Supplemental Tables

Fusing Dual-Event Datasets for Mycobacterium Tuberculosis Machine Learning Models and their Evaluation

Sean Ekins[†,‡*], Joel S. Freundlich[§,‖] and Robert C. Reynolds[⊥]

[†] Collaborative Drug Discovery, 1633 Bayshore Highway, Suite 342, Burlingame, CA 94010, USA.

[‡] Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA.

[§] Department of Medicine, Center for Emerging and Reemerging Pathogens, Rutgers University – New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA.

[‖] Department of Pharmacology & Physiology, Rutgers University – New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA.

[⊥] University of Alabama at Birmingham, College of Arts and Sciences, Department of Chemistry, 1530 3rd Avenue South, Birmingham, AL 35294-1240, USA.

*To whom correspondence should be addressed. E-mail: ekinssean@yahoo.com

Running title: Fusing datasets for *Mtb*

Supplemental Table 1. Mean (± sd) leave-one-out and leave-out 50% x 100 cross-validation of Bayesian models (ROC = receiver operator characteristic). Best Split for ARRA = -2.292, best split for combined = -0.712.

Dataset (number of molecules)	Leave-one-out ROC	Leave-out 50% x 100 External ROC Score	Leave-out 50% x 100 Internal ROC Score	Leave-out 50% x 100 Concordance	Leave-out 50% x 100 Specificity	Leave-out 50% x 100 Sensitivity
ARRA dose response & cytotoxicity (1924)	0.87	0.85 ± 0.02	0.86 ± 0.02	80.13 ± 5.32	80.31 ±5.85	75.85 ± 8.74
Combined MLSMR, TAACF-CB2	0.71	0.74 ± 0.01	0.75 ± 0.01	69.39 ± 1.98	70.79 ± 3.85	65.34 ± 4.1

and TAACF Kinase dose response and cytotoxicity (5304)						

Supplemental Table 2. GSK Mtb hits scored – see excel file

Supplemental Table 3. Clustering GSK *Mtb* hits using MDL fingerprints to infer *Mtb* targets – see excel file.