechanism of resistance in drug resistance clinical strains<sup>10-14</sup>. Since the lack of understanding of compound mode of action (MOA) has become the major barrier to the development of potential novel drug to treatment tuberculosis, the WGS technology has been extended to identify and characterize the candidate drug targets in novel TB drug discovery program; the genomes of spontaneous drug-resistant mutants of *M. bovis* BCG or *M. smegmatis* from screening have been characterized followed by variants annotation as well as target identification and validation<sup>15-18</sup>.

In this study, we seek to identify novel/candidate drug targets/hit compound mode of action and further understand the mechanisms of resistance to various chemical entities. We took advantage of the hit compound libraries and the intracellular drug screening assays. We postulated the MOA of selected hit compounds based upon their chemical properties. We further generated and sequenced the genomes of 53 resistant mutants of *M. tuberculosis* H37Rv on media containing 2-5 · MIC90 of various hit compounds to assist identification of their corresponding MOA. The identified mutations lead us to identification of novel MOA candidate proteins that include possible drug targets which are critically important to identifying new antibiotics for the long-term control of TB disease.

## Results And Discussion

### Identification of hit compounds

Previous screening of GSK-proprietary libraries had identified a set of diverse chemical entities. High-throughput Screening (HTS) was performed using 5µM single shot and 1-10µM dose response for 84,000 compounds from predefined in-house GSK libraries; first, the “TB box”, having 11,000 compounds coming out from *in vitro* phenotypic screening of 2,000,000 compounds against *M. bovis* BCG with hit confirmation in Mtb and second, a library of 73,000 compounds drawn from GSKChem with “ideal” medicinal chemistry characteristics termed “Small&Beautiful”.

Data has been analysed and intracellular MIC50 and MIC90 have been extrapolated for all compounds tested. 523 hit compounds belonged to the GSK “TB box”, with intracellular MIC90 <3 µM, and 31 hit compounds belonged to the Small&Beautiful, with MIC90 <10 µM. A total of 564 hits from the HTS campaign were identified in both aforementioned libraries (unpublished data).

After removal of duplicates, compounds with known MOA, and compounds cytotoxic in HepG2 cells, 265 hit compounds were identified that were active at low micro-molar concentration. The majority of hits were more active intracellularly with 85 compounds (about 30%) shown to be active only in an intracellular assay (unpublished data). Structure-based hit clustering identified 85 hit compounds to known chemical structures, which belonged to 19 previously identified compounds with either known activity, MOA or targets in Mtb (Fig. 1). These include 18 compounds that are Dihydrofolate reductase (DHFR) like inhibitors, 11 Inosine 5'-monophosphate (IMP), 10 N-benzyl-6',7'-dihydrospiro[piperidine-4,4'-thieno[3,2-c]pyrans] (SPIRO), 10 tetrahydropyrazolo[1,5-a]pyrimidine-3-carboxamides (THPP) and 8 were previously identified as mmpL3 inhibitors; other targets include Mtb gyrase inhibitors (MGI) and oxazolidinone. Based on the MOA or target in Mtb, 12 of these chemical entities are targeting DNA gyrase and 11 targeting Inosine 5'-monophosphate (IMP) (Fig. 1).

Rationalizing that it would be difficult to obtain mutants for intracellular Mtb, we took advantage of the observation that various carbon sources can mimic the intracellular environment within the macrophages. The 85 hit compounds selected were subject to *in vitro* MIC in different carbon sources (glucose, cholesterol and acetate). Twenty-seven compounds (34%) of compounds tested had MIC90 in cholesterol. Most of the compounds active in cholesterol were also active in acetate<sup>6</sup>.

## **Generation of Mutants, Mutant characterization and whole genome sequencing**

Out of these 85 hit compounds, the properties of 16 compounds were summarized in Table 1. These compounds had demonstrated potent activities against Mtb in defined carbon source media.<sup>5,6</sup> Although some compounds showed antitubercular activity based on structural analysis, the targets/MOA of some compounds has been unknown or undefined.

Table 1  
 Characteristics of the chemical compounds and conditions for mutant screening.

ID	Molecular Weight	Intracellular MIC90 (uM)	Resistant mutant selection media	# of resistant colonies
213A	365	0.6	5X MIC Glucose	5
267A	433	0.16	2X MIC90 ADC	3
290A	476	0.6	2X MIC90 ADS	2
950A	436	1.26	2X MIC90 ADC	4
739A	381	2	2X MIC90 Acetate	3
472A	355	2	2X MIC90 Acetate	4
412A	380	0.32	2X MIC90 ADC	3
412A	380	0.32	2X MIC90 Glucose	6
296A	420	1	2X MIC90 Acetate	6
648X	386	1.58	2X MIC90 ADC	2
454A	496	2	2X MIC90 ADC	3
1114A	347	2.51	5X MIC90 Glucose	2
486X	220	0.16	5X MIC90 ADC	3
912A	394	1.58	2X MIC90 Acetate	4
622A	288	2	5X MIC90 Glucose	3
705A	310.8	1.26	2X MIC Acetate	1

Drug resistance in Mtb is always the result of mutations in existing genes that are passed along through vertical descent, that is, passed from parental (mother) to mutants (daughter cells). Following phenotypic screening, about 3-6 resistant colonies were obtained for each compound. 1-6 resistant mutants were selected for whole genome sequencing, which allowed genetic polymorphism comparison among different colonies. The laboratory parental strain H37Rv was also sequenced to exclude SNPs derived from in-house passage in comparison to the standard reference genome (NC\_000962.3). For each mutant, 151782 - 2071066 reads were generated; the depth of coverage varied among samples and ranged from 14 to 186 (average coverage =65x) (supple. Table 1). Over 95% of the trimmed reads were mapped to the reference genome in NCBI showing a very low level of contaminants (supple. Table 1). While two mutants 648X-1 and 648X-2 contained reads matching to *Alteromonas macleodii* and other bacteria, these contaminated reads were excluded before mapping to the reference genome. The size of the Mtb assembled genome was ca. 4.4 Mb in agreement with reference genome size.

# Identification of candidate intracellular drug targets in resistant mutants

We discovered 74 nonsynonymous SNPs and 13 indels (insertions or deletions) located in 30 different genes that arose in 53 *Mtb* mutant genomes (Table 2). Most polymorphisms (28) localized in genes involved in cell wall and cell processes, followed by intermediary metabolism and respiration (16) and regulatory proteins (14) (Fig. 2).

Table 2

List of SNPs and indels in various genes recovered from various mutants after WGS. Gene names and features were displayed according to Mycobrowser annotations (<https://mycobrowser.epfl.ch/>).

Compound	# Mutants	Gene/ ORFs	Genetic Polymorphisms (frequency)	Relevant Codon change (frequency)	Product
213A	5	<i>mmpL3</i>	755A>G; 758G>A; 875T>C; 1985C>A; 2051T>C	Tyr252Cys; Gly253Glu; Ile292Thr; Ala662Glu; Val684Ala	Conserved membrane transport protein
		<i>sugI</i>	16C>T	Gln6*, gained stop codon (2)	Involved in transport of sugar across the membrane. Responsible for the translocation of the substrate across the membrane.
267A	3	<i>mmpL3</i>	763T>C, 765C>G,  1932 C>A	Phe255Leu (2) ; Phe644Leu	Conserved membrane transport protein
290A	2	<i>mmpL3</i>	1909C>A	Leu637Ile (2)	Conserved membrane transport protein
		<i>Rv0370c</i>	474G>T	Val158Val (2)	Unknown. possible Oxidoreductase
950A	4	<i>rpoB</i>	1720G>A	Asp574Asn (4)	Transcription of DNA into RNA
		<i>Rv3629</i>	641G>A	Gly214Glu (4)	Probable conserved integral membrane protein
739A	3	<i>ethA</i>	611T>C	Met204Thr (2)	Monooxygenase that activates the pro-drug ethionamide (ETH)
		<i>rpsO</i>	157C>T	Arg53Trp	30S ribosomal protein S15
		<i>Rv1024</i>	154C>T	Pro52Ser	Possible conserved membrane protein
		<i>Rv3220c</i>	746C>CA (2)	indels, frameshift variant	Probable two component sensor kinase
472A	4	<i>ethA</i>	611T>C	Met204Thr (3)	Monooxygenase that activates the pro-drug ethionamide (ETH)

Compound	# Mutants	Gene/ ORFs	Genetic Polymorphisms (frequency)	Relevant Codon change (frequency)	Product
		<i>rpsO</i>	157C>T	Arg53Trp	30S ribosomal protein S15
		<i>Rv3220c</i>	746C>CA (3),	indels, frameshift variant	Probable two component sensor kinase
412A	9	<i>prpB</i>	452T>C (7); 548C>T; 875A>G;	Leu151Pro (7); Thr183Ile; Gln292Arg;	Two component regulatory system PRRB/PRRA
		<i>moaC3</i>	392A>G	Asp131Gly	Probable molybdenum cofactor biosynthesis protein
		<i>iniB</i>	290C>T	Thr97Ile	Isoniazid inducible gene protein.
		<i>narL</i>	298G>C	Ala100Pro	Possible nitrate/nitrite response transcriptional regulatory protein
		<i>ctpl</i>	3113TGCGAG>T	indels, frameshift variant	Probable cation-transporter ATPase I
296A	6	<i>TB18.5</i>	145G>C; 236A>G; 243C>A; 277A>G	Val49Leu (2); Tyr79Cys; His81Gln; Thr93Ala (2)	Conserved protein
		<i>Rv1948</i>	122C>A	Ala41Glu	Hypothetical protein
		<i>pks6</i>	667G>A	Val223Ile (2)	Probable membrane bound polyketide synthase
648X	2	<i>ctpC</i>	1511C>T	Ser504Phe	Probable metal cation-transporting P-type ATPase C
454A	3	<i>Rv0678</i>	466 G>GC (3)	indels, frameshift variation	Conserved protein
		<i>mbtA</i>	1369CT>C (3)	Indels, frameshift variation	Bifunctional enzyme MbtA: salicyl-AMP ligase (SAL-AMP ligase) + salicyl-S-ArCP synthetase
1114A	2	<i>dnaE1</i>	2215A>G	Met739Val	Probable DNA polymerase III (alpha chain) DnaE1 (DNA nucleotidyltransferase)

Compound	# Mutants	Gene/ ORFs	Genetic Polymorphisms (frequency)	Relevant Codon change (frequency)	Product
		<i>virS</i>	983C>A	Pro328His	Virulence-regulating transcriptional regulator VirS (AraC/XylS family)
		<i>Rv0585c</i>	1202A>C	Asp401Ala	conserved integral membrane protein
		<i>sugI</i>	16C>T	Gln6* (2), gained stop codon	Involved in transport of sugar across the membrane. Responsible for the translocation of the substrate across the membrane.
486X	3	<i>phoR</i>	661G>C	Ala221Pro,	Possible two component system response sensor kinase membrane associated PhoR
		<i>fbiC</i>	1082C>A	Thr361Lys,	Probable F420 biosynthesis protein FbiC
		<i>fbiA</i>	866T>A	Leu289Gln (2),	Probable F420 biosynthesis protein FbiA
		<i>Rv3327</i>	296C>G	Pro100Ala	Probable transposase fusion protein
912A	3	<i>ethA</i>	205T>C; 190T>C	Trp69Arg; Phe64Ile	Monooxygenase that activates the pro-drug ethionamide (ETH)
		<i>Rv2542</i>	1042G>A	Ala348Thr (2)	Conserved hypothetical protein
		<i>Rv3083</i>	783G>A; 806T>C	Trp261*, gained stop codon; Leu269Pro	Probable monooxygenase (hydroxylase)
622A	3	<i>ftsK</i>	1192T>C	Ser398Pro,	Possible cell division transmembrane protein
		<i>virS</i>	712G>T	Val238Phe,	Virulence-regulating transcriptional regulator VirS (AraC/XylS family)

Compound	# Mutants	Gene/ ORFs	Genetic Polymorphisms (frequency)	Relevant Codon change (frequency)	Product
		<i>sugI</i>	16C>T	Gln6* (3), gained stop codon	Involved in transport of sugar across the membrane. Responsible for the translocation of the substrate across the membrane.
705A	1	<i>Rv3083</i>	380_381 G>GA, indels, Frame shift variant	Glu127_Thr128fs,	Probable monooxygenase (hydroxylase)
		<i>sugI</i>	16C>T	Gln6*, gained stop codon	Involved in transport of sugar across the membrane. Responsible for the translocation of the substrate across the membrane.

Annotation of these mutations revealed several known genes or targets that are relevant to drug tolerance/resistance; several mutations were reported in targets of the first- or second-line drugs in treating tuberculosis. We found a mutation in *rpoB* (Asp574Asn) encoding the beta subunit of RNA polymerase in mutants to compound 950A. Mutations in *rpoB* known to confer resistance to rifampicin are commonly reported in MDR and XDR *M. tuberculosis* strains (Gygli et al. 2017); approximately 95% of rifampicin resistant clinical isolates carry a mutation in the resistance determining region in *rpoB* in *M. tuberculosis*. Alternatively, six resistant strains to three compounds (472A, 739A, 912A) were found to possess mutations (3 different nonsynonymous mutations) in *ethA* (Table 2), an FAD-containing monooxygenase, which is a mycobacterial enzyme responsible for bio-activation of ethionamide (ETH), an antibiotic prodrug in tuberculosis treatment<sup>19-22</sup>. Loss-of-function mutations in *ethA* result in ethionamide resistance<sup>20</sup>. Interestingly for 4 mutants resistant to 472A and 739A and bearing substitutions in *ethA*, they also possessed indels in *Rv3220c*, which belongs to two component regulatory system that enable the organisms to make coordinated changes in gene expression in response to various environmental stimuli<sup>23</sup>. However, *Rv3220c* did not appear to contribute to *M. tuberculosis* virulence because *Rv3220c* knock out mutant did not demonstrate any severe infection in mice compared with the H37Rv wild type in previous study<sup>23</sup>. For the two mutants resistant to 472A and 739A but without mutations/ indels in *ethA* and *Rv3220c*, mutations were observed in *rpsO*, which encodes the ribosomal protein S15 and is important in protein translation<sup>24</sup>.

Apart from *ethA*, mutants resistant to 912A and 705A had mutations or gained a stop codon in *Rv3083* (*MymA*), which also plays a role in activating ethionamide; loss of *MymA* function resulted in ethionamide resistant *Mtb*<sup>25,26</sup>. Grant et al. (2016) found that *MymA*, a Baeyer-Villiger monooxygenases (BVMO) not previously described as an activating enzyme, is required to oxidize compounds to the corresponding

sulfoxide for its replicating and non-replicating activity<sup>25</sup>; loss of *MymA* function is proposed to confer resistance comparable to loss of *ethA* function<sup>26</sup>.

Gene targets with independent mutations may confer a fitness advantage to Mtb strains in the presence of antimicrobial drugs. Importantly, we identified *mmpL3* as a target of independent mutation in mutants resistant to compound 267A, 213A<sup>27</sup> and 290A. *mmpL3* is a membrane transporter in the resistance-nodulation-cell division family and has been shown to be the target of several small molecules and antimycobacterial compounds<sup>28-31</sup>. Similarly, multiple strains resistant to 412A developed mutations in *prfB*, which belongs to a two-component regulatory system composed of PrfB histidine kinase and PrfA response regulator<sup>32-34</sup>. This gene has been shown to be critical for viability of *M. tuberculosis* cells and is required for the initial phase of macrophage infection; *prfBA* is conserved among all mycobacterial species pointing towards its critical function in mycobacterial physiology. Recently, *prfB* was also reported as the target of a hit compound Diarylthiazole<sup>33</sup>.

In addition, mutations were reported in the target of new anti-tuberculous drugs. A frameshift insertion leading to loss-of-function in Rv0678 has been observed in mutants resistant to 454A. *Rv0678* is a gene that regulate the expression of the MmpS5-MmpL5 efflux pump, of which the variants could confer resistance to bedaquiline, leading to 2- to 8- fold increases in bedaquiline MIC as well as 2- to 4-fold increases in clofazimine MIC<sup>35</sup>. They have been isolated *in vitro* upon exposure to clofazimine<sup>3</sup> or bedaquiline. Recently, cross-resistance between clofazimine (CFZ) and bedaquiline (BDQ) was shown to be due to mutations within *Rv0678*<sup>35-37</sup>, a transcriptional repressor, which results in derepression and upregulation of the multi-substrate efflux pump mmpL5. Similar genetic polymorphisms (SNPs and indels) in *Rv0678* had also been reported in resistant mutants to “compound 5” in drug target discovery study<sup>38</sup>. Interestingly, all the mutants bearing indels in *Rv0678* also had deletion in *mbtA*, which is an adenylating enzyme that catalyzes the first step in the biosynthesis of the mycobactins<sup>39</sup>. In the last decade, the siderophore biosynthesis has been pursued as a drug target to tuberculosis<sup>40</sup>. The analog, 5'-O[N-(salicyl)sulfamoyl]adenosine was shown independently by three groups to inhibit MbtA with a Ki value of ~6 nM<sup>41,42</sup>.

Our study also revealed mutations in genes that are not well investigated or unknown in function. Some of these targets could be newly discovered drug targets. For instance, independent mutations were observed in *TB18.5* in mutants to 296A. TB18.5 is a conserved hypothetical protein without known function, but it has been predicted to be an outer membrane protein (<https://mycobrowser.epfl.ch/>). Other candidate targets include *narL* discovered in mutants to 412A. NarL belongs to one of the two-component regulatory systems and regulates the synthesis of formate dehydrogenase-N and nitrate reductase enzymes during aerobic nitrate metabolism<sup>43</sup>. Lead compound targeting NarL is being explored for Mtb treatment<sup>44</sup>. Mutations were observed in polyketide synthase *pkv6*, in two mutants (to compound 296A) having mutations in TB18.5. Pkv6 is involved in human infection<sup>45</sup>. Also, substitution was reported in *CtpC* from mutant to 648X; *ctpC* appeared to be important in the transport of heavy metal zinc and contributed to the survival of Mtb in macrophage<sup>46,47</sup>. In addition, mutants from compound 486X were found to have

mutations in *phoR* and *fbiA/fbiC*, which have important functions. PhoPR is a well-known 2-component regulator of pathogenic phenotypes, including secretion of the virulence factor ESAT-6, biosynthesis of acyltrehalose-based lipids, and modulation of antigen export<sup>48–50</sup>. Clinical mutants resistant to delamanid, a drug for *M. tuberculosis* were found to possess mutations in *fbiA* and *fbiC*, as well as *fbiB*<sup>51</sup>. Likewise, mutations in *fbiA* and *fbiC* have been related to resistance to delamanid in *M. bovis* BCG mutants; mutations in *fbiA* and *fbiC* likely to impact the F420 pathway, and delamanid requires a bio-activation by the F420-dependent nitro-reduction pathway to exert its anti-tuberculous activity<sup>51</sup>.

Mutants to 622A and 1114A have mutations in *virS* which is important - the expression of *mymA* operon genes may be regulated through PknK-mediated phosphorylation of VirS<sup>52</sup>. VirS is important for *M. tuberculosis* to block phagosomal-lysosomal fusion in the activated macrophages and to survive in acidic conditions<sup>53</sup>. Another mutant to compound 1114A has mutations in *dnaE1*, which is essential for high-fidelity DNA replication and is considered a potential drug target<sup>54,55</sup>. A mutant to 412A has a mutation in isoniazid inducible gene *iniB*, which is involved in cell wall synthesis<sup>56</sup>, while *moaC3* (mutant to 412A) is part of the Molybdenum cofactor (Moco) biosynthesis pathway, which may be significant to pathogenesis<sup>57</sup>.

Multiple mutants to 213A, 622A and 1114A gained the N-terminal stop codons in *sugI*, which encodes a sugar-transport membrane protein in *M. tuberculosis*<sup>58</sup>. The same mutation may be associated with the resistance to the second line drug D-cycloserine; Chen et al.<sup>13</sup> suggested the loss-of-function mutation discovered from a mutant may result in a lower uptake of cycloserine inside the cell, therefore leading to higher resistance to d-cycloserine.

The role of mutations in the genes of unknown functions or “relatively low-abundance” genes such as *Rv0370c*, *Rv3629*, *Rv1948*, *Rv1825*, *Rv0585c*, *Rv3175*, *Rv3327* is unclear; these mutations may be random or involved in compensating for resistance mutations or providing an additional level of resistance<sup>59</sup>. Fitness costs caused by chemical resistance mutations could be ameliorated by compensatory mutations, which do not contribute directly to drug resistance<sup>59</sup>. In fact, whole genome sequencing of MDR and XDR strains also revealed lots of mutations, and some of them may be trade-off or involved in compensation of fitness costs<sup>10</sup>.

## Conclusion

Bioinformatics analysis of 53 mutants screened against various compounds identified several promising genes that confer resistance to given chemical entities and as such may provide novel drug targets. Some targets of these chemical libraries are consistent with those that have been tied to the proposed mechanism of action or resistance (e.g. *rpoB*, *mmpL3*, *ethA*) and potential new pathway identified in our analysis (e.g. *prfB*). The analysis has extended our understanding of the biological basis for the anti-tuberculous actions. Future studies are needed to address the role of the identified mutations in genes of unknown functions and how they might be involved in the mode of action/ resistance of these compounds to tuberculosis.

# Materials And Methods

## Preparation of chemical compounds.

Hit compounds with potential anti-tuberculous activities have been identified and the chemicals were prepared under similar conditions as previously described in Sorrentino et al. <sup>6</sup>.

## Libraries.

The TB box library collection of 11,000 compounds that have shown a biological effect (DR curve) from any of the phenotypic HTS campaigns run against *M. tuberculosis* and *M. bovis*. Compounds with structure related to mode of action (MOA) known as antitubercular have been removed. Compounds showing pTOX50 values higher than 6.3 (< 0.5 uM) have been removed. The Small & Beautiful Library is composed of compounds drawn from GSKchem (not GSKscreen), Filtered on size and lipophilicity  $10 \leq \text{HAC} \leq 28$  (HAC = heavy atom count),  $-2 \leq \text{ClogP} \leq 3$ , Filtered on "promiscuity" (multiple targets, side effects, unsuitable DMPK)  $\text{IFI} \leq 3\%$  (IFI = Inhibition Frequency Index) and filtered using other physicochemical properties

$\text{MW} \leq 400$ ,  $\text{RotBonds} \leq 5$ ,  $0 \leq \text{HBD} < 8$ ,  $0 \leq \text{HBA} \leq 40$ ,  $\text{Neg} \leq 2$ ,  $0 \leq \text{Pos} \leq 2$ ,  $\text{AromRings} \leq 2$ ,  $\text{TotRings} \leq 3$

- Filtered on "shapeliness" (roundness)  $\text{fCsp3} \geq 0.3$  (i.e.  $\geq 30\%$  of carbon atoms must be sp<sup>3</sup>)
- Filtered on ALS availability, 150uL minimum in UP ALS required

Diversity selection & redundancy elimination were also performed

## Phenotypic screening

We utilized our recently developed *ex vivo* high-throughput screening assay to test extract and peptide activity against intracellular Mtb. Initially, we used the raw intensity luciferase method as described recently <sup>5,6</sup>. Briefly, Mtb cultures were opsonized and used to infect the THP-1 cells at a multiplicity of infection (MOI). 3h following infection, the medium was aspirated and replaced with a medium containing the extracts. After 24, 48 or 72 h, infected macrophages were harvested and lysed. Bright-Glo reagent (Promega TM052) was added to each culture and luciferase activity was measured with an aTropix TR7171 luminometer (Applied Biosystems, Foster City, CA) available in the BCL-3 facility. The process was repeated with fractionated and purified peptide until a pure homogenized compound was identified.

## Bacterial culture and mutant conditions.

*M. tuberculosis* H37Rv (ATCC 25618) was utilized for all experiments. Mtb strains (WT and all chemical mutants) were grown in 7H9 broth (Difco) supplemented with albumin-dextrose-catalase enrichment (ADC), 0.05% Tween 80. *M. tuberculosis* H37Rv was grown to mid-logarithmic phase in 7H9 broth supplemented with ADC (10%) and Tween 80 (0.05%). OD600 was measured, and the bacteria was spun

and resuspended in media to a final concentration of  $\sim 1 \times 10^8$  CFU/mL. The entire volume was plated onto Middlebrook 7H10 agar supplemented with 10% ADC 0.05% tween 80, 0.2% glycerol, carbon source of interest and 2X MIC90 compound concentration. Bacteria was also plated onto compound-free plates as controls. Plates were incubated for 4-6 weeks at 37°C containing 5% CO<sub>2</sub> or until colonies were observed. The spontaneous rate of resistance was calculated as the number of colonies on compound containing plates divided by the total number of viable bacteria estimated on compound free plates. Isolated resistant colonies were picked from compound containing plates and replated on to fresh 2x MIC90 compound containing Middlebrook 7H10 plates for confirmation of resistance. Once confirmed, colonies were picked and grown in 7H9 broth supplemented with 10% ADC and 2x MIC90 compound concentration and grown to mid-logarithmic phase. Genomic DNA was extracted using the lysozyme method according to published protocol<sup>60</sup>.

### **Whole genome sequencing of resistant mutants.**

Wild type *M. tuberculosis* H37Rv and mutants were characterized by WGS as described in<sup>61</sup>. Briefly, the DNA libraries were constructed with Nextera XT DNA kit (Illumina, San Diego, USA). The DNA was fragmented and purified with AMPure XP beads. DNA libraries of H37Rv wild type and mutant samples were normalized and sequenced using the MiSeq platform with 2 × 250 cycles (MiSeq Reagent Kit v2, Illumina) at British Columbia Center for Disease Control Public Health Laboratory and at Genome Québec Innovation Centre in McGill University.

### **Bioinformatics analysis.**

The quality of the reads was assessed by Fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were quality trimmed by Trim Galore ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)). Trimmed sequence reads were aligned to the reference genome sequence of H37Rv (NC\_000962.3) using BWA-mem<sup>62</sup>, and SNVs and indels were called using GATK v.3<sup>63</sup>. The SNVs generated using GATK were filtered using vcftools (10) to ensure high confidence. The parameters for filtering were (i) minimum read depth of 10; (ii) maximum base quality of 30 for every nucleotide in the sample; (iii) minimum mapping quality of 20. SnpEff<sup>64</sup> was used to annotate and to output the SNVs changes in mutants according to the reference genome and GFF files of Mtb H37Rv in NCBI. Unique variants in mutants were identified by examining the discordant SNVs between wild type and mutants that differed from the H37Rv reference in NCBI<sup>65</sup>. To avoid false-positive SNVs, the unique variant in mutants was inspected through Tablet<sup>66</sup>. SNPs and indels occurring in PPE/PE\_PGRS genes, which contain repetitive elements, were excluded to avoid inaccuracies in the read mapping and alignment in those portions of the genome<sup>10,38</sup>. Mutations arising from the comparison between the parental strain *M. tuberculosis* H37Rv and the standard reference genome (NC\_000962.3) were also excluded from the analysis. Reads from contamination data and other bacteria were excluded after analysis by Kraken v.1<sup>67</sup>. Raw reads were also assembled using SPAdes v.3.9.0<sup>68</sup> with *k-mer* sizes of 21, 33, 55, 77, and 99.

## Accession numbers

This whole genome shotgun project has been deposited in DDBJ/ENA/GenBank under BioProject PRJNA558545.

## Declarations

### ACKNOWLEDGMENTS

The project is funded by the Canadian Institutes for Health Research (CIHR project PJT-148646), The British Columbia Lung association and the Tres Cantos Open Lab foundation to Dr. Av-Gay. We are grateful to Julia Castro-Pichel (GSK, Madrid, Spain) for assistance in the chemical compounds, Mary Ko (Faculty of Medicine, UBC), Will Hsiao, Jun Duan, Miguel Uyaguari-Diaz, Tracy Lee, Neil Chin, Mel Krajdén and Linda Hoang (British Columbia Center for Disease Control Public Health Lab) for their technical assistance and discussion.

### Contributions

YA and AML conceived the idea of study. FS, NG, AML, RG, EPH, AL, AB, XZ performed the Mtb culture and phenotypic screening work. CKMT performed the whole genome sequencing and bioinformatics analysis. CKMT, MJR-B and YA drafted the manuscript. All authors have reviewed the manuscript.

## References

1. Guthrie, J. L. & Gardy, J. L. A brief primer on genomic epidemiology: lessons learned from *Mycobacterium tuberculosis*. *Ann. N. Y. Acad. Sci*, **1388**, 59–77 (2017).
2. Hmama, Z., Peña-Díaz, S., Joseph, S. & Av-Gay, Y. Immuno-evasion and immunosuppression of the macrophage by *Mycobacterium tuberculosis*. *Immunol. Rev*, **264**, 220–232 (2015).
3. Deoghare, S. Bedaquiline: a new drug approved for treatment of multidrug-resistant tuberculosis. *Indian J. Pharmacol*, **45**, 536–537 (2013).
4. Pethe, K. *et al.* A chemical genetic screen in *Mycobacterium tuberculosis* identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy. *Nat. Commun*, **1**, 57 (2010).
5. Zheng, X. & Av-Gay, Y. System for Efficacy and Cytotoxicity Screening of Inhibitors Targeting Intracellular *Mycobacterium tuberculosis*. *J. Vis. Exp*, <https://doi.org/10.3791/55273> (2017).
6. Sorrentino, F. *et al.* Development of an Intracellular Screen for New Compounds Able To Inhibit *Mycobacterium tuberculosis* Growth in Human Macrophages. *Antimicrob. Agents Chemother*, **60**, 640–645 (2016).
7. Galagan, J. E. Genomic insights into tuberculosis. *Nat. Rev. Genet*, **15**, 307–320 (2014).
8. Zeng, X. *et al.* Whole genome sequencing data of 1110 *Mycobacterium tuberculosis* isolates identifies insertions and deletions associated with drug resistance. *BMC Genomics*, **19**, 365 (2018).

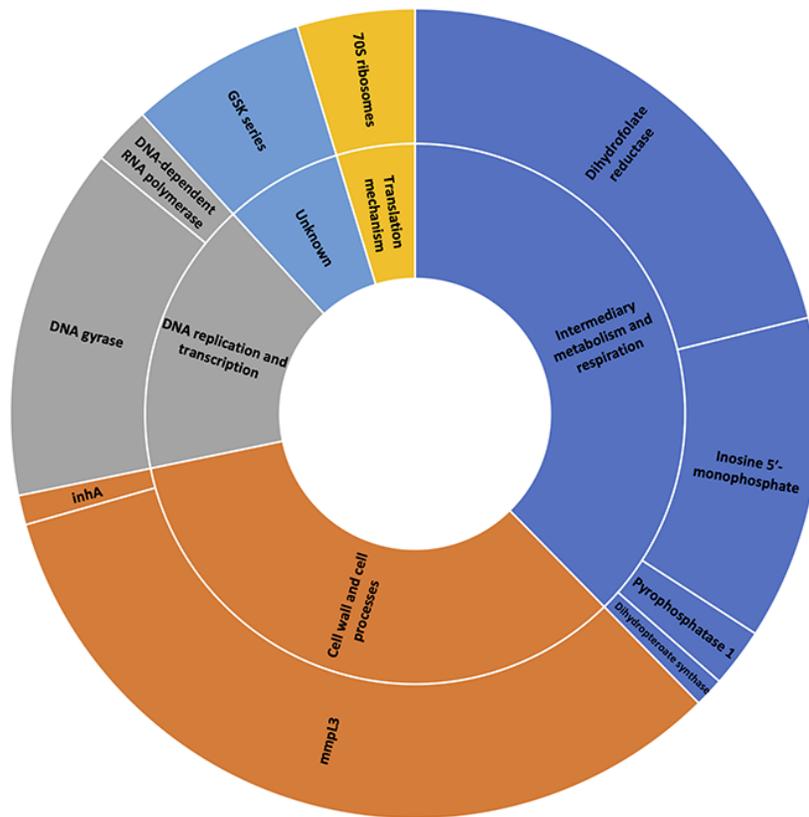
9. Gardy, J. L. *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med*, **364**, 730–739 (2011).
10. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet*, **45**, 1183–1189 (2013).
11. Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet*, **50**, 307–316 (2018).
12. Mortimer, T. D., Weber, A. M. & Pepperell, C. S. Signatures of Selection at Drug Resistance Loci in *Mycobacterium tuberculosis*. *mSystems* 3(2018).
13. Chen, J. *et al.* Identification of novel mutations associated with cycloserine resistance in *Mycobacterium tuberculosis*. *J. Antimicrob. Chemother*, **72**, 3272–3276 (2017).
14. Zhang, S. *et al.* Mutations in *panD* encoding aspartate decarboxylase are associated with pyrazinamide resistance in *Mycobacterium tuberculosis*. *Emerg. Microbes Infect*, **2**, e34 (2013).
15. Gurcha, S. S. *et al.* Biochemical and structural characterization of mycobacterial aspartyl-tRNA synthetase AspS, a promising TB drug target. *PLoS One*, **9**, e113568 (2014).
16. Abrahams, K. A. *et al.* Identification of KasA as the cellular target of an anti-tubercular scaffold. *Nat. Commun*, **7**, 12581 (2016).
17. Abrahams, K. A. *et al.* Inhibiting mycobacterial tryptophan synthase by targeting the inter-subunit interface. *Sci. Rep*, **7**, 9430 (2017).
18. Abrahams, K. A. *et al.* Identification of novel imidazo[1,2-a]pyridine inhibitors targeting *M. tuberculosis* QcrB. *PLoS One*, **7**, e52951 (2012).
19. Baulard, A. R. *et al.* Activation of the pro-drug ethionamide is regulated in mycobacteria. *J. Biol. Chem*, **275**, 28326–28331 (2000).
20. DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G. & Barry, C. E. 3 rd. Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9677–9682(2000).
21. Dover, L. G. *et al.* EthA, a common activator of thiocarbamide-containing drugs acting on different mycobacterial targets. *Antimicrob. Agents Chemother*, **51**, 1055–1063 (2007).
22. Tatum, N. J. *et al.* New active leads for tuberculosis booster drugs by structure-based drug discovery. *Org. Biomol. Chem*, **15**, 10245–10255 (2017).
23. Parish, T. *et al.* Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect. Immun*, **71**, 1134–1140 (2003).
24. Slinger, B. L., Deiorio-Hagggar, K., Anthony, J. S., Gilligan, M. M. & Meyer, M. M. Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. *BMC Genomics*, **15**, 657 (2014).
25. Grant, S. S. *et al.* Baeyer-Villiger Monooxygenases EthA and MymA Are Required for Activation of Replicating and Non-replicating *Mycobacterium tuberculosis* Inhibitors. *Cell Chem Biol*, **23**, 666–677 (2016).

26. Moure, A. L. *et al.* MymA Bioactivated Thioalkylbenzoxazole Prodrug Family Active against Mycobacterium tuberculosis. *J. Med. Chem*, **63**, 4732–4748 (2020).
27. Herrán, E. *et al.* Ko Mary, Losana Alfonso Mendoza, Remuinan-Blanco Modesto Jesus, Av-Gay Yossef. Oxadiazols active against Mycobacterium tuberculosis; resistant mutants are mapped to MmpL3. (in prep)
28. Grzegorzewicz, A. E. *et al.* Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane. *Nat. Chem. Biol*, **8**, 334–341 (2012).
29. Meneghetti, F. *et al.* Iron Acquisition Pathways as Targets for Antitubercular Drugs. *Curr. Med. Chem*, **23**, 4009–4026 (2016).
30. Tahlan, K. *et al.* SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. *Antimicrob. Agents Chemother*, **56**, 1797–1809 (2012).
31. Xu, Z., Meshcheryakov, V. A., Poce, G. & Chng, S. S. MmpL3 is the flippase for mycolic acids in mycobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 114, 7993–7998(2017).
32. Nowak, E. *et al.* Structural and functional aspects of the sensor histidine kinase PrrB from Mycobacterium tuberculosis. *Structure*, **14**, 275–285 (2006).
33. Bellale, E. *et al.* Diarylthiazole: an antimycobacterial scaffold potentially targeting PrrB-PrrA two-component system. *J. Med. Chem*, **57**, 6572–6582 (2014).
34. Ewann, F. *et al.* Transient requirement of the PrrA-PrrB two-component system for early intracellular multiplication of Mycobacterium tuberculosis. *Infect. Immun*, **70**, 2256–2263 (2002).
35. Villellas, C. *et al.* Unexpected high prevalence of resistance-associated Rv0678 variants in MDR-TB patients without documented prior use of clofazimine or bedaquiline. *J. Antimicrob. Chemother*, **72**, 684–690 (2017).
36. Hartkoorn, R. C., Uplekar, S. & Cole, S. T. Cross-resistance between clofazimine and bedaquiline through upregulation of MmpL5 in Mycobacterium tuberculosis. *Antimicrob. Agents Chemother*, **58**, 2979–2981 (2014).
37. Radhakrishnan, A. *et al.* Crystal structure of the transcriptional regulator Rv0678 of Mycobacterium tuberculosis. *J. Biol. Chem*, **289**, 16526–16540 (2014).
38. loerger, T. R. *et al.* Identification of new drug targets and resistance mechanisms in Mycobacterium tuberculosis. *PLoS One*, **8**, e75245 (2013).
39. Duckworth, B. P. *et al.* Development of a selective activity-based probe for adenylating enzymes: profiling MbtA Involved in siderophore biosynthesis from Mycobacterium tuberculosis. *ACS Chem. Biol*, **7**, 1653–1658 (2012).
40. Ferguson, L. *et al.* Integrated Target-Based and Phenotypic Screening Approaches for the Identification of Anti-Tubercular Agents That Bind to the Mycobacterial Adenylating Enzyme MbtA. *ChemMedChem*, **14**, 1735–1741 (2019).
41. Nelson, K. M. *et al.* Synthesis and Pharmacokinetic Evaluation of Siderophore Biosynthesis Inhibitors for Mycobacterium tuberculosis. *J. Med. Chem*, **58**, 5459–5475 (2015).

42. Lun, S. *et al.* Pharmacokinetic and in vivo efficacy studies of the mycobactin biosynthesis inhibitor salicyl-AMS in mice. *Antimicrob. Agents Chemother*, **57**, 5138–5140 (2013).
43. Malhotra, V., Agrawal, R., Duncan, T. R., Saini, D. K. & Clark-Curtiss, J. E. Mycobacterium tuberculosis response regulators, DevR and NarL, interact in vivo and co-regulate gene expression during aerobic nitrate metabolism. *J. Biol. Chem*, **290**, 8294–8309 (2015).
44. Kumar, N., Srivastava, R., Prakash, A. & Lynn, A. M. Structure-based virtual screening, molecular dynamics simulation and MM-PBSA toward identifying the inhibitors for two-component regulatory system protein NarL of Mycobacterium Tuberculosis. *J. Biomol. Struct. Dyn.* 1–15(2019).
45. Kumar, M. *et al.* Identification of Mycobacterium tuberculosis genes preferentially expressed during human infection. *Microb. Pathog*, **50**, 31–38 (2011).
46. Botella, H. *et al.* Mycobacterial p(1)-type ATPases mediate resistance to zinc poisoning in human macrophages. *Cell Host Microbe*, **10**, 248–259 (2011).
47. Padilla-Benavides, T., Long, J. E., Raimunda, D., Sasseti, C. M. & Argüello, J. M. A novel P(1B)-type Mn<sup>2+</sup>-transporting ATPase is required for secreted protein metallation in mycobacteria. *J. Biol. Chem*, **288**, 11334–11347 (2013).
48. Cimino, M. *et al.* Identification of DNA binding motifs of the Mycobacterium tuberculosis PhoP/PhoR two-component signal transduction system. *PLoS One*, **7**, e42876 (2012).
49. Xing, D. *et al.* Asymmetric Structure of the Dimerization Domain of PhoR, a Sensor Kinase Important for the Virulence of Mycobacterium tuberculosis. *ACS Omega*, **2**, 3509–3517 (2017).
50. Broset, E., Martín, C. & Gonzalo-Asensio, J. Evolutionary landscape of the Mycobacterium tuberculosis complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *MBio*, **6**, e01289–15 (2015).
51. Fujiwara, M., Kawasaki, M., Hariguchi, N., Liu, Y. & Matsumoto, M. Mechanisms of resistance to delamanid, a drug for Mycobacterium tuberculosis. *Tuberculosis*, **108**, 186–194 (2018).
52. Kumar, P. *et al.* The Mycobacterium tuberculosis protein kinase K modulates activation of transcription from the promoter of mycobacterial monooxygenase operon through phosphorylation of the transcriptional regulator VirS. *J. Biol. Chem*, **284**, 11090–11099 (2009).
53. Singh, S., Goswami, N., Tyagi, A. K. & Khare, G. Unraveling the role of the transcriptional regulator VirS in low pH-induced responses of Mycobacterium tuberculosis and identification of VirS inhibitors. *J. Biol. Chem*, **294**, 10055–10075 (2019).
54. Baños-Mateos, S. *et al.* High-fidelity DNA replication in Mycobacterium tuberculosis relies on a trinuclear zinc center. *Nat. Commun*, **8**, 855 (2017).
55. Nasir, N. & Kisker, C. Mechanistic insights into the enzymatic activity and inhibition of the replicative polymerase exonuclease domain from Mycobacterium tuberculosis. *DNA Repair*, **74**, 17–25 (2019).
56. Alland, D., Steyn, A. J., Weisbrod, T., Aldrich, K. & Jacobs, W. R. Jr. Characterization of the Mycobacterium tuberculosis iniBAC promoter, a promoter that responds to cell wall biosynthesis inhibition. *J. Bacteriol*, **182**, 1802–1811 (2000).

57. Williams, M., Mizrahi, V. & Kana, B. D. Molybdenum cofactor: a key component of Mycobacterium tuberculosis pathogenesis? *Crit. Rev. Microbiol*, **40**, 18–29 (2014).
58. Titgemeyer, F. *et al.* A genomic view of sugar transport in Mycobacterium smegmatis and Mycobacterium tuberculosis. *J. Bacteriol*, **189**, 5903–5915 (2007).
59. Gygli, S. M., Borrell, S., Trauner, A. & Gagneux, S. Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS Microbiol. Rev*, **41**, 354–373 (2017).
60. *Molecular genetics of mycobacteria*. (American Society of Microbiology, 2014).
61. Tsui, C. K. M. *et al.* Genome Sequences of the Mycobacterium tuberculosis H37Rv-ptkA Deletion Mutant and Its Parental Strain. *Genome Announc*, **5**, (2017).
62. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform., **26**, 589–595 (2010).
63. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, **20**, 1297–1303 (2010).
64. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3., **6**, 80–92 (2012).
65. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, **393**, 537–544 (1998).
66. Milne, I., Bayer, M., Stephen, G., Cardle, L. & Marshall, D. Tablet: Visualizing Next-Generation Sequence Assemblies and Mappings. *Methods Mol. Biol*, **1374**, 253–268 (2016).
67. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*, **15**, R46 (2014).
68. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*, **19**, 455–477 (2012).

## Figures



**Figure 1**

Clustering analysis of 85 hit compounds into different known targets or chemical entities.



**Figure 2**

The function of candidate target genes identified from *M. tuberculosis* H37Rv mutants from screening.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytable1.xlsx](#)