

Supplemental Figures

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.

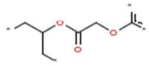
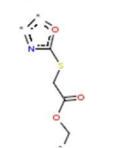
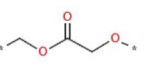
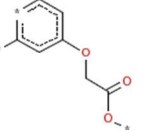
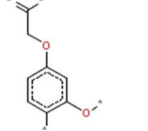
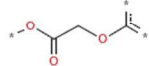
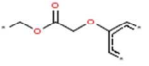
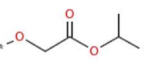
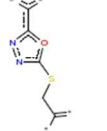
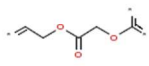
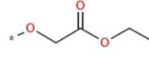
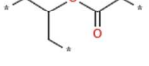
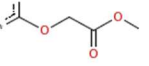
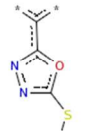
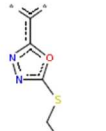
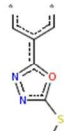
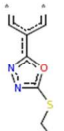
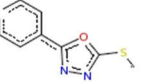
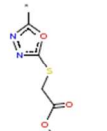
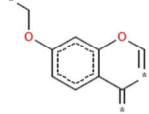
^{||} Department of Pharmacology & Physiology, Rutgers University – New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA.

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

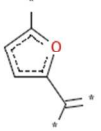
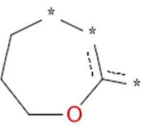
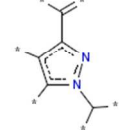
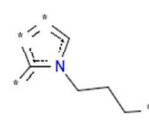
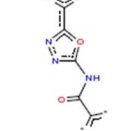
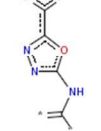
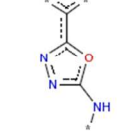
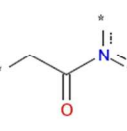
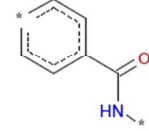
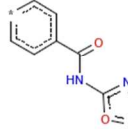
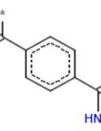
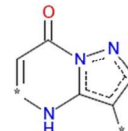
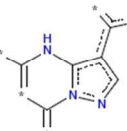
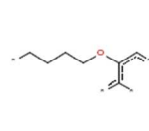
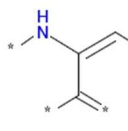
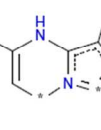
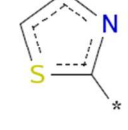
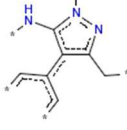
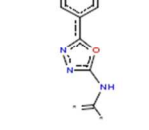
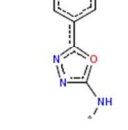
*To whom correspondence should be addressed. E-mail: ekinssean@yahoo.com

Running title: Fusing datasets for *Mtb*

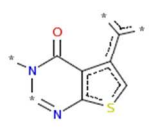
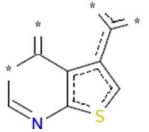
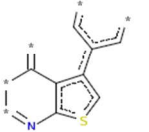
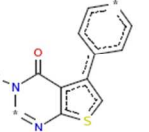
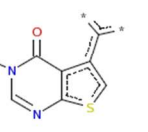
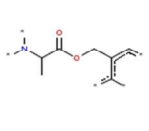
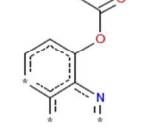
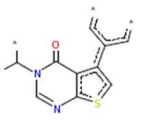
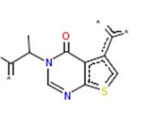
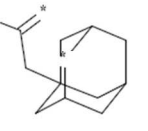
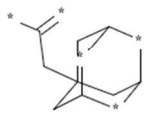
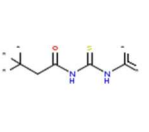
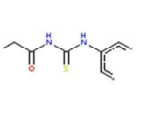
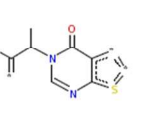
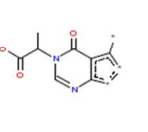
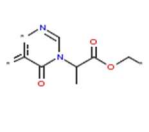
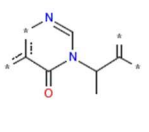
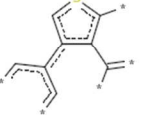
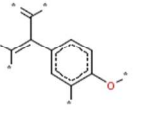
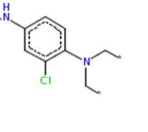
Supplemental Figure 1. ARRA dose response and cytotoxicity Bayesian model: good features from FCFP_6.

 <p>G1: 1673799978 12 out of 21 good Bayesian Score: 1.945</p>	 <p>G2: -1449793464 12 out of 23 good Bayesian Score: 1.902</p>	 <p>G3: -60038150 47 out of 151 good Bayesian Score: 1.900</p>	 <p>G4: 1209850658 27 out of 79 good Bayesian Score: 1.890</p>	 <p>G5: -278994470 26 out of 85 good Bayesian Score: 1.797</p>
 <p>G6: 363977085 48 out of 175 good Bayesian Score: 1.793</p>	 <p>G7: -312381965 35 out of 128 good Bayesian Score: 1.753</p>	 <p>G8: -1883081175 8 out of 14 good Bayesian Score: 1.744</p>	 <p>G9: -1387481105 14 out of 45 good Bayesian Score: 1.664</p>	 <p>G10: 1670875748 16 out of 56 good Bayesian Score: 1.642</p>
 <p>G11: 854938188 10 out of 29 good Bayesian Score: 1.616</p>	 <p>G12: -428284881 12 out of 39 good Bayesian Score: 1.611</p>	 <p>G13: 1669687439 13 out of 44 good Bayesian Score: 1.609</p>	 <p>G14: -1048017265 14 out of 49 good Bayesian Score: 1.608</p>	 <p>G15: 1941608633 14 out of 49 good Bayesian Score: 1.608</p>
 <p>G16: 1476473713 14 out of 49 good Bayesian Score: 1.608</p>	 <p>G17: -1700730929 14 out of 49 good Bayesian Score: 1.608</p>	 <p>G18: -180488964 12 out of 40 good Bayesian Score: 1.596</p>	 <p>G19: -891643182 13 out of 45 good Bayesian Score: 1.595</p>	 <p>G20: 1953499806 30 out of 137 good Bayesian Score: 1.546</p>

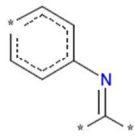
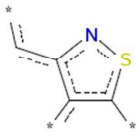
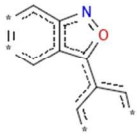
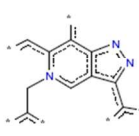
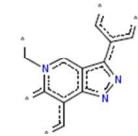
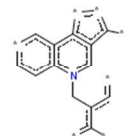
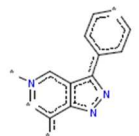
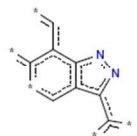
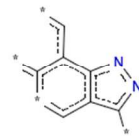
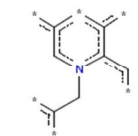
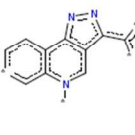
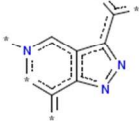
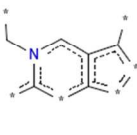
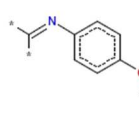
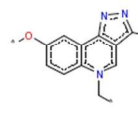
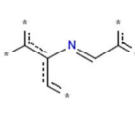
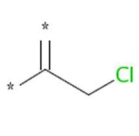
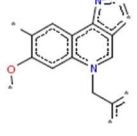
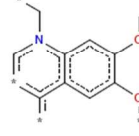
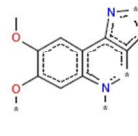
Supplemental Figure 2. ARRA dose response and cytotoxicity Bayesian model: bad features from FCFP_6.

 <p>B1: -1280036918 0 out of 154 good Bayesian Score: -1.988</p>	 <p>B2: 368675221 0 out of 147 good Bayesian Score: -1.948</p>	 <p>B3: -743321149 0 out of 128 good Bayesian Score: -1.830</p>	 <p>B4: -1076532720 0 out of 126 good Bayesian Score: -1.817</p>	 <p>B5: 1704324327 1 out of 258 good Bayesian Score: -1.754</p>
 <p>B6: 1160440068 1 out of 258 good Bayesian Score: -1.754</p>	 <p>B7: -1775135141 1 out of 258 good Bayesian Score: -1.754</p>	 <p>B8: 193617513 0 out of 114 good Bayesian Score: -1.734</p>	 <p>B9: -581879738 1 out of 252 good Bayesian Score: -1.732</p>	 <p>B10: -1388275145 1 out of 242 good Bayesian Score: -1.696</p>
 <p>B11: 1790572653 0 out of 107 good Bayesian Score: -1.682</p>	 <p>B12: 2015833643 0 out of 106 good Bayesian Score: -1.674</p>	 <p>B13: 1919278075 0 out of 106 good Bayesian Score: -1.674</p>	 <p>B14: -1915183124 0 out of 105 good Bayesian Score: -1.667</p>	 <p>B15: 436975416 0 out of 101 good Bayesian Score: -1.635</p>
 <p>B16: 2109546364 0 out of 98 good Bayesian Score: -1.611</p>	 <p>B17: -1956934524 0 out of 96 good Bayesian Score: -1.595</p>	 <p>B18: 1593831848 0 out of 96 good Bayesian Score: -1.595</p>	 <p>B19: 2006426266 1 out of 213 good Bayesian Score: -1.580</p>	 <p>B20: -1266529233 1 out of 213 good Bayesian Score: -1.580</p>

Supplemental Figure 3. Combined dose response and cytotoxicity Bayesian model: good features from FCFP_6.

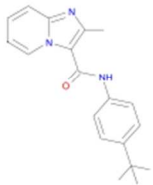
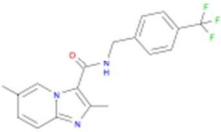
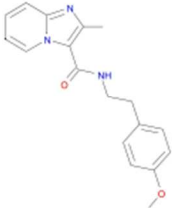
 <p>G1: 713865785 10 out of 11 good Bayesian Score: 1.084</p>	 <p>G2: -862029858 10 out of 11 good Bayesian Score: 1.084</p>	 <p>G3: -1974830073 10 out of 11 good Bayesian Score: 1.084</p>	 <p>G4: 92779600 10 out of 11 good Bayesian Score: 1.084</p>	 <p>G5: 921426640 7 out of 7 good Bayesian Score: 1.075</p>
 <p>G6: 529664119 7 out of 7 good Bayesian Score: 1.075</p>	 <p>G7: -634910424 7 out of 7 good Bayesian Score: 1.075</p>	 <p>G8: 628932230 7 out of 7 good Bayesian Score: 1.075</p>	 <p>G9: -92657515 7 out of 7 good Bayesian Score: 1.075</p>	 <p>G10: 504464533 16 out of 20 good Bayesian Score: 1.051</p>
 <p>G11: -583648804 16 out of 20 good Bayesian Score: 1.051</p>	 <p>G12: 1023493548 13 out of 16 good Bayesian Score: 1.039</p>	 <p>G13: 1925706152 15 out of 19 good Bayesian Score: 1.033</p>	 <p>G14: 883653880 8 out of 9 good Bayesian Score: 1.026</p>	 <p>G15: -1355870860 8 out of 9 good Bayesian Score: 1.026</p>
 <p>G16: -861283934 8 out of 9 good Bayesian Score: 1.026</p>	 <p>G17: 1426666980 8 out of 9 good Bayesian Score: 1.026</p>	 <p>G18: -648390934 10 out of 12 good Bayesian Score: 1.020</p>	 <p>G19: 1773007393 12 out of 15 good Bayesian Score: 1.016</p>	 <p>G20: 900585349 17 out of 23 good Bayesian Score: 0.990</p>

Supplemental Figure 4. Combined dose response and cytotoxicity Bayesian model: bad features from FCFP_6.

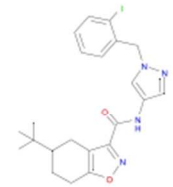
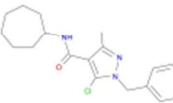
 <p>B1: -1699460258 0 out of 59 good Bayesian Score: -2.746</p>	 <p>B2: 1087307120 0 out of 43 good Bayesian Score: -2.454</p>	 <p>B3: -437082084 0 out of 37 good Bayesian Score: -2.317</p>	 <p>B4: 1813774889 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B5: -1528325941 0 out of 36 good Bayesian Score: -2.292</p>
 <p>B6: -2024954534 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B7: -1236091207 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B8: -438280287 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B9: 267940428 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B10: -1731186150 0 out of 36 good Bayesian Score: -2.292</p>
 <p>B11: 2056975573 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B12: -1227636663 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B13: 2101570543 0 out of 36 good Bayesian Score: -2.292</p>	 <p>B14: -1185638464 0 out of 33 good Bayesian Score: -2.215</p>	 <p>B15: 100031611 0 out of 30 good Bayesian Score: -2.130</p>
 <p>B16: -1058561716 0 out of 29 good Bayesian Score: -2.100</p>	 <p>B17: -1208154531 0 out of 28 good Bayesian Score: -2.070</p>	 <p>B18: 747352596 0 out of 26 good Bayesian Score: -2.005</p>	 <p>B19: 91575904 0 out of 26 good Bayesian Score: -2.005</p>	 <p>B20: 374776331 0 out of 26 good Bayesian Score: -2.005</p>

Supplemental Figure 5. Examples of GSK compounds (see Molecule Name) clustering with TB Mobile compounds known to target specific *Mtb* targets (see Target Gene 1).

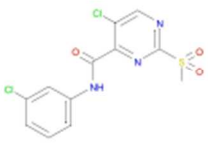
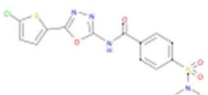
A

Structure	Molecule Name	Cluster(1)	TargetGene1
	GSK1829660A	21	
	GSK762874A	21	qcrB (Rv2196)
	SB-204804-A	23	panC (Rv3602c)

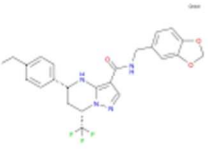
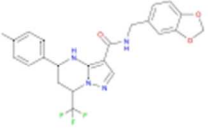
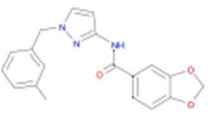
B

	GSK636544A	23	panC (Rv3602c)
	SB-204804-A	23	
	GSK636544A	23	

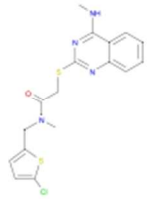
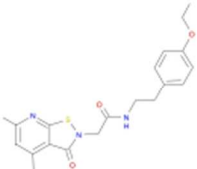
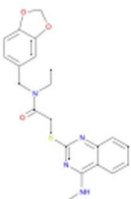
C

		36	alr (Rv3423c)
	GSK991960A	36	
		36	ivG (Rv1820)

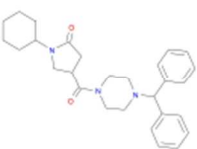
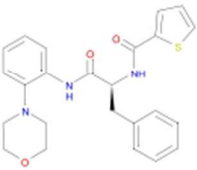
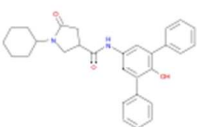
D

		39	mmpL3 (Rv0206c)
	GSK937733A	39	
	GSK798463A	39	

E

	GSK1925843A	39	
		39	alr (Rv3423c)
	GSK1859936A	39	

F

		49	inhA (Rv1484)
	GSK1729177A	49	
		49	inhA (Rv1484)