Electronic Supplementary Material (ESI) for Molecular BioSystems

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1 Exploring the mechanism of the resolution of K-NN predictions. **A)** 100 training datasets with 4 drug mechanism subnetworks and 3 training examples were simulated alongside 1000 test examples per drug mechanism subnetwork. The performance of diverse classifiers for these 100 training sets was compared. **B)** 100 training datasets with 4 drug mechanism subnetworks and 10 training examples were simulated alongside 1000 test examples per drug mechanism subnetwork. The performance of diverse classifiers for these 100 training sets was compared.

Supplementary Figure 2 Examining the accuracy of K-NN predictions across experimental error. **A)** Schematic of the simulations: Doxorubicin replicates (N=30) over three years are plotted in the heatmap. These experimental replicates are used to plot an eCDF, which is then used to inverse transform sample 1000 simulated signatures. This is done for 4 drugs; Doxorubicin (Dox), Chlorambucil (CBL)(N=9), Vincristine (Vin)(N=7), and 5-Flourouracil (5-FU)(N=31). **B)** The accuracy of the K-NN algorithm upon the simulated signatures.

Supplementary Figure 3 Examining chemical intervention datasets. **A)** After model training using topoisomerase II and I poisons, spindle poisons, alkylating agents, proteasome inhibitors and HDAC inhibitors, these other drugs in the Wolpaw et al. dataset were successfully recognized as “New” drugs. **B)** Clustering of the “New” drugs from A. Colors represent different drug categories. Brown lines indicate potential clustering solutions. **C)** Different edge thresholded topologies. The colors correspond to B. Boxed is the final solution following network consensus clustering and supervised refinement. **D)** A heatmap and surface plot of network consensus clustering scores for all edge thresholds and 2-5 clusters. A tie is indicated at three different cluster-edge interpretations. **E)** Statistical and biological generalization in the Wolpaw et al. dataset using our supervised methodology.

Supplementary Figure 4 Examples of connectivity map searches. **A)** The “best” topoisomerase II poison search in MCF 7 arrays. Large font indicates a drug that is also in the table in Figure 4. False positives are indicated in red. True positives are indicated in black. **B)** The “best” topoisomerase II poison search in MCF 7 arrays. Large font indicates a drug that is also in the table in Figure 4. False positives are indicated in red. True positives are indicated in black.

Supplementary Figure 5 Connectivity map arrays identified as “New” drug mechanisms. The cell line, array id, drug name, and prediction/statistics for a variety of drugs from the connectivity map. Colors indicate drugs with the same mechanisms of action, or array replicates.