Supplementary Tables.

Drug	Abbreviation	Tier 1 Genes
Isoniazid	INH	*fabG1 (Rv1483), inhA (Rv1484), katG (Rv1908c), furA (Rv1909c), ahpC (Rv2428)*
Rifampicin	RIF	*rpoB (Rv0667)*
Ethambutol	EMB	*aftA (Rv3792), embC (Rv3793), embA (Rv3794), embB (Rv3795), ubiA (Rv3806c)*
Pyrazinamide	PZA	*pncA (Rv2043c), clpC1 (Rv3596c), panD (Rv3601c)*
Levofloxacin	LFX	*gyrB (Rv0005), gyrA (Rv0006)*
Moxifloxacin	MXF	*gyrB (Rv0005), gyrA (Rv0006)*
Bedaquiline	BDQ	*mmpL5 (Rv0676c), mmpS5 (Rv0677c), Rv0678, atpE (Rv1305), pepQ (Rv2535c)*
Clofazimine	CFZ	*mmpL5, mmpS5, Rv0678, Rv1979c, pepQ*
Linezolid	LZD	*rplC (Rv0701), rrl (MTB000020)*
Delamanid	DLM	*fgd1, ddn (Rv3547), fbiC, Rv2983, fbiA, fbiB*
Ethionamide	ETO	*mshA, fabG1, inhA, ethA (Rv3854c)*
Amikacin	AMK	*rrs (MTB000019), eis (Rv2416c), whiB7 (Rv3197A)*
Streptomycin	STM	*rpsL (Rv0682), Rv1258c, rrs, whiB7, gid (Rv3919c)*
Kanamycin	KAN	*rrs, eis, whiB7*
Capreomycin	CAP	*rrs, tlyA (Rv1694)*

Supplementary Table 1. Drugs and Tier 1 Genes. Variants in these genes and the upstream promoter regions were used as input into the regression models. Adapted from Table 21 in the 2nd edition of the WHO mutation catalogue.

	Genotypes (Tier 1 genes only)		
Model	Non-silent variants	LoF pooling	Silent variants
base	yes	no	no
pool	yes	yes	no
base + silent	yes	no	yes

Supplementary Table 2. Nested regression models. Nine models were built for each drug using a combination of variant types included in the model (columns) and data subgroups with different phenotype quality criteria and availability (rows). Each model was fit on each set of phenotypes ("WHO dataset," "ALL dataset," and MICs). Red text highlights the model used for grading each specific group of genotypes. Full model sizes, with numbers of isolates and variants, are in **Supplementary Data 1**. The MIC models were not used for grading; they were only used to add supportive evidence for associations found in the binary drug resistance datasets.

Row	**WHO Grading**	**ALL Grading**	**Final Grading**	**Reason**	**# of Variants**
1	Assoc w R	Assoc w R	1) Assoc w R	Strong evidence from both datasets in agreement.	587
2	Not assoc w R, Neutral	Not assoc w R, Neutral	5) Not assoc w R	Strong evidence from both datasets in agreement.	66

3	Neutral	Uncertain	5) Not assoc w R	Neutral variants in the WHO dataset.	28
4	Uncertain*	Neutral	4) Not assoc w R - Interim	Prioritize Neutrals from the WHO dataset, so grade these Interim.	53
5	Assoc w R	Uncertain	3) Uncertain significance	ALL grading adds uncertainty to WHO grading. Downgrade to Uncertain to avoid overcalling resistance-associated variants due to potential sampling bias.	12
6	Not assoc w R	Uncertain	4) Not assoc w R - Interim	ALL grading adds uncertainty to WHO grading. Downgrade to interim to be less strict for Not assoc w R variants.	16
7	Uncertain	Assoc w R	2) Assoc w R - Interim	Evidence only from the ALL dataset, so grade these interim.	149
8	Uncertain	Not assoc w R	4) Not assoc w R - Interim	Evidence only from the ALL dataset, so grade these interim.	56
9	Assoc w R	Not assoc w R, Neutral	3) Uncertain significance	Discrepant directions of estimated statistical effects	0

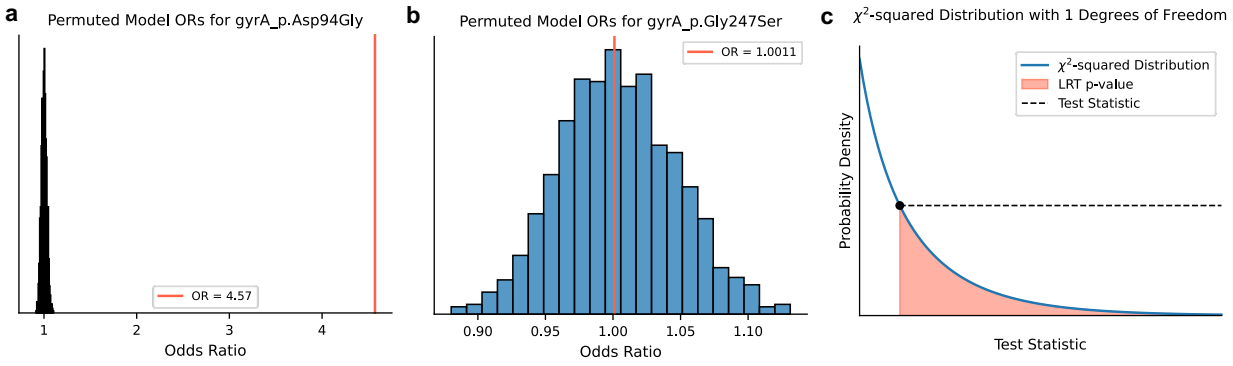
10	Not assoc w R, Neutral	Assoc w R			0

Supplementary Table 3. WHO-ALL Grading Integration. The number of variants across the dataset of 21,589 unique (drug, variant) pairs that meet these criteria are shown in the "# of Variants" column. If a variant is absent from a dataset, it is graded "Uncertain." *: Must be present in the WHO dataset; to upgrade variants that are called "Neutral" only in the ALL dataset to 4) Not assoc w R - Interim, they must not be absent from the WHO dataset. There were no cases when a variant was graded "Neutral" or "Not assoc w R" in one dataset and "Assoc w R" in the other dataset.

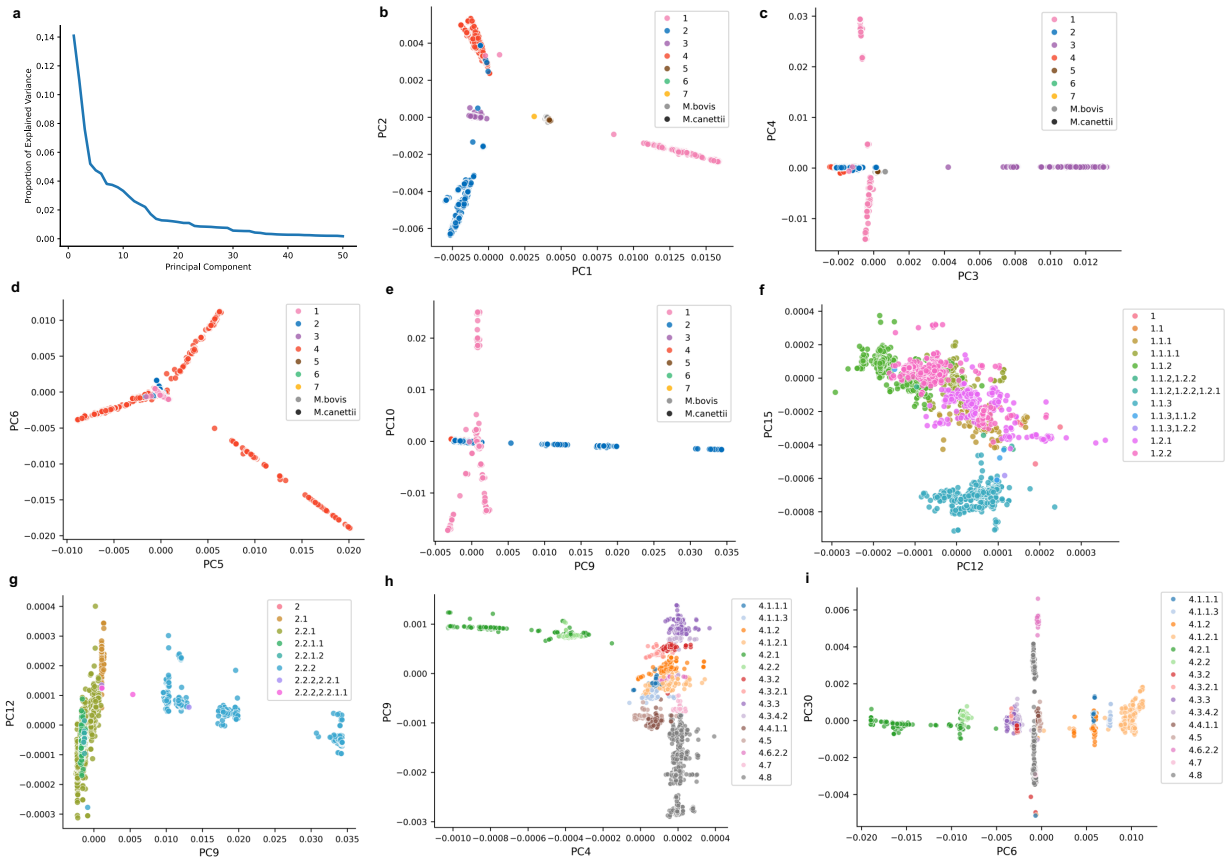
Ordered Media List	**Type**
7H10 > LJ > 7H11	Solid
MGIT > MODS > BACTEC™ 460 > 7H9	Liquid
Frozen Broth Microdilution Plate (PMID31969421) > UKMYC > REMA > MYCOTB > MABA > MABA24 > MABA48 > non-colorimetric > M24 BMD	Plate

Supplementary Table 4. MIC Media Hierarchy. Prioritization of MIC testing media for the regression models. For isolates with MICs tested in multiple media, the media with the highest priority rating was kept. The order is based on media prioritization in sections 5.4-5.5 in the 2nd edition of the mutation catalogue.

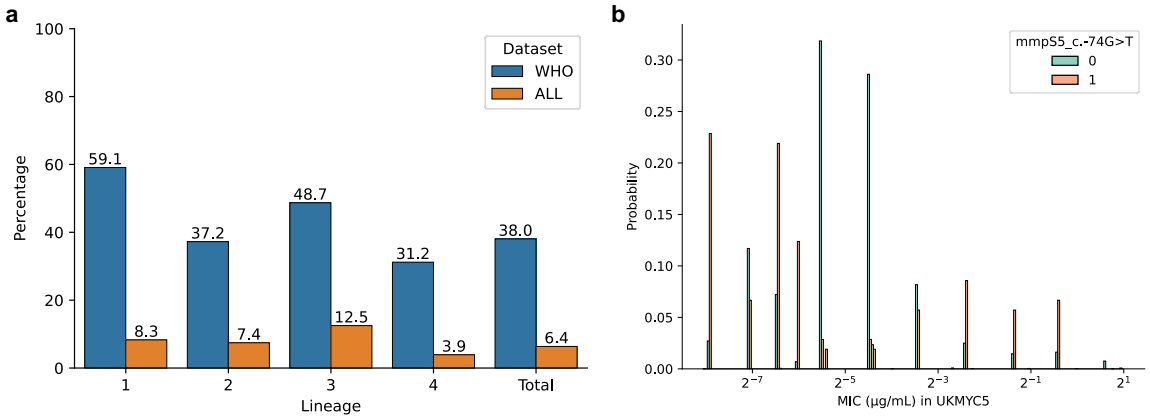
Supplementary Figures.



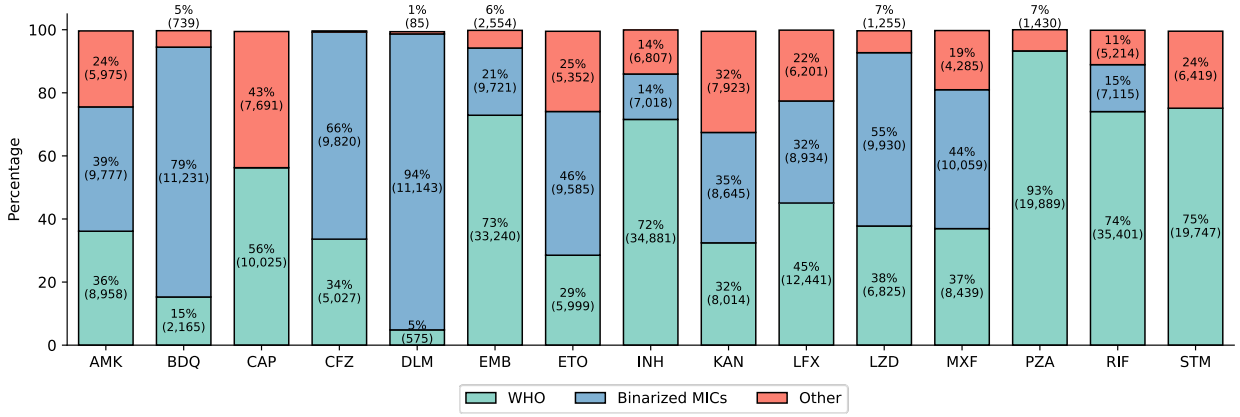
<u>Supplementary Figure 1.</u> Examples of results used to perform the permutation tests and LRT. a: Odds ratios for gyrA_p.Asp94Gly (MXF), which is R-associated, in the 1,000 permuted models. **b:** Odds ratios for gyrA_p.Gly247Ser (LFX), a neutral masked variant by SOLO, in the 1,000 permuted models. **c:** Example of chi-squared distribution with one degree of freedom (dof) and the area corresponding with the p-value of the LRT. Larger test statistics result in smaller p-values. The dof for the LRT as applied here is one because each time, one variant is removed to test the change in model goodness-of-fit.



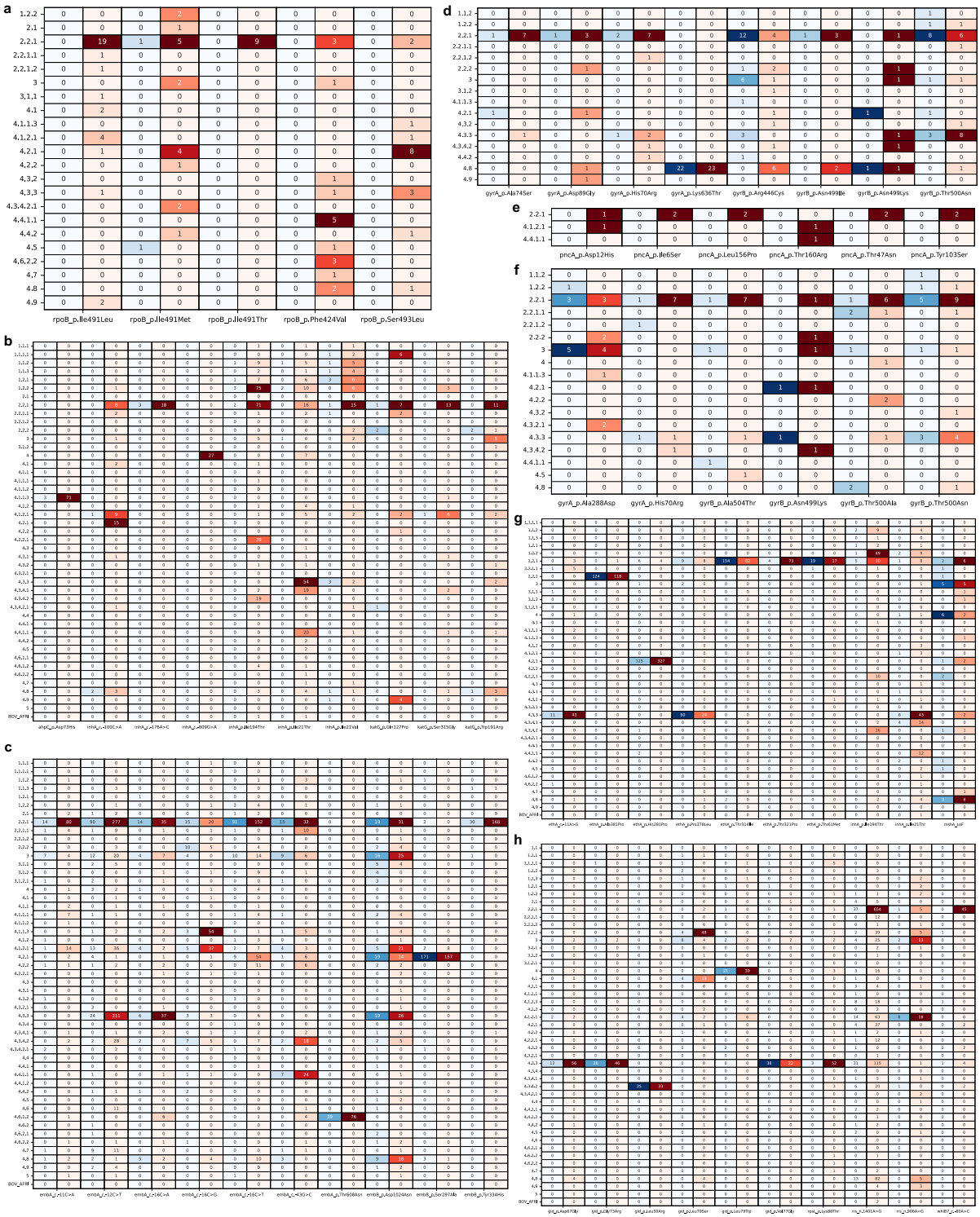
<u>Supplementary Figure 2.</u> Principal components analysis of SNPs to correct for lineage structure in regression models. a: Plot of the first 50 principal components and the genetic variation each one represents. Combined, they represent 94.5% of the genetic variation in the dataset. **b-e:** First several principal components, colored by MTBC lineage to show separation of the primary lineages. High principal components show separation between sublineages of L1 (N = 4,607) **(f)** and L2 (N = 16,801) **(g). h-i:** Selected principal components highly correlated with sublineages of L4, showing the diversity of this MTBC lineage. 24,149 isolates have a single L4 sublineage based on the Coll 2014 scheme. For readability, only sublineages that make up at least 2% of this group were plotted, leaving 20,346 isolates across 15 L4 sublineages. Only isolates with a single lineage (but not necessarily a single sublineage) assigned by the Coll *et al.,* 2014 scheme (N = 52,172, 99.25%) are shown here. All lineage classifications are in **Supplementary Table 4.**



<u>Supplementary Figure 3.</u> Bedaquiline resistance is significantly enriched in the WHO dataset over the ALL dataset. a: Percentage of Bedaquiline-resistant isolates (pDST data only), stratified by lineage and phenotypic group. Only isolates with a single primary lineage group assigned are shown here. There are no L5 and *M. bovis* isolates that are BDQ-susceptible. **b:** Distribution of 12,379 available Bedaquiline MICs, stratified by the presence of mmpS5_c.-74G>T (0 = absent, 1 = present). MICs were normalized to the UKMYC5 scale because it is the most common medium in the dataset. The counts of isolates with and without mmpS5_c.-74G>T are normalized separately to better view the much smaller set of isolates that have the variant (N=105). The UKMYC critical concentration for BDQ is 0.25 µg/mL.

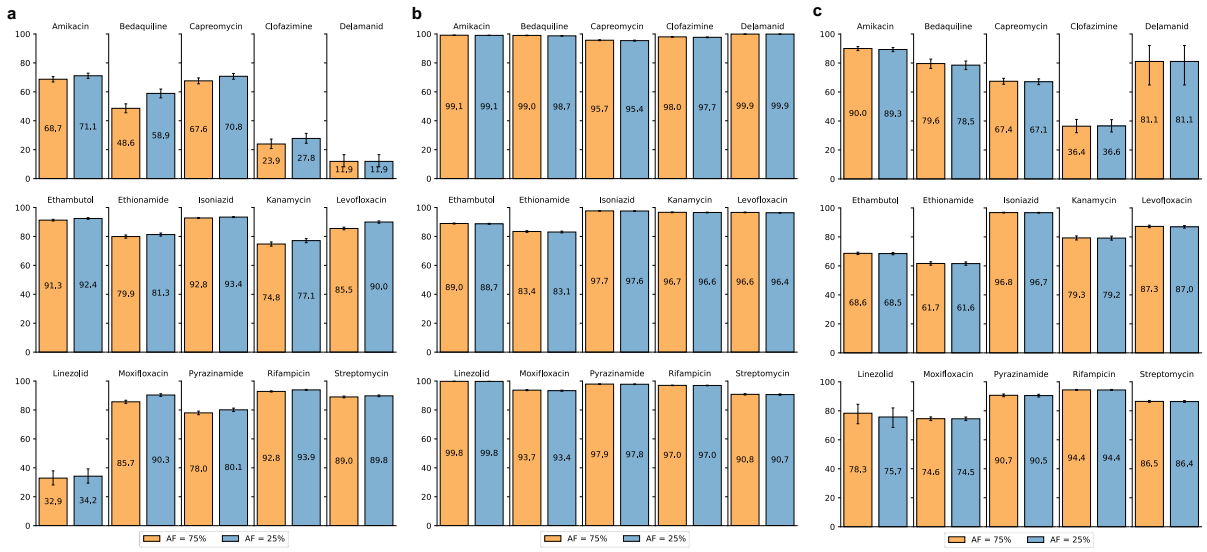


<u>Supplementary Figure 4.</u> Composition of the ALL dataset for each drug. The ALL dataset is composed of WHO-approved pDST results ("WHO dataset"), binarized MICs, and other pDST results not tested using WHO-approved methods. All phenotypic results are shown here, not just those that were included in the regression models. Counts are labeled on the bars, with percentages of each drug's ALL dataset in parentheses. All binarized MICs are part of the ALL dataset. There are no binarized MICs for Capreomycin, Pyrazinamide, or Streptomycin.



<u>Supplementary Figure 5.</u> Lineage and phenotypic distributions for variants graded in Groups 1-2 by regression and Uncertain by SOLO + GR -- RIF **(a),** INH **(b),** EMB **(c),** LFX **(d),** PZA **(e),** MXF **(f),** ETO **(g),** and STM **(h).** For readability, only the 10 most common variants are

shown, and *ahpC* promoter variants covered by the Xpert MTB/XDR test for INH are excluded. The lineage distribution is plotted in two columns for each variant: the left column is drug-susceptible isolates (blue), and the right column is drug-resistant isolates (red).



<u>Supplementary Figure 6.</u> Comparison of binary prediction metrics for regression between with different AF thresholds. Sensitivity **(a)**, specificity **(b)**, and PPV **(c)** between AF = 75% (orange) and 25% (blue) as the exclusive cutoff for variant occurrence. Only results for the regression mutation list are shown here. Bars are the computed sensitivity, specificity, and PPV, as percent, for each drug and model type. Error bars are 95% exact binomial confidence intervals computed using the Clopper-Pearson method. Source data are in **Supplementary Data 9**, including the dataset sizes.