RESEARCH ARTICLE

Systems Biology-Based Identification of *Mycobacterium tuberculosis* Persistence Genes in Mouse Lungs

Noton K. Dutta,[a] Nirmalya Bandyopadhyay,[b] Balaji Veeramani,[b]* Gyanu Lamichhane,[a] Petros C. Karakousis,[a,c] Joel S. Bader[b]

Division of Infectious Diseases, Department of Medicine, Center for Tuberculosis Research, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA[a]; Department of Biomedical Engineering, High-Throughput Biology Center and Institute of Computational Medicine, Johns Hopkins University, Baltimore, Maryland, USA[b]; Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA[c]

* Present address: Balaji Veeramani, Dow AgroSciences, Indianapolis, Indiana, USA.

N.K.D. and N.B. contributed equally to this article.

ABSTRACT **Identifying *Mycobacterium tuberculosis* persistence genes is important for developing novel drugs to shorten the duration of tuberculosis (TB) treatment. We developed computational algorithms that predict *M. tuberculosis* genes required for long-term survival in mouse lungs. As the input, we used high-throughput *M. tuberculosis* mutant library screen data, mycobacterial global transcriptional profiles in mice and macrophages, and functional interaction networks. We selected 57 unique, genetically defined mutants (18 previously tested and 39 untested) to assess the predictive power of this approach in the murine model of TB infection. We observed a 6-fold enrichment in the predicted set of *M. tuberculosis* genes required for persistence in mouse lungs relative to randomly selected mutant pools. Our results also allowed us to reclassify several genes as required for *M. tuberculosis* persistence *in vivo*. Finally, the new results implicated additional high-priority candidate genes for testing. Experimental validation of computational predictions demonstrates the power of this systems biology approach for elucidating *M. tuberculosis* persistence genes.**

IMPORTANCE ***Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), has a genetic repertoire that permits it to persist in the face of host immune responses. Identification of such persistence genes could reveal novel drug targets and elucidate mechanisms by which the organism eludes the immune system and resists drugs. Genetic screens have identified a total of 31 persistence genes, but to date only 15% of the ~4,000 *M. tuberculosis* genes have been tested experimentally. In this paper, as an alternative to brute force experimental screens, we describe computational methods that predict new persistence genes by combining known examples with growing databases of biological networks. Experimental testing demonstrated that these predictions are highly accurate, validating the computational approach and providing new information about *M. tuberculosis* persistence in host tissues. Using the new experimental results as additional input highlights additional genes for testing. Our approach can be extended to other data types and target organisms to characterize host-pathogen interactions relevant to this and other infectious diseases.**

Received 23 December 2013 **Accepted** 30 December 2013 **Published** 18 February 2014

Citation Dutta NK, Bandyopadhyay N, Veeramani B, Lamichhane G, Karakousis PC, Bader JS. 2014. Systems biology-based identification of *Mycobacterium tuberculosis* persistence genes in mouse lungs. mBio 5(1):e01066-13. doi:10.1128/mBio.01066-13.

Editor Eric Rubin, Harvard School of Public Health

Address correspondence to Joel S. Bader, joel.bader@jhu.edu, or Petros C. Karakousis, petros@jhmi.edu.

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), has evolved adaptive mechanisms to avoid killing by host immune responses. Identifying metabolic and regulatory pathways required for *M. tuberculosis* persistence in host tissues may elucidate novel strategies to eradicate TB infection. The availability of the *M. tuberculosis* genome sequence has enabled high-throughput screens using subsaturated transposon (Tn) mutant libraries (1, 2). Such libraries have been used to study the genetic requirements of the pathogen under physiologically relevant stress conditions, including during infection of macrophages (3), mice (4–6), guinea pigs (7, 8), and nonhuman primates (9).

Recently, there has been substantial interest in developing computational algorithms for accurately predicting genes essential for *M. tuberculosis* growth and survival. Flux balance analysis uses the stoichiometry of biochemical reactions to predict growth requirements but is limited to metabolic enzymes (10–12). Other approaches have enhanced flux balance analysis by including transcriptional profiles and regulatory relationships to constrain fluxes through metabolic reactions (13, 14). These approaches have been used to predict drug effects on *M. tuberculosis* mycolic acid biosynthesis capacity and transcription factor knockout phenotypes (13, 14). Approaches to predict genetic requirements beyond metabolism would have great value, particularly since only 660 *M. tuberculosis* genes (~17% of the genome) are represented in metabolic reconstructions.

Alternative approaches described here combine actual physical interactions, including enzyme-substrate and protein-protein interactions, with functional associations. The resulting networks

can be exploited to predict protein function and mutant phenotypes (15). Simple metrics, such as shortest distance to known genes of interest, have been used previously to predict *M. tuberculosis* drug resistance genes (16). Graph diffusion kernels, introduced first for searching Web pages, additionally account for multiple independent network paths and improve performance. Successes have included predicting epistatic genetic interactions in yeast (17, 18), predicting protein function through protein-protein interactions (19), and identifying candidate genes for disease (20, 21). Biological networks with different interaction types can provide complementary information, and integrative approaches modeling biological functions have been used to predict protein-protein interactions (22, 23), synthetic lethal interactions (17), co-complexed pairs (24), and driver missense mutations (25). In this study, we combined known *M. tuberculosis* persistence genes and transcriptional profiles with networks from metabolic reconstructions and functional associations to make genome-wide predictions of genes required for mycobacterial persistence in the host (26, 27). The top-ranked predictions were then tested experimentally to confirm their accuracy. Further, we developed new computational algorithms, incorporating recently published data sets (28–31), which together with our new experimental results highlight additional genes for testing. This study extends our knowledge of *M. tuberculosis* persistence and identifies potential novel drug targets, with the ultimate goal of shortening the duration of TB treatment. This systems biology approach, combining computational predictions with experimental validation, is general and readily extended to new data types and other target organisms, including host-pathogen interactions relevant to this and other infectious diseases.

RESULTS

Computational predictions. Computational predictions (see Data Set S1 and S2 in the supplemental material) were used to prioritize mutants for experimental tests in mice (Fig. 1; see Data Set S3A and B). The predictions propagated gene-based phenotypes (Table 1), including known persistence defects and additional informative phenotypes, through *M. tuberculosis* gene networks to generate gene-based features for predicting additional persistence mutants with logistic regression (see Data Set S3C).

Known *in vivo* persistence genes were derived from a Tn mutant screen using designer arrays for defined mutant analysis (DeADMAn) (5). This screen identified 31 persistence genes and 474 genes not required for persistence in mouse lungs. These genes served as known positives and negatives, respectively. Additional relevant gene data sets included Tn site hybridization (TraSH) data derived from mouse spleen (6) and murine macrophages (3). Genes required for *in vitro* growth were obtained from Tn mutagenesis screens (1, 2). Genes differentially expressed during infection were obtained from transcription profiling studies (32, 33).

Networks of functional associations were obtained from publicly available metabolic reconstructions (11) and data integration approaches (27). A steady-state graph diffusion kernel propagated the gene data (persistence genes, essential genes, and differentially expressed genes) through the networks to create features for logistic regression and support vector machine classifiers (see Data Set S1). The full logistic regression model included all 28 features; stepwise selection with the Akaike information criterion (AIC) eliminated redundant and uninformative features. Twentyfold cross-validation was used to assess performance based on the known positives and negatives, with area under the receiver operating curve (AUROC) and the maximum harmonic mean of precision and recall (*F* score) serving as quantitative criteria. Ten different random 20-way splits were performed to ensure robust results.

Stepwise logistic regression and full logistic regression were equivalent, and both regression methods were superior to support vector machines (Fig. 2). The *F* score for all methods is maximal near 20 to 30% recall. Stepwise regression at 20% recall is predicted to have a mean precision of ~50%, an approximately 8-fold enrichment compared to the overall estimate of *in vivo* persistence genes within the entire genome (6%) (5). Stepwise logistic regression was chosen as the most parsimonious model and used to predict genome-wide persistence requirements based on the 11 features selected for the full data (Table 2). Known positives and negatives ranked by cross-validation provided empirical estimates of precision and recall as a function of ranking. Predicted values are provided genome-wide (see Data Set S2).

Gene selection for experimental verification. The top 75 computationally predicted genes were selected in rank order, in addition to the positive and negative controls, *pknF* (*Rv1746*) and *Rv1863c* (9), respectively, yielding 77 candidate genes. Of these 77 genes, 7 had unfavorable rankings as the prediction method was being developed, and 1 known positive was not selected for testing, leaving 69 genes selected for testing. Of the 70 corresponding mutant strains, 7 failed to grow sufficiently *in vitro*, yielding 63 *M. tuberculosis* Tn mutants corresponding to 62 unique genes in the infection pool.

Experimental verification in the murine model of TB infection. On the day after aerosol infection of BALB/c mice, the implantation dose was determined to be 2.71 ± 0.01 $\log_{10}$ bacilli. The output time point of 14 weeks was selected to evaluate mutant persistence in mouse lungs for consistency with previous studies used for statistical modeling (5). In addition, earlier (day 49) and day 196 time points were included to permit a kinetic analysis of individual mutant survival.

Total lung bacillary counts increased and mice gained weight as expected (see Data Set S3A and B). Gross examination of mouse lungs 49 days postinfection and beyond revealed discrete tubercle lesions. Histological evaluation showed cellular aggregates comprising primarily lymphocytes, with few histiocytes and plasma cells. Acid-fast bacilli were localized primarily within foamy macrophages (data not shown).

The ability of each mutant to survive in the host was ascertained by quantitative real-time PCR (qPCR). PCR primers failed to amplify 5 of the mutants. Of 63 mutants used, 5 (the *Rv0099*, *Rv0101*, *Rv1183*, *Rv1821*, *Rv3823c* mutants) repeatedly failed to amplify and were removed from further analysis. Data were available for a total of 58 mutants corresponding to 57 unique genes, including 6 known positives previously characterized as having a persistence phenotype, 12 known negatives previously characterized as not required for persistence in mouse lung, and 39 mutants previously uncharacterized by DeADMAn. The mean predicted precision was 32%.

Wild-type (null) mutants showed no change in representation over time. On the other hand, attenuated mutants showed an increase in cycle threshold (C_T) number over time, and "hypervirulent" mutants showed a decreasing C_T over time, indicating a population fraction increase. Mutants having a multiple-testing-

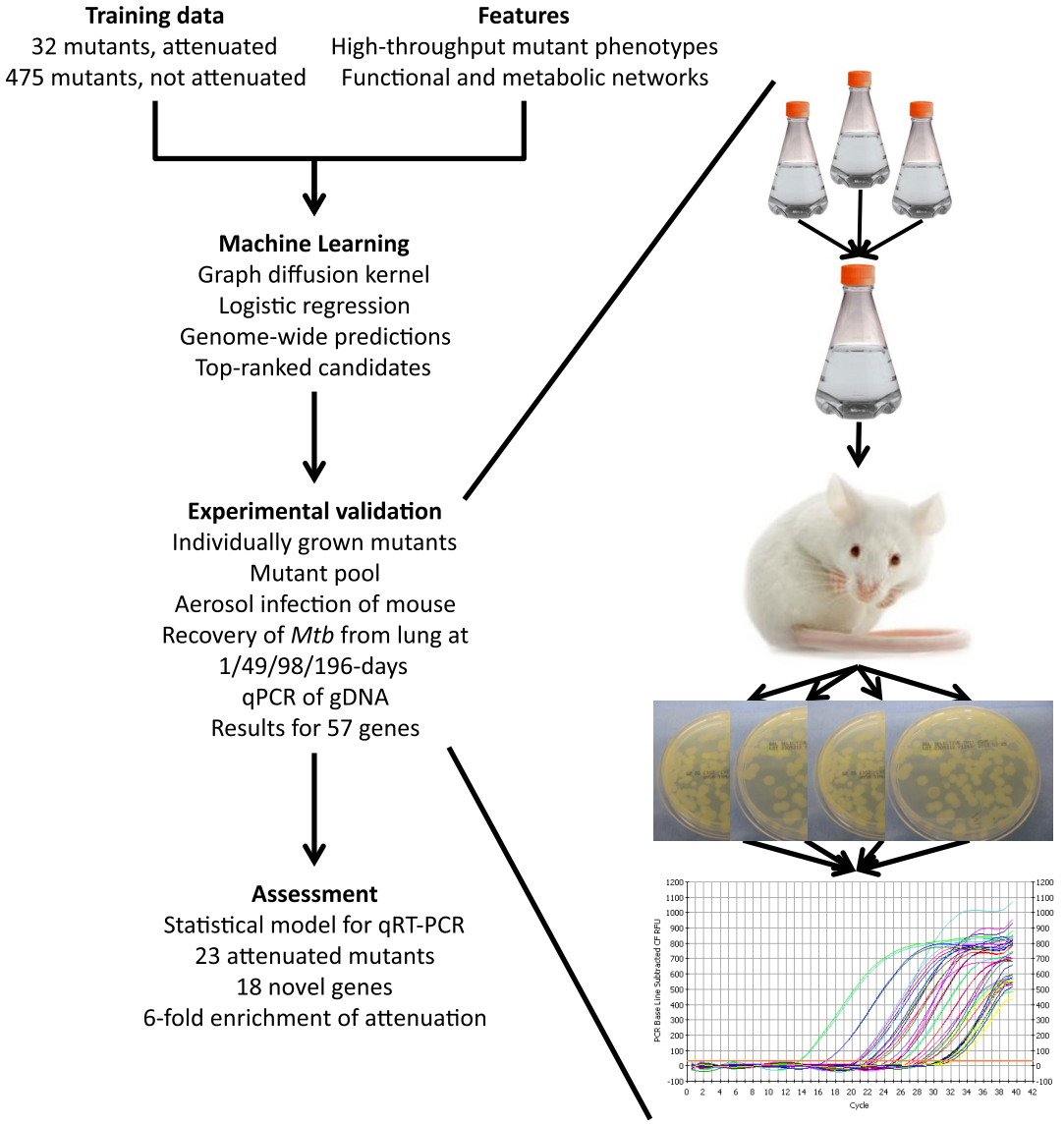

FIG 1 Overview of study design. Phenotypes from previous studies of *M. tuberculosis* persistence in mouse lungs were combined with high-throughput data and functional and metabolic networks to predict new candidate genes for experimental testing. Mutants corresponding to the top-ranked genes were grown, pooled, and used for aerosol infection of mouse. Mutants were recovered from lungs at 1, 49, 98, and 196 days postinfection, and abundance for 57 mutants was characterized by qPCR. Statistical models identified 23 of the 57 mutants as attenuated, including 18 novel genes, representing a 6-fold enrichment over the fraction required for persistence genome-wide.

corrected *P* value of 0.05 were classified as either attenuated or virulent; both replicates of *Rv0169* had concordant null phenotypes. Of the 57 unique genes tested, 23 were found to be attenuated, 3 virulent, and 31 null (Table 3). Roughly equivalent results are obtained using a threshold of 95% posterior probability for a mutant to belong to the attenuated class. These thresholds correspond to a change of about 1 C_T unit between measurements or an average change of 3 C_T units (~8-fold attenuation) from the first to the last of the 4 time points.

Statistical assessment of performance on known genes and novel predictions. Of the 6 known positives that were tested, 5 gave growth defects in this test. The single known positive with no growth defect was *lldD2* (*Rv1872c*). However, the previously studied *Rv1872c* mutant was in an *sigF* deletion background (5), perhaps accounting for the persistence phenotype. Of the 12 known negatives that were tested, 8 remained negative. Four, however, were attenuated: *atsd* (*Rv0663*), *hrca* (*Rv2374c*), *fadA6* (*Rv3556c*), and *Rv3870*. All four have been tested previously in related TraSH studies, and all but *Rv0663* were required for growth in mouse spleen (2). The overall concordance for previously characterized mutants is at least (5 + 8)/(6 + 12), or 72%, and may be closer to (5 + 11)/(5 + 12), or 94%.

Of the 39 unique novel genes tested, 22 had no persistence defect and 17 were found to have a non-wild-type phenotype, 14 with persistence defects and three with increased growth relative to the wild type (Table 3). The attenuation ranged from 8-fold (the lower limit for statistical significance) to over 100,000-fold (the dynamic range of qPCR) (Fig. 3). Counting only the attenuated

TABLE 1 Sources of data input into computational models

Source	Description	Edge wt	Gene wt
M. tuberculosis networks			
STRING functional associations (27)	3,964 nodes, 496,278 edges	Combined score ∈ [0, 1]	
BiGG metabolic reconstruction (11)	661 nodes, 217,470 edges	Poisson score mapped to [0, 1]	
M. tuberculosis essential genes			
Transposon mutants (1)	3,795 genes, Gibbs sampling posterior probability		Pr(essential) ∈ [0, 1]
TraSH (2)	3,172 genes		Log(input/output)
M. tuberculosis persistence genes			
DeADMAn in mouse (5)	31 persistence genes, 474 nonpersistence genes		+1 (persistence), −1 (nonpersistence), 0 (untested)
TraSH in mouse (6)	2,967 genes, measured 8 weeks after infection		Log(input/output)
TraSH in mouse macrophage (3)	2,859 genes, unactivated macrophage		Log(input/output)
TraSH in mouse macrophage (3)	2,859 genes, activated with IFN-γ before infection		Log(input/output)
TraSH in mouse macrophage (3)	2,859 genes, activated with IFN-γ after infection		Log(input/output)
M. tuberculosis differentially expressed genes			
Mouse infection (33)	Weeks 1, 2, 4, and 8 after infection		Log(input/output)
Macrophage infection (32)	Hours 4 and 24 after infection		Log(input/output)

strains as correct predictions, this 14/39 or 36% success rate is close to the 32% success rate predicted by the statistical model and represents a 6-fold enrichment over the 6% estimate of *in vivo* persistence genes (5).

The 23 genes required for persistence in mouse lungs in this assay include 5 that were previously known to be required and 18 novel genes that were either not tested or likely false negatives in previous mouse lung screens (Table 3).

Concordance of experimental model systems. This and a previous study (5) used medium-throughput assays to test 545 genotypically characterized mutants for persistence in mouse lungs following aerosol infection (see Data Set S4A to E). Similar mutants have also been tested as part of high-throughput, complex libraries using TraSH to study bacillary survival in macrophages and in mouse spleen following intravenous infection (3, 6). Of the 459 genes tested by all three systems, 76 of the corresponding mutants have a defect in at least one of the three systems: 8 are attenuated in all three systems, and an additional 18 are attenuated in two of the three systems (see Data Set S4F).

All pairwise comparisons of mutant phenotypes with 2-by-2 contingencies are highly significant (see Data Set S4G and H). It does appear, however, that the *in vivo* DeADMAn system is more similar to the corresponding *in vivo* TraSH mouse system (odds ratio of 13.3, Fisher's exact one-sided *P* value of 1.3×10^{-10}) than

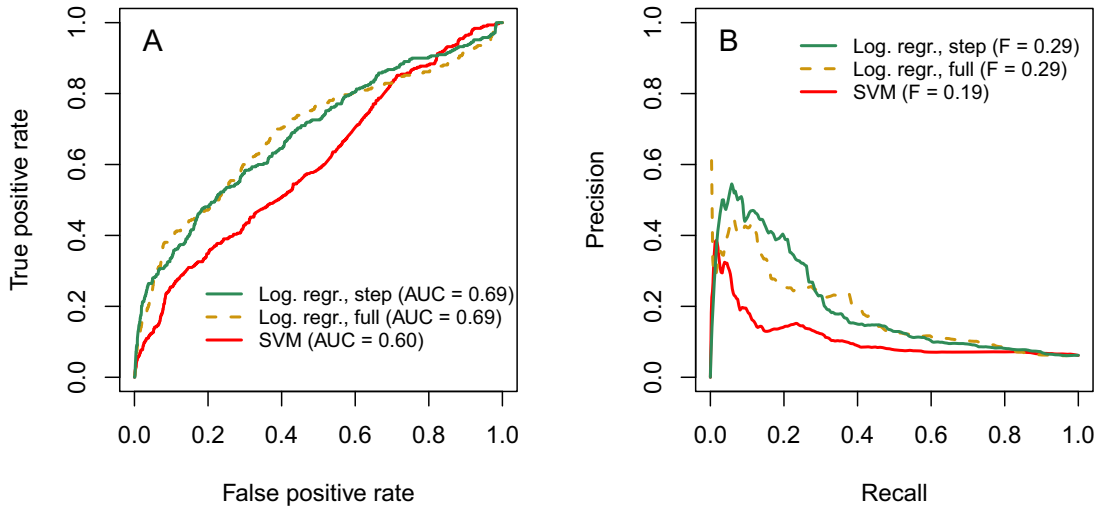

FIG 2 Statistical assessment of prediction methods. Predictions using logistic regression with stepwise selection by AIC (solid, green), logistic regression with a full model (dashed, orange), and a support vector machine (solid, red) are assessed by receiver operating characteristic (A) and precision recall using 20-fold cross-validation (B). Logistic regression with a full model or stepwise selection provide equivalent performance and are superior to the support vector machine.

TABLE 2 Stepwise logistic regression model

Feature	Coefficient	*P* value
Intercept	$-17.55 \pm 1{,}067.87$	9.86×10^{-1}
GDK (STRING, DeADMAn in mouse)[a]	8.37 ± 2.21	1.57×10^{-4}
GDK (STRING, TraSH in mouse macrophage after IFN-γ)	0.66 ± 0.37	7.47×10^{-2}
GDK (STRING, TraSH in mouse)	1.62 ± 0.42	1.14×10^{-4}
GDK (STRING, TraSH essential genes)	0.40 ± 0.23	8.95×10^{-2}
GDK (metabolic, DeADMAn in mouse)	-110.06 ± 78.75	1.62×10^{-1}
GDK (metabolic, TraSH in mouse macrophage unactivated)	-1.35 ± 0.76	7.62×10^{-2}
GDK (metabolic, transposon mutants)	$1{,}232.09 \pm 793.44$	1.20×10^{-1}
Mouse infection day 14	-0.63 ± 0.42	1.39×10^{-1}
Mouse infection day 21	0.65 ± 0.22	3.42×10^{-3}
Indicator (mouse infection day 7)	-3.04 ± 1.44	3.50×10^{-2}
Indicator (mouse infection day 14)	$17.82 \pm 1{,}067.88$	9.86×10^{-1}

[a] GDK (network, gene data) indicates features from a graph diffusion kernel with the given network and gene data.

to TraSH in macrophages (odds ratio of 7.4, *P* value of 3.4×10^{-5}). The two TraSH systems are also significantly correlated (odds ratio of 14.2, *P* value of 5.2×10^{-9}). Of genes attenuated by TraSH overall, 37% are also required for persistence in mouse lungs, similar to the predictive performance of the statistical model. It is important to note, however, that this study identified 12 of the genes attenuated in both. Prior to this study, only 24% of genes attenuated by TraSH were also found to be attenuated using DeADMAn. Furthermore, of the mutants tested across all three systems, distinct sets are attenuated in only a single system: 20 are unique to DeADMAn, 18 are unique to TraSH in mice, and 12 are unique to TraSH in macrophages. These results suggest corresponding distinct mechanisms. The number of mutants unique to TraSH in macrophages is smallest, possibly because macrophage infection is common to all three systems.

Predictions with updated external data and new results from this study. We investigated (see Data Set S4 to S6) whether recently reported external data improved our predictions (28–31). Incorporating four new external data sets with improved annotation of essential genes did not improve the predictions: the area under the curve (AUC) remained close to 0.69 and the *F* score remained close to 0.30 (see Data Set S4A and B). We also updated the predictions by including the new experimental results of this study, which update gene labels from "untested" to either "attenuated" or "null," together with the four new external data sets (Fig. 4). In the three cases where the new experimental results conflicted with previous results (*lprK* [*Rv0173*], *lldD2* [*Rv1872c*], *tig* [*Rv2462c*]), we used the new results for cross-validation tests. Here, the prediction performance improved substantially, with a new AUC of 0.77 and a new *F* score of 0.42 (see Data Set S4C and D). Three genes are particularly noteworthy in rising substantially in priority and also having mutants available for testing: *Rv1410c*, *fadD21* (*Rv1185c*), and *pheA* (*Rv3838c*).

DISCUSSION

Although many studies have highlighted the importance of various adaptive mechanisms in promoting the long-term persistence of *M. tuberculosis* in host tissues, the *M. tuberculosis* molecular pathways underlying long-term survival in the infected host remain largely undefined (34–36). This information is not only important for improving our understanding of TB pathogenesis but could also serve as the basis for the rational development of novel

TABLE 3 Experimental results of Tn mutant survival in mice and comparison with prior high-throughput studies

Gene(s)	Count	Result		
		This screen	DeADMAn (5)	TraSH (3, 6)
mkl (*Rv0655*)	1	Attenuated	Attenuated	Attenuated
mmpL11 (*Rv0202c*), *fadD26* (*Rv2930*)	2	Attenuated	Attenuated	Null
mmpL4 (*Rv0450c*), *pknF* (*Rv1746*)	2	Attenuated	Attenuated	Untested
hrcA (*Rv2374c*), *fadA6* (*Rv3556c*), *Rv3870*	3	Attenuated	Null	Attenuated
atsD (*Rv0663*)	1	Attenuated	Null	Null
pks16 (*Rv1013*), *Rv1045*, *lprG* (*Rv1411c*), *bioA* (*Rv1568*), *aceE* (*Rv2241*), *cpsA* (*Rv3484*), *Rv3683*, *Rv3723*, *Rv3871*	9	Attenuated	Untested	Attenuated
pntB (*Rv0157*), *Rv1226c*, *Rv1591*, *mez* (*Rv2332*), *tig* (*Rv2462c*)	5	Attenuated	Untested	Null
lldD2 (*Rv1872c*)	1	Null	Attenuated	Null
mce1A (*Rv0169*), *Rv2707*	2	Null	Null	Attenuated
mmpL6 (*Rv1557*), *Rv1863c*, *Rv2674*, *ppsE* (*Rv2935*), *pks1* (*Rv2946c*)	5	Null	Null	Null
fadD28 (*Rv2941*)	1	Null	Null	Untested
lprK (*Rv0173*), *Rv0176*, *pcaA* (*Rv0470c*), *lpqY* (*Rv1235*), *ppgK* (*Rv2702*), *drrA* (*Rv2936*), *Rv3236c*, *Rv3616c*, *Rv3910*	9	Null	Untested	Attenuated
gca (*Rv0112*), *Rv0203*, *Rv0660c*, *Rv0662c*, *fabG* (*Rv2766c*), *dinF* (*Rv2836c*), *ppsC* (*Rv2933*), *Rv3253c*, *aspB* (*Rv3565*)	9	Null	Untested	Null
ppsA (*Rv2931*), *ppsB* (*Rv2932*), *ppsD* (*Rv2934*), *Rv3273*	4	Null	Untested	Untested
rodA (*Rv0017c*)	1	Virulent	Untested	Attenuated
gabD2 (*Rv1731*), *pgsA2* (*Rv1822*)	2	Virulent	Untested	Null

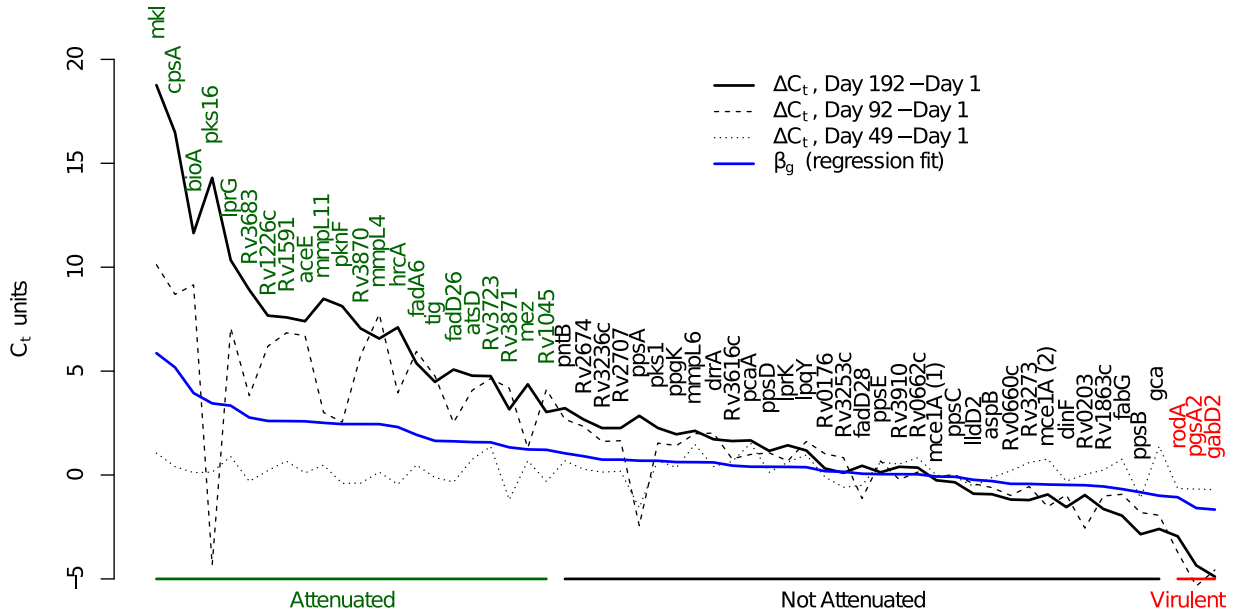

FIG 3 *M. tuberculosis* Tn mutant survival, as assessed by qPCR. Genes are sorted in decreasing order of β_g (blue line), the regression fit of the change in ΔC_T over 3 time intervals; large positive values correspond to attenuated mutants (green), and large negative values correspond to virulent mutants (red). The ΔC_T values at day 49 (dotted line), day 98 (dashed line), and day 196 (solid line) are shown relative to the day 1 baseline.

sterilizing drugs to shorten the duration of TB chemotherapy. The computational methods developed here provide a genome-scale ranking of bacterial mutants by predicting persistence phenotypes. The predictions are then validated by medium-scale tests of tens to hundreds of mutants in a mouse model. Using this approach, we observed a 6-fold enrichment in the predicted set of *M. tuberculosis* genes required for persistence in mouse lungs relative to randomly selected mutant pools.

We identified 18 genes, which were previously not characterized as *M. tuberculosis* persistence in animal lungs. Of these genes, *Rv1013*, *Rv1411c*, *Rv2374c*, *Rv2462c*, *Rv3484*, *Rv3556c*, *Rv3683*, *Rv3870*, and *Rv3871* were found to be significantly differentially expressed during nutrient deprivation of *M. tuberculosis* (37, 38), consistent with the hypothesis that the encoded products are involved in adaptation of *M. tuberculosis* to the nutrient-deprived environment of mouse lungs during chronic infection. The novel persistence genes *Rv1226c*, *Rv2462c*, *Rv3556c*, *Rv3683*, and *Rv3723* were shown to be significantly differentially regulated by *M. tuberculosis* upon inorganic phosphate limitation, suggesting that the cognate products may contribute to bacillary survival within the phosphate-starved environment of the macrophage phagolysosome during chronic infection (3, 39). These genes represent potential novel drug targets but require further validation in individual infections.

The *M. tuberculosis* genome contains a number of genes belonging to the family of polyketide synthases (PKSs), which catalyze the formation of polyketide secondary metabolites (40). The PKSs are structurally and mechanistically related to the fatty acid synthases (FASs), which are involved in the biosynthesis of fatty acids. Recent reports suggest that proteins encoded by the three-operon *fadD26-mmpL7* locus (*fadD26 ppsA-ppsE*, *drrA-drrC*, *papA5 mas fadD28 mmpL7*) play major roles in phthiocerol dimycocerosate (PDIM) biosynthetic and transport pathways, which are required for virulence (41–44). Out of 13 genes in this locus, we tested 7 genes in the current study: *fadD26*, a known positive, and *ppsA-ppsE* and *drrA*, all previously untested in mouse lungs, except for the known negative *ppsE*. While attenuation of the *fadD26* mutant was confirmed, none of the remaining genes was required for persistence in mouse lungs. Although the *drrA and drrB* genes are required for macrophage infection (3), our data suggest that they are not required for *M. tuberculosis* survival in mouse lungs.

The PKS genes *pks1*, *pks10* (45), and *pks7* (46), which are involved in dimycocerosyl phthiocerol synthesis, were reported to be required for *M. tuberculosis* persistence in mice (45, 46). In the current study, a *pks16*-deficient mutant showed reduced persistence in mouse lungs, while the *pks1*-deficient mutant showed no survival defect. The discrepancy between our findings and those of Sirakova et al. may be due to the different strains of mice (BALB/c and C57BL6/J, respectively), different routes of infection (aerosol and intranasal, respectively), different inoculating dose (10^2 and 10^4 CFU, respectively), or model system (pooled and individual infection, respectively) (45). It is unlikely that the function of the Pks1 protein was not abrogated in our mutant, since the Tn insertion is at 2,869 bp (total gene length = 4,863). Although *pks7* was previously reported to be an essential gene (2), our data are consistent with other studies demonstrating that the gene is dispensable for *in vitro* growth but essential for *M. tuberculosis* survival in mice (4).

Of the 12 *M. tuberculosis* genes designated mycobacterial membrane protein large (*mmpL1* to *mmpL12*), we studied three (*mmpL4*, *mmpL11*, *mmpL6*) and confirmed the results of earlier high-throughput screens demonstrating that the first two genes are required for long-term bacillary survival in mouse lungs (5, 41). MmpL4 and MmpL11 are predicted to serve as lipid transporters and have been shown to have a role in *M. tuberculosis* virulence in mice (47).

The genes *Rv3870* and *Rv3871*, which together with *Rv3877*

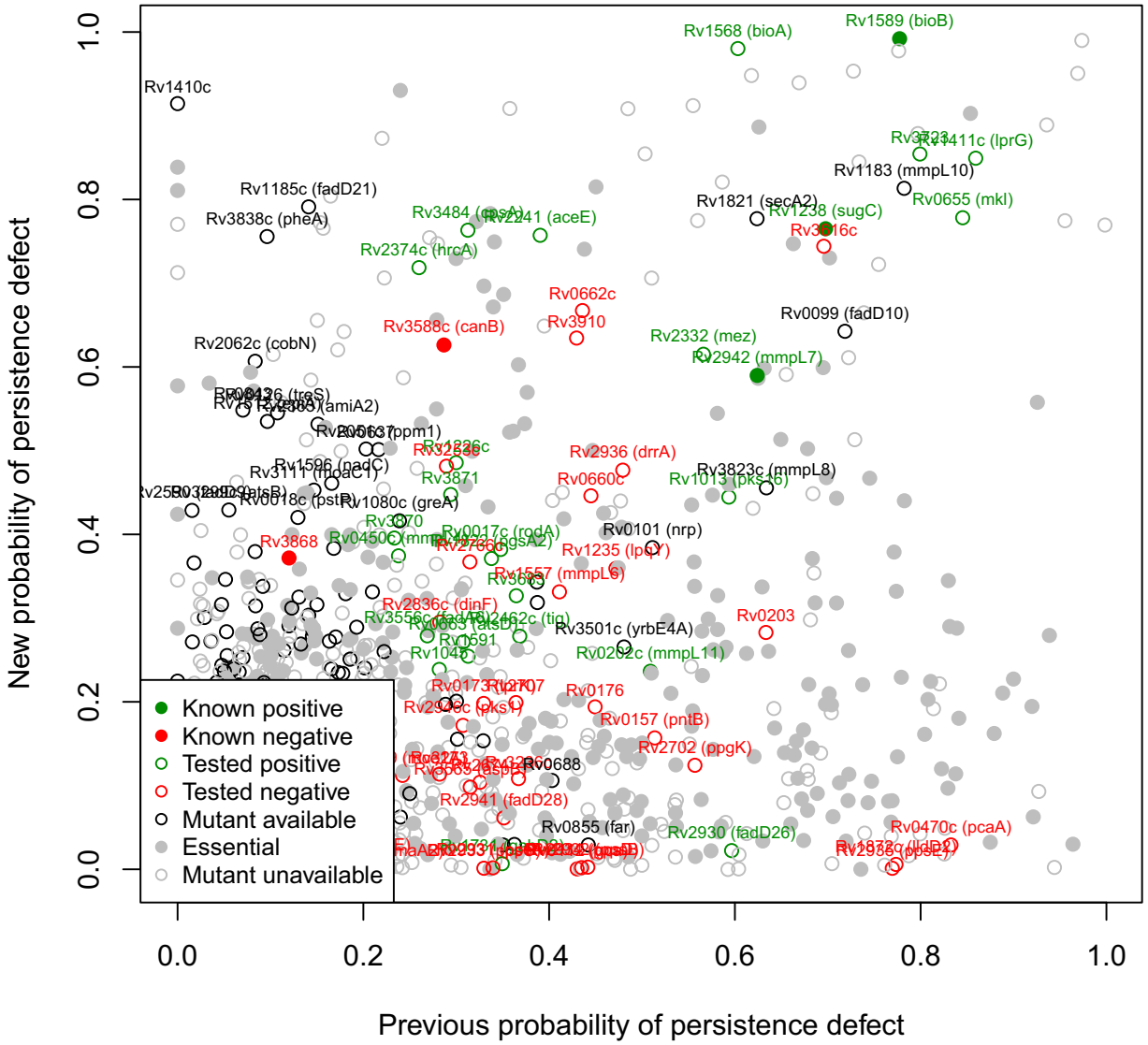

FIG 4 Predictions of probabilities of persistence defects for deletion mutants, including new results from this study (*y* axis), are compared with original predictions at the start of this study (*x* axis). Colors indicate previously known and new positive attenuated mutants (solid and open green), previously known and new negative nonattenuated mutants (solid and open red), untested mutants available for testing (black circles), and mutants unavailable for testing because they are essential (solid gray) or otherwise unavailable (open gray).

encode cytosolic or membrane-bound components of the ESX-1 secretion machinery, were found to be required for persistence in mouse lungs in the current study. Our findings are consistent with prior studies demonstrating a requirement for *Rv3871* in *M. tuberculosis* survival in murine macrophages (3) and lungs (2), as well as in nonhuman primate lungs (9). Together, these results indicate the central role for the ESX secretion pathway in *M. tuberculosis* virulence (48).

Interestingly, four mutants (*Rv0017c*::Tn, *Rv0112*::Tn, *Rv1731*::Tn, and R*v1822*::Tn) were more abundant in the mouse lungs at days 98 and 196 relative to day 49. Data for two mutants (*Rv0017c*::Tn and *Rv0112*::Tn) appear to conflict with earlier TraSH-based studies reporting that *Rv0112* is an essential gene (2) and that *Rv0017c* is required for *M. tuberculosis* survival in primary murine macrophages (3). Since the Tn insertion in our mutant, the *Rv0112*::Tn mutant, is at base pair position 91 (total gene length = 957 bp), gene function is expected to be disrupted, indicating that it is, in fact, a nonessential gene (1). The discrepancy in our findings and those of Rengarajan et al. (3) regarding *Rv0017c*, which encodes a probable cell division protein RodA, may be due to differences in models (mouse versus macrophages) or techniques used to assess mutant growth and survival (qPCR versus microarrays).

The current study demonstrates that a network-based computational approach integrating diverse high-throughput data sets may be used to predict genes essential for *M. tuberculosis* persistence in mouse lungs. These computational predictive algorithms can be further improved by iterative refinement through active learning or by including data from additional relevant model systems, *M. tuberculosis* regulatory networks (49), and operon structure. To test this hypothesis, we updated the external data by including four new essential gene data sets and updated the training

data by using the new experimental results from this study. The new experimental results highlighted three additional genes as high-priority candidates for testing. Additional rounds of experimentation and modeling could therefore lead to even greater knowledge of the genetic requirements for *M. tuberculosis* persistence. We believe future work should focus on the development of small molecule inhibitors of the most promising candidates identified through such systems biology-based approaches, with the ultimate goal of shortening the duration of TB chemotherapy.

MATERIALS AND METHODS

Network data. A functional association network for the *M. tuberculosis* H37Rv strain was obtained from the STRING database (27). A metabolic reconstruction for H37Rv (11) was converted to a functional association network using the log-likelihood ratio ρ for shared metabolites (18) and then mapped to the weight: $1/(1 + e^{2 - \rho})$. Protein-protein interactions from yeast two-hybrid screens (50) are included in STRING and did not improve performance when used as a separate feature.

Essential genes *in vitro*. Probabilities that genes are essential for *M. tuberculosis* growth in nutrient-rich broth were compiled from two random mutagenesis studies and Gibbs sampling with mutant survival data (1, 2, 51). Probabilities were recalculated using the "negenes" R-package (http://www.biostat.wisc.edu/~kbroman/software/) from current data available from the Tuberculosis Animal Research and Gene Evaluation Taskforce (TARGET) (http://webhost.nts.jhu.edu/target/).

Persistence genes *in vivo*. Genes required for *M. tuberculosis* survival in mouse tissues (persistence genes) were obtained from two previous studies (5, 6). In addition, data were extracted from a Tn mutant study in macrophages derived from C57BL/6J bone marrow with and without gamma interferon (IFN-γ) activation (3). Persistence genes from *M. tuberculosis* strain CDC 1551 were mapped to H37Rv orthologs from TubercuList (52). Scores s_g were log output pool/input pool for each gene g, and $s_g = 0$ for untested genes. The 8-week time point from the Sassetti et al. study (6) was selected as the closest match to the 49-day time point in the Lamichhane et al. study (5). Class totals for each study were

$$S_{\pm} = \sum_g \frac{|s_g \pm |s_g||}{2}$$

and normalized weights w_g were $s_g S_{tot}/S_{\pm}$ for $\pm s_g > 0$ and $S_{tot} = S_+ + S_-$.

Transcriptional profiling. Transcriptional data of *M. tuberculosis* H37Rv during infection of mouse lungs and bone marrow-derived macrophages were obtained from the TB database (53). Features, defined as positive or negative weights w_g for each gene g, were the log ratios of the transcriptional profiles obtained at 1, 2, 3, and 4 weeks (33) or 4 and 24 h (32) postinfection.

Features from graph diffusion kernels Please see Data Set S1 for a detailed description.

Classification and cross-validation performance assessment. Please see Data Set S1 for a detailed description. Software and data sets are available in the supplemental material (see Data Set S1 and reference 54).

Mutant pool generation for experimental studies. A library of 5,126 unique transposon (Tn) insertion mutants in 2,246 unique genes in CDC 1551 was generated previously (1). The top 75 genes with Tn mutants available were considered in rank order, and 67 were selected for testing. A positive control, JHU1746-380, an *in vivo* persistence mutant containing a Tn insertion in gene *Rv1746*/MT1788, and a negative control, JHU1863c-275, a fully virulent mutant containing a Tn insertion in gene *Rv1863c*/MT1912, were also added to the pool. JHU0169-511 and JHU0169-573 mutants were internal controls with Tn insertions in the same gene but at different positions (511 bp and 573, respectively). Each mutant was grown individually at 37°C in supplemented Middlebrook 7H9 medium (Difco) containing 20 μg/ml kanamycin (Sigma) to mid-log phase (optical density at 600 nm [OD_{600}] of ~0.6). The 63 different mutants in 62 unique genes were pooled by combining an equal volume of each strain.

Mouse infection. All procedures involving animals were performed in compliance with the U.S. Animal Welfare Act regulations and Public Health Service Policy according to protocols approved by the Institutional Animal Care and Use Committee at Johns Hopkins University. All mice were maintained and bred under specific-pathogen-free conditions and fed water and chow *ad libitum*. Female BALB/c mice (5 to 6 weeks old; Charles River) were infected via the aerosol route using an inhalation exposure system (Glas-Col) with 2 $\log_{10}$ bacilli. Five mice per group were sacrificed at days 1, 49, 98, and 196 postinfection. Both lungs were homogenized in phosphate-buffered saline (PBS), plated on supplemented Middlebrook 7H10 solid medium (Difco) containing 20 μg of kanamycin/ml, and incubated at 37°C at least 3 weeks before colony enumeration or DNA extraction.

Real-time PCR. For each time point, approximately 1,000 colonies were scraped and pooled, and genomic DNA (gDNA) was prepared (4, 5, 7). The gDNA preparations from each experimental group were pooled, and qPCR was performed in duplicate using iCycler iQ (version 3.1.7050; Bio-Rad). Mutant-specific primer sets, each composed of a generic Tn primer and a gene-specific primer, were designed to amplify 150- to 200-bp DNA fragments and validated by amplifying the correct-sized fragment by conventional PCR. For a given qPCR run, the cycle threshold (C_T) for Tn mutant g is $C_T(g)$ and for the housekeeping gene *sigA* is $C_T(h)$. The difference $C_T(g) - C_T(h)$ is $\Delta C_T(gtr)$, where g labels the mutant, t labels the four time points (day 1, 49, 98, or 196), and r labels the technical replicate (1 or 2). Finally, y_{gt} is the average of the replicates: $y_{gt} = [\Delta C_T(gt1) + \Delta C_T(gt2)]/2$. A detailed description of the qPCR data analysis is provided in Data Set S1. Software, data, and expectation-maximization detailed methods are available in the supplemental material (see Data Set S1).

New predictions based on additional experimental data sets and new experimental results. We collected essentiality data sets from four papers published after the initial selection of candidates for testing (28–31). Three of these new data sets rely on improved experimental methods using next-generation sequencing to identify TA sites lacking transposon insertions. Different methods characterize essential genes based on the number of consecutive TA sites without observed insertions (29) or identify overlapping genome regions lacking transposon insertions and then identify genes overlapping these essential regions (29, 31). New Bayesian methods using extreme value distributions to describe runs of TA sites have also been applied to estimate posterior probabilities of essentiality for each gene (28, 29). In addition to these experimental approaches, a recent computational method employed a metabolic reconstruction and flux balance analysis (FBA) to identify essential metabolic genes (30). These data sets generally identify 700 genes overall as essential, of which about 200 are metabolic (see Data Set S5A). These four data sets were incorporated as additional essential gene features and propagated through the biological networks using graph diffusion kernels.

New predictions also relied on updated "attenuated" and "nonattenuated" gene labels according to the new results for mutants tested experimentally. Mutants found to be virulent were labeled as nonattenuated.

We generated new predictions in two stages: first, we included just the new external data; then, we also included the updated gene labels. These predictions used the same methods as described for the original set of predictions used to prioritize genes for testing (see Data Set S6).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01066-13/-/DCSupplemental.

Data Set S1, DOCX file, 0 MB.
Data Set S2, XLSX file, 0.9 MB.
Data Set S3, PPTX file, 0.1 MB.
Data Set S4, DOCX file, 0.2 MB.
Data Set S5, DOCX file, 0 MB.
Data Set S6, XLSX file, 0.3 MB.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (AI064229 and AI083125 to P.C.K. and HL106786 to J.S.B. and P.C.K.) and by the Robert J. and Helen C. Kleberg Foundation (to J.S.B.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

J.S.B. and P.C.K. conceived the experiment. N.B., B.V., and J.S.B. developed the prediction model. G.L. contributed materials and assistance with experimental methods. N.K.D. performed the experiments. All authors contributed to analyzing the results and writing the manuscript.

REFERENCES

1. **Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR.** 2003. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. **100:** 7213–7218. http://dx.doi.org/10.1073/pnas.1231432100.
2. **Sassetti CM, Boyd DH, Rubin EJ.** 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. **48:** 77–84. http://dx.doi.org/10.1046/j.1365-2958.2003.03425.x.
3. **Rengarajan J, Bloom BR, Rubin EJ.** 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. Proc. Natl. Acad. Sci. U. S. A. **102:**8327–8332. http://dx.doi.org/10.1073/pnas.0503272102.
4. **Be NA, Lamichhane G, Grosset J, Tyagi S, Cheng QJ, Kim KS, Bishai WR, Jain SK.** 2008. Murine model to study the invasion and survival of *Mycobacterium tuberculosis* in the central nervous system. J. Infect. Dis. **198:**1520–1528. http://dx.doi.org/10.1086/592447.
5. **Lamichhane G, Tyagi S, Bishai WR.** 2005. Designer arrays for defined mutant analysis to detect genes essential for survival of *Mycobacterium tuberculosis* in mouse lungs. Infect. Immun. **73:**2533–2540. http://dx.doi.org/10.1128/IAI.73.4.2533-2540.2005.
6. **Sassetti CM, Rubin EJ.** 2003. Genetic requirements for mycobacterial survival during infection. Proc. Natl. Acad. Sci. U. S. A. **100:**12989–12994. http://dx.doi.org/10.1073/pnas.2134250100.
7. **Jain SK, Hernandez-Abanto SM, Cheng QJ, Singh P, Ly LH, Klinkenberg LG, Morrison NE, Converse PJ, Nuermberger E, Grosset J, McMurray DN, Karakousis PC, Lamichhane G, Bishai WR.** 2007. Accelerated detection of *Mycobacterium tuberculosis* genes essential for bacterial survival in guinea pigs, compared with mice. J. Infect. Dis. **195:** 1634–1642. http://dx.doi.org/10.1086/517526.
8. **Klinkenberg LG, Sutherland LA, Bishai WR, Karakousis PC.** 2008. Metronidazole lacks activity against *Mycobacterium tuberculosis* in an *in vivo* hypoxic granuloma model of latency. J. Infect. Dis. **198:**275–283. http://dx.doi.org/10.1086/589515.
9. **Dutta NK, Mehra S, Didier PJ, Roy CJ, Doyle LA, Alvarez X, Ratterree M, Be NA, Lamichhane G, Jain SK, Lacey MR, Lackner AA, Kaushal D.** 2010. Genetic requirements for the survival of tubercle bacilli in primates. J. Infect. Dis. **201:**1743–1752. http://dx.doi.org/10.1086/652497.
10. **Bordbar A, Lewis NE, Schellenberger J, Palsson BØ, Jamshidi N.** 2010. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. Mol. Syst. Biol. **6:**422. http://dx.doi.org/10.1038/msb.2010.68.
11. **Jamshidi N, Palsson BØ.** 2007. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. BMC Syst. Biol. **1:**26. http://dx.doi.org/10.1186/1752-0509-1-26.
12. **Orth JD, Thiele I, Palsson BØ.** 2010. What is flux balance analysis? Nat. Biotechnol. **28:**245–248.
13. **Chandrasekaran S, Price ND.** 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. **107:** 17845–17850. http://dx.doi.org/10.1073/pnas.1005139107.
14. **Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE.** 2009. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput. Biol. **5:**e1000489. http://dx.doi.org/10.1371/journal.pcbi.1000489.
15. **Sharan R, Ulitsky I, Shamir R.** 2007. Network-based prediction of protein function. Mol. Syst. Biol. **3:**88. http://dx.doi.org/10.1038/msb4100129.
16. **Raman K, Chandra N.** 2008. *Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance. BMC Microbiol. **8:**234. http://dx.doi.org/10.1186/1471-2180-8-234.
17. **Qi Y, Suhail Y, Lin YY, Boeke JD, Bader JS.** 2008. Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. Genome Res. **18:**1991–2004. http://dx.doi.org/10.1101/gr.077693.108.
18. **Veeramani B, Bader JS.** 2010. Predicting functional associations from metabolism using bi-partite network algorithms. BMC Syst. Biol. **4:**95. http://dx.doi.org/10.1186/1752-0509-4-95.
19. **Tsuda K, Noble WS.** 2004. Learning kernels from biological networks by maximizing entropy. Bioinformatics **20**(Suppl 1):i326–i333. http://dx.doi.org/10.1093/bioinformatics/bth906.
20. **Köhler S, Bauer S, Horn D, Robinson PN.** 2008. Walking the interactome for prioritization of candidate disease genes. Am. J. Hum. Genet. **82:**949–958. http://dx.doi.org/10.1016/j.ajhg.2008.02.013.
21. **Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R.** 2010. Associating genes and protein complexes with disease via network propagation. PLoS Comput. Biol. **6:**e1000641. http://dx.doi.org/10.1371/journal.pcbi.1000641.
22. **Lu LJ, Xia Y, Paccanaro A, Yu H, Gerstein M.** 2005. Assessing the limits of genomic data integration for predicting protein networks. Genome Res. **15:**945–953. http://dx.doi.org/10.1101/gr.3610305.
23. **Qi Y, Klein-Seetharaman J, Bar-Joseph Z.** 2007. A mixture of feature experts approach for protein-protein interaction prediction. BMC Bioinformatics **8**(Suppl 10):S6. http://dx.doi.org/10.1186/1471-2105-8-S10-S6.
24. **Qiu J, Noble WS.** 2008. Predicting co-complexed protein pairs from heterogeneous data. PLoS Comput. Biol. **4:**e1000054. http://dx.doi.org/10.1371/journal.pcbi.1000054.
25. **Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R.** 2009. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res. **69:**6660–6667. http://dx.doi.org/10.1158/0008-5472.CAN-09-1133.
26. **Schellenberger J, Park JO, Conrad TM, Palsson BØ.** 2010. BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics **11:**213. http://dx.doi.org/10.1186/1471-2105-11-213.
27. **Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C.** 2009. String 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. **37:**D412–D416. http://dx.doi.org/10.1093/nar/gkn760.
28. **DeJesus MA, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, Ioerger TR.** 2013. Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. Bioinformatics **29:**695–703. http://dx.doi.org/10.1093/bioinformatics/btt043.
29. **Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM.** 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. **7:**e1002251. http://dx.doi.org/10.1371/journal.ppat.1002251.
30. **Lofthouse EK, Wheeler PR, Beste DJ, Khatri BL, Wu H, Mendum TA, Kierzek AM, McFadden J.** 2013. Systems-based approaches to probing metabolic variation within the mycobacterium tuberculosis complex. PLoS One **8:**e75913. http://dx.doi.org/10.1371/journal.pone.0075913.
31. **Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, Rubin EJ.** 2012. Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. PLoS Pathog. **8:**e1002946. http://dx.doi.org/10.1371/journal.ppat.1002946.
32. **Fontán PA, Aris V, Alvarez ME, Ghanny S, Cheng J, Soteropoulos P, Trevani A, Pine R, Smith I.** 2008. *Mycobacterium tuberculosis* sigma factor E regulon modulates the host inflammatory response. J. Infect. Dis. **198:**877–885. http://dx.doi.org/10.1086/591098.
33. **Talaat AM, Lyons R, Howard ST, Johnston SA.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc. Natl. Acad. Sci. U. S. A. **101:**4602–4607. http://dx.doi.org/10.1073/pnas.0306023101.
34. **Bloch H, Segal W.** 1956. Biochemical differentiation of *Mycobacterium tuberculosis* grown *in vivo* and *in vitro*. J. Bacteriol. **72:**132–141.
35. **Dahl JL, Kraus CN, Boshoff HI, Doan B, Foley K, Avarbock D, Kaplan G, Mizrahi V, Rubin H, Barry CE, III.** 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *My-*

cobacterium tuberculosis in mice. Proc. Natl. Acad. Sci. U. S. A. **100:** 10026–10031. http://dx.doi.org/10.1073/pnas.1631248100.
36. **Primm TP, Andersen SJ, Mizrahi V, Avarbock D, Rubin H, Barry CE, III.** 2000. The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. J. Bacteriol. **182:**4889–4898. http://dx.doi.org/10.1128/JB.182.17.4889-4898.2000.
37. **Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K.** 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol. Microbiol. **43:** 717–731. http://dx.doi.org/10.1046/j.1365-2958.2002.02779.x.
38. **Hampshire T, Soneji S, Bacon J, James BW, Hinds J, Laing K, Stabler RA, Marsh PD, Butcher PD.** 2004. Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: a model for persistent organisms? Tuberculosis **84:**228–238. http://dx.doi.org/10.1016/j.tube.2003.12.010.
39. **Rifat D, Bishai WR, Karakousis PC.** 2009. Phosphate depletion: a novel trigger for *Mycobacterium tuberculosis* persistence. J. Infect. Dis. **200:** 1126–1135. http://dx.doi.org/10.1086/605700.
40. **Gokhale RS, Saxena P, Chopra T, Mohanty D.** 2007. Versatile polyketide enzymatic machinery for the biosynthesis of complex mycobacterial lipids. Nat. Prod. Rep. **24:**267–277. http://dx.doi.org/10.1039/b616817p.
41. **Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C.** 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol. Microbiol. **34:**257–267. http://dx.doi.org/10.1046/j.1365-2958.1999.01593.x.
42. **Cox JS, Chen B, McNeil M, Jacobs WR, Jr..** 1999. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature **402:**79–83. http://dx.doi.org/10.1038/47042.
43. **Pinto R, Saunders BM, Camacho LR, Britton WJ, Gicquel B, Triccas JA.** 2004. *Mycobacterium tuberculosis* defective in phthiocerol dimycocerosate translocation provides greater protective immunity against tuberculosis than the existing Bacille Calmette-Guerin vaccine. J. Infect. Dis. **189:** 105–112. http://dx.doi.org/10.1086/380413.
44. **Rousseau C, Winter N, Pivert E, Bordat Y, Neyrolles O, Avé P, Huerre M, Gicquel B, Jackson M.** 2004. Production of phthiocerol dimycocerosates protects *Mycobacterium tuberculosis* from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. Cell. Microbiol. **6:**277–287. http://dx.doi.org/10.1046/j.1462-5822.2004.00368.x.
45. **Sirakova TD, Dubey VS, Cynamon MH, Kolattukudy PE.** 2003. Attenuation of *Mycobacterium tuberculosis* by disruption of a *mas*-like gene or a chalcone synthase-like gene, which causes deficiency in dimycocerosyl phthiocerol synthesis. J. Bacteriol. **185:**2999–3008. http://dx.doi.org/10.1128/JB.185.10.2999-3008.2003.
46. **Rousseau C, Sirakova TD, Dubey VS, Bordat Y, Kolattukudy PE, Gicquel B, Jackson M.** 2003. Virulence attenuation of two Mas-like polyketide synthase mutants of *Mycobacterium tuberculosis*. Microbiology **149:**1837–1847. http://dx.doi.org/10.1099/mic.0.26278-0.
47. **Domenech P, Reed MB, Barry CE, III.** 2005. Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect. Immun. **73:**3492–3501. http://dx.doi.org/10.1128/IAI.73.6.3492-3501.2005.
48. **Ramage HR, Connolly LE, Cox JS.** 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. PLoS Genet. **5:**e1000767. http://dx.doi.org/10.1371/journal.pgen.1000767.
49. **Balázsi G, Heath AP, Shi L, Gennaro ML.** 2008. The temporal response of the *Mycobacterium tuberculosis* gene regulatory network during growth arrest. Mol. Syst. Biol. **4:**225.
50. **Wang Y, Cui T, Zhang C, Yang M, Huang Y, Li W, Zhang L, Gao C, He Y, Li Y, Huang F, Zeng J, Huang C, Yang Q, Tian Y, Zhao C, Chen H, Zhang H, He ZG.** 2010. Global protein-protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. J. Proteome Res. **9:**6665–6677. http://dx.doi.org/10.1021/pr100808n.
51. **Geman S, Geman D.** 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6:**721–741.
52. **Lew JM, Kapopoulou A, Jones LM, Cole ST.** 2011. TubercuList—10 years after. Tuberculosis **91:**1–7. http://dx.doi.org/10.1016/j.tube.2010.09.008.
53. **Galagan JE, Sisk P, Stolte C, Weiner B, Koehrsen M, Wymore F, Reddy TB, Zucker JD, Engels R, Gellesch M, Hubble J, Jin H, Larson L, Mao M, Nitzberg M, White J, Zachariah ZK, Sherlock G, Ball CA, Schoolnik GK.** 2010. TB database 2010: overview and update. Tuberculosis **90:** 225–235. http://dx.doi.org/10.1016/j.tube.2010.03.010.
54. **Sing T, Sander O, Beerenwinkel N, Lengauer T.** 2005. ROCR: visualizing classifier performance in R. Bioinformatics **21:**3940–3941. http://dx.doi.org/10.1093/bioinformatics/bti623.