Description of Additional Supplementary Files

File Name: Supplementary Data 1
Description: Isolate IDs, lineages and F2 strain mixing metrics, and MICs for 8 drugs, including the bedaquiline augmented and pyrazinamide binary datasets, which are separate sheets. All MICs are in 7H10 medium, except for bedaquiline (7H11) and pyrazinamide (MGIT), and both the MIC bounds and midpoint are listed. The train-validation-test splits are also given. Each sheet corresponds with a single drug dataset. Samples that have since been made publicly available have a public ID in the BIOSAMPLE_ACCESSION column.

File Name: Supplementary Data 2
Description: List of data contributors, affiliations, and email addresses from the MIC-ML consortium, who provided as-yet unpublished genotypic and phenotypic data for model training. Some of the genotypic data are now public (see **Supplementary Data 1**).

File Name: Supplementary Data 3
Description: Model performances for 5 cross-validation splits for both CNN and L2 regression for all drugs, with and without lineage SNPs, amino acid properties (without amino acid properties only for pyrazinamide), additional tier 2 loci (ethionamide only), and augmentation (bedaquiline only). Both binned mean error and mean error computed relative to MIC midpoints are listed for each quantitative model. The binary pyrazinamide model results are also in this table (Model = “Binary CNN” or “LogReg”) with AUCs.

File Name: Supplementary Data 4
Description: In silico MIC predictions for mutations from the WHO catalog. Sheet 1: all predictions for quantitative models. Sheet 2: *In silico* predictions for mutations from the WHO catalog by the binary pyrazinamide model. Predictions are resistance probabilities, which have been binarized to a resistant vs. susceptible label. Sheet 3: mutations not associated with resistance based on literature curation for the 2nd version of the WHO mutation catalog. Sheet 3: 28 mutations with literature evidence against association with resistance. Sheet 4: 82 Group 1 mutations with predicted MICs below the respective critical concentrations. Sheet 5: 25 mutations associated with low-level resistance to isoniazid, moxifloxacin, and rifampicin.

File Name: Supplementary Data 5
Description: 247 nucleotide sites across all 8 drugs (244 unique sites due to locus redundancy across some drugs) newly associated with resistance (in the 70th score percentile or higher within a single model, occur in at least 5 isolates, and are not in the WHO catalog.

File Name: Supplementary Data 6
Description: Binary classification metrics (sensitivity, specificity, and F1 score) between the WHO catalog classification method and the quantitative CNNs (or binary, in the case of pyrazinamide). The “_LB” and “_UB” suffixes indicate lower and upper bounds of 95% exact binomial confidence intervals computed using the Clopper-Pearson method.

File Name: Supplementary Data 7
Description: Site-saturation mutagenesis results for rpoB, gyrA, gyrB, pncA, katG, ethA, and Rv0678 and Getis-Ord clustering scores (“G_score” column) for average predicted $\log_2$FC per site. For gyrA and gyrB, G scores for only residues solved in the crystal structure were computed. “BH_pval” = FDR-corrected p-value using Benjamini-Hochberg correction. “clustering_result” is 1 for hotspots, -1 for coldspot, and NA for all else. “average” is the average predicted $\log_2$FC for each codon.

File Name: Supplementary Data 8
Description: All variants in genes graded in the 2023 WHO MTBC mutation catalog in the samples from participants in the TRUST study. Quality control fields output by the pilon variant calling software – FILTER, IMPRECISE, AF (allele fraction), DP (read depth), BQ (base quality), MQ (mapping quality), IC (reads supporting an insertion), and DC (reads supporting a deletion) – are included.