Science Signaling
AAAS

peopwww.sciencesignaling.org/cgi/content/full/3/118/ra30/DC1

Supplementary Materials for

Systems Pharmacology of Arrhythmias

Seth I. Berger, Avi Ma'ayan, Ravi Iyengar*

*To whom correspondence should be addressed. E-mail: Ravi.Iyengar@mssm.edu

Published 20 April 2010, *Sci. Signal.* **3**, ra30 (2010)
DOI: 10.1126/scisignal.2000723

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencesignaling.org/cgi/content/full/3/118/ra30/DC1)

Figure S1

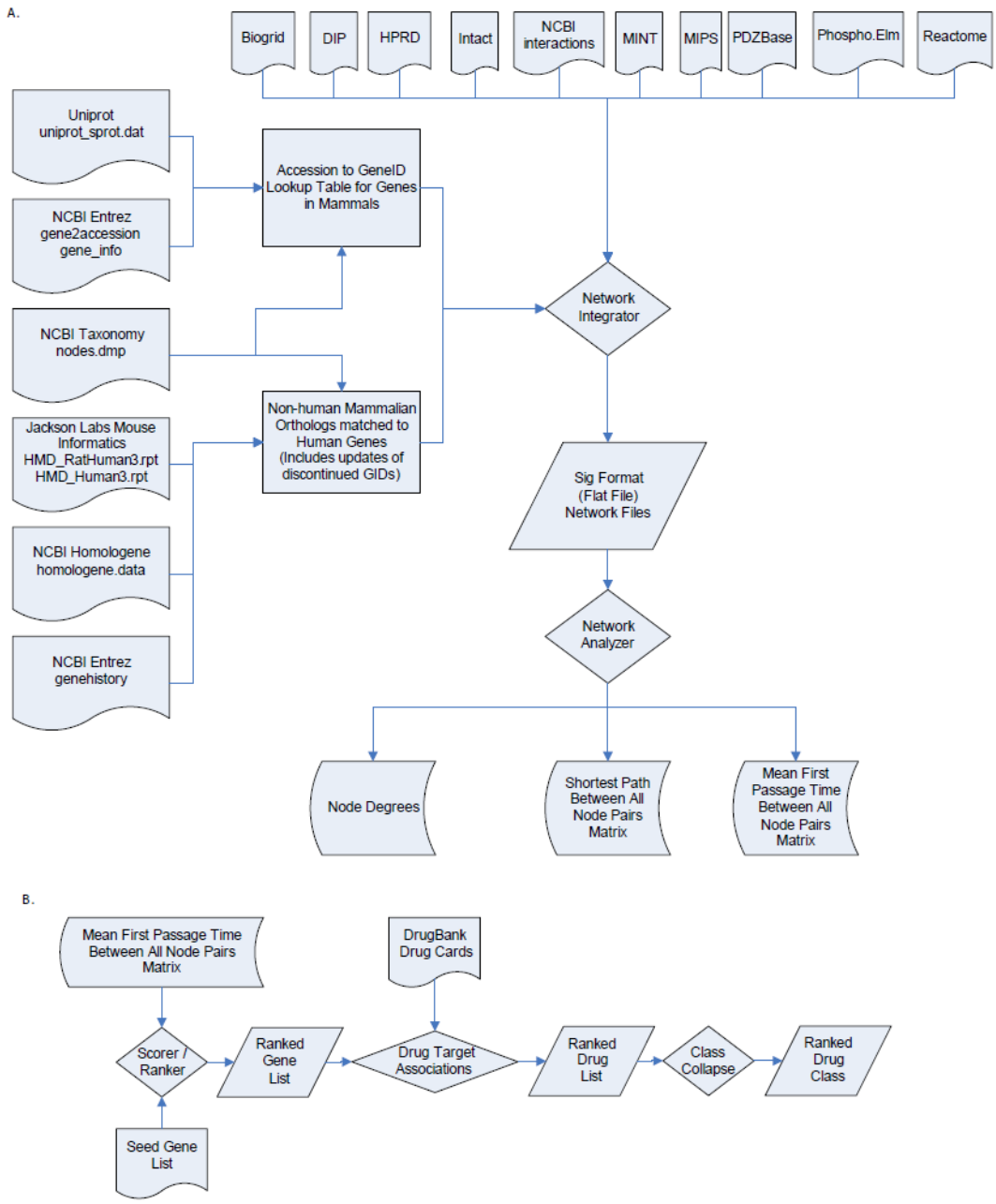

Fig. S1. Overview of data integration and analysis. (A) A flowchart shows the method for network construction. Information from NCBI and Uniprot was used to construct a lookup table from various Entrez gene accession symbols for all mammalian species. Data from NCBI Homologene, Jackson Labs Mouse Informatics, and NCBI Gene History was used to construct a lookup table that matched each mammalian gene to its human ortholog. Curated data from Jackson labs took precedence over predicted orthology from Homologene. These lookup tables were used to process the protein interaction data taken from Biogrid, DIP, HPRD, Intact, NCBI, MINT, MIPS, PDZBase, PhosphoElm, and Reactome. Each protein from a mammalian species was matched to its gene and subsequently to its human ortholog. Using a breadth-first search, all interactions that were not reachable from KCNH2 were discarded from the network to leave a large fully connected cluster. The network was then analyzed to give a table of node degrees, a matrix of shortest paths between all node pairs, and a matrix of mean first passage times (MFPTs) between all node pairs. (B) For a given seed gene list, the MFPT matrix was used to calculate a score for all nodes in the network. Sorting based on the scores provided a ranked list of genes. With drug target data from DrugBank, all the drugs in DrugBank were assigned the score of their highest ranked target, providing a ranked drug list. Drug classes are then collapsed by grouping all drugs with the same highest ranked target (ignoring KCNH2 if it was the highest target) into a single class.

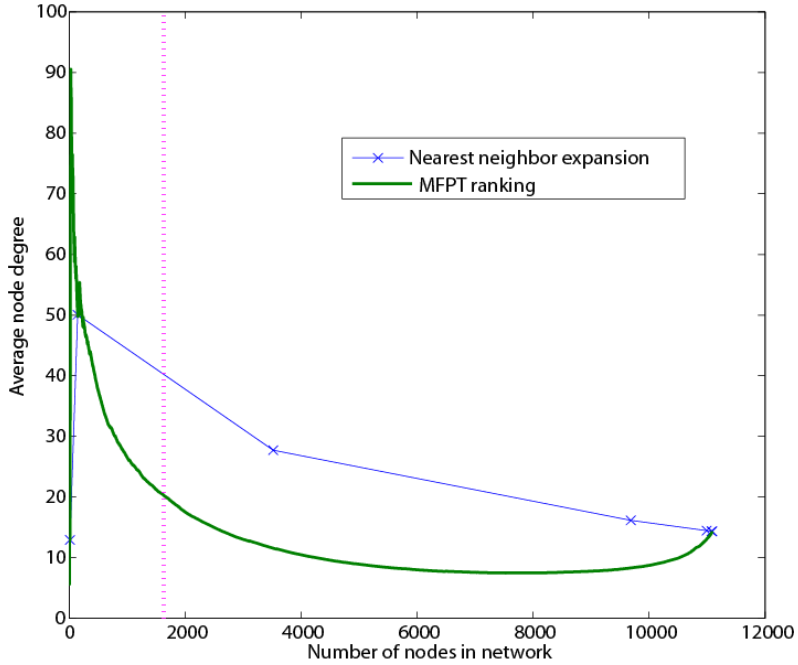

Fig. S2. Comparison of node connectivity in networks created by nearest-neighbor expansion or through module distance score based on MFPT. The average node degree of nodes in subnetworks generated by nearest neighbor expansion (blue) were generally greater than for nodes in networks based on ranking by the MFPT score (green). The pink vertical line labels the cutoff for the LQTS neighborhood. Networks including only the seed genes, the highest ranked nodes, have the same average degree. This demonstrates that the LQTS neighborhood deemphasized hubs (highly connected nodes) in the network.

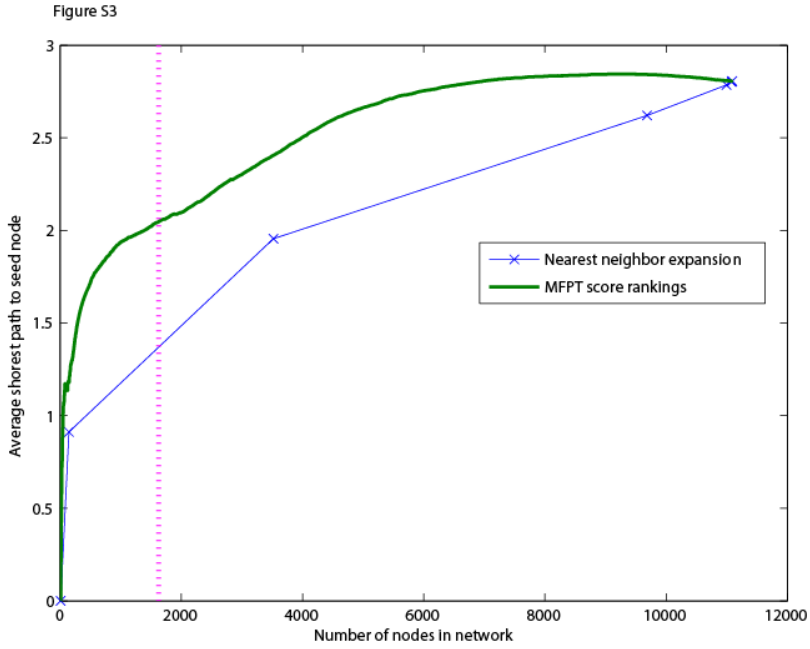

Fig. S3. Average shortest path to seed node as a function of network size. The average shortest path to a LQTS seed node in subnetworks generated by nearest neighbor expansion (blue) were shorter than for networks based on MFPT ranking (green). The pink vertical line labels the cutoff for the LQTS neighborhood. Nodes in the network based on MFPT ranking tend have a longer shortest path to the seed list nodes because nodes connected through short paths through hubs are ranked lower than nodes connected through longer hubless paths. The LQTS neighborhood contains half as many nodes as a network containing all the first and second neighbors of the LQTS seed genes. However, the nodes in the LQTS neighborhood are on average 2 steps away from the nearest LQTS seed gene.

Fig S4

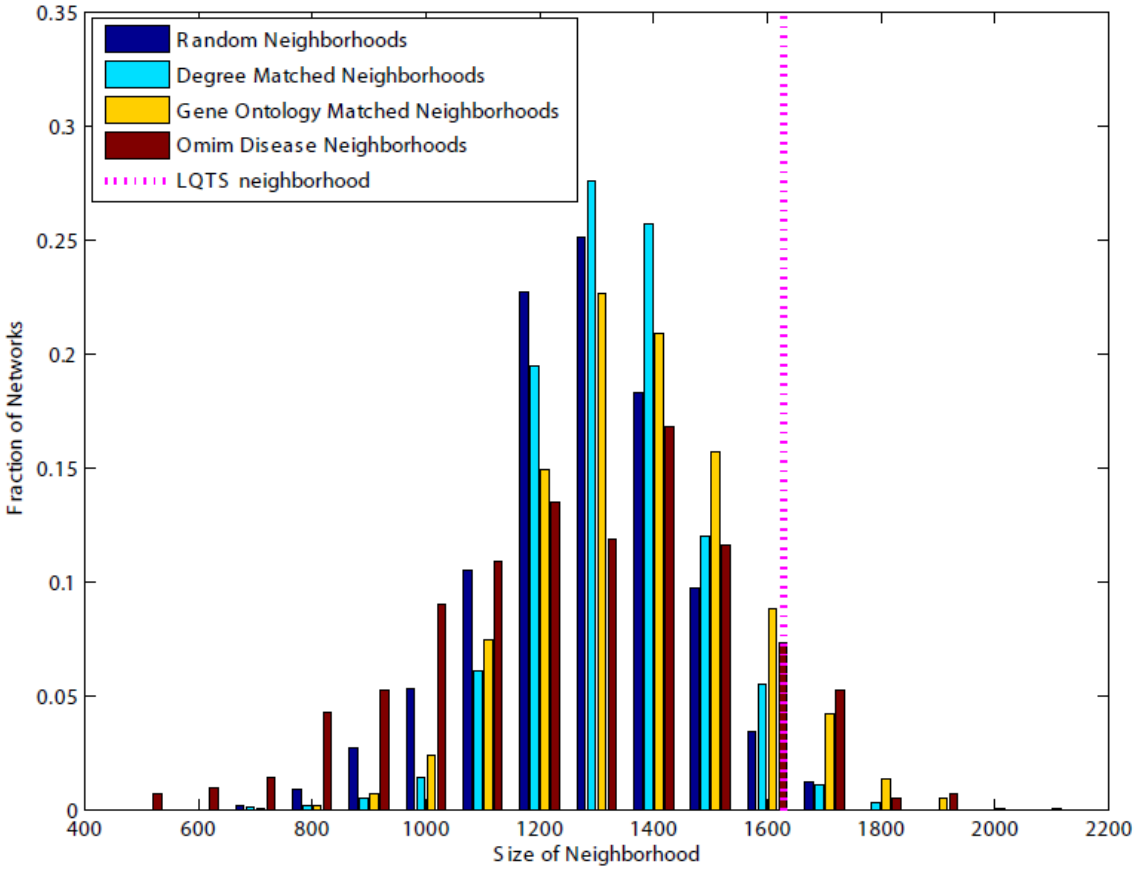

Fig. S4. Network size distribution. Neighborhood size depends on the list of seed genes used to construct the neighborhood. Random neighborhoods subnetworks and degree-matched neighborhood subnetworks were generated from 1,000 lists of 13 seeds. The Gene Ontology (GO)-matched neighborhood subnetworks were generated from 2,500 lists of gens with the same distribution of GO Cellular Compeont terms. The neighborhood sizes for difference diseases in OMIM has a larger range than the random neighborhoods due to the variation in seed list size for the disease networks. The LQTS neighborhood falls at the larger end of the neighborhood size distribution but is not at an extreme.

Fig. S5

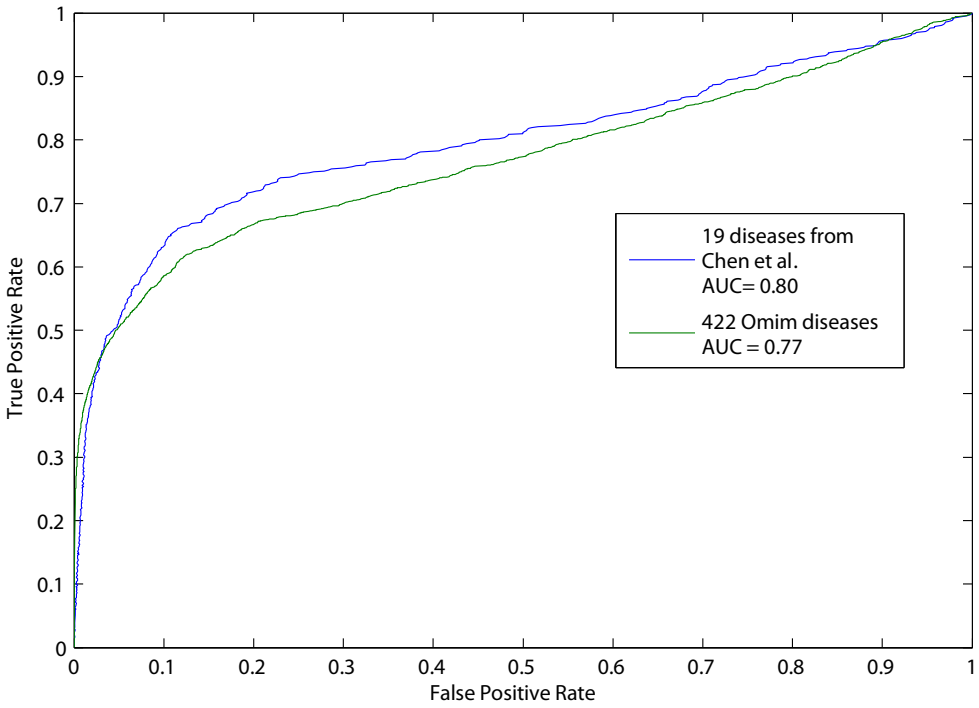

Fig. S5. Receiver operating characteristic (ROC) curve for classification of disease genes. ROC curves were generated for candidate gene prediction by leave-one-out analysis on the set of OMIM disease gene sets and the disease gene sets used by Chen *et al*. For candidate gene prediction, the MFPT score ranking approach performs as well as the method published by Chen *et al*. on their disease gene sets without the need to tune the algorithm with an arbitrary parameter (blue line). On the larger number of disease sets with fewer genes per disease, the MFPT score ranking approach also performs well at candidate disease gene prediction (green line). It performs slightly worse on this set as the smaller seed lists provide less information about each disease.

Chen, J., Aronow, B.J., Jegga, A.G. (2009) Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* **10**: 73.

Fig. S6.

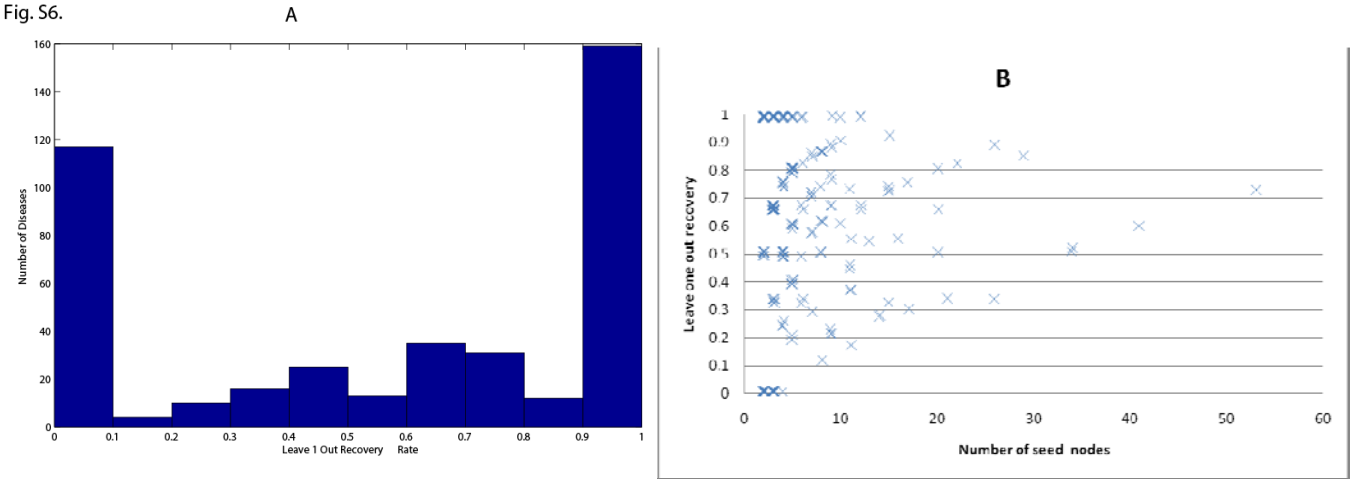

Fig. S6. Network modular versus nonmodular diseases: Leave-one-out recovery rate distribution. A) For each disease in OMIM, we calculated the fraction of the seed genes that achieved a positive score when left off the seed list. In most cases, a majority of the genes associated with a disease fell within the neighborhood generated from the remaining disease genes. B) The large peaks in A at 100% and 0% are due to diseases with very small seed lists of less than 4 genes as seen on this scatterplot.

Figure S7

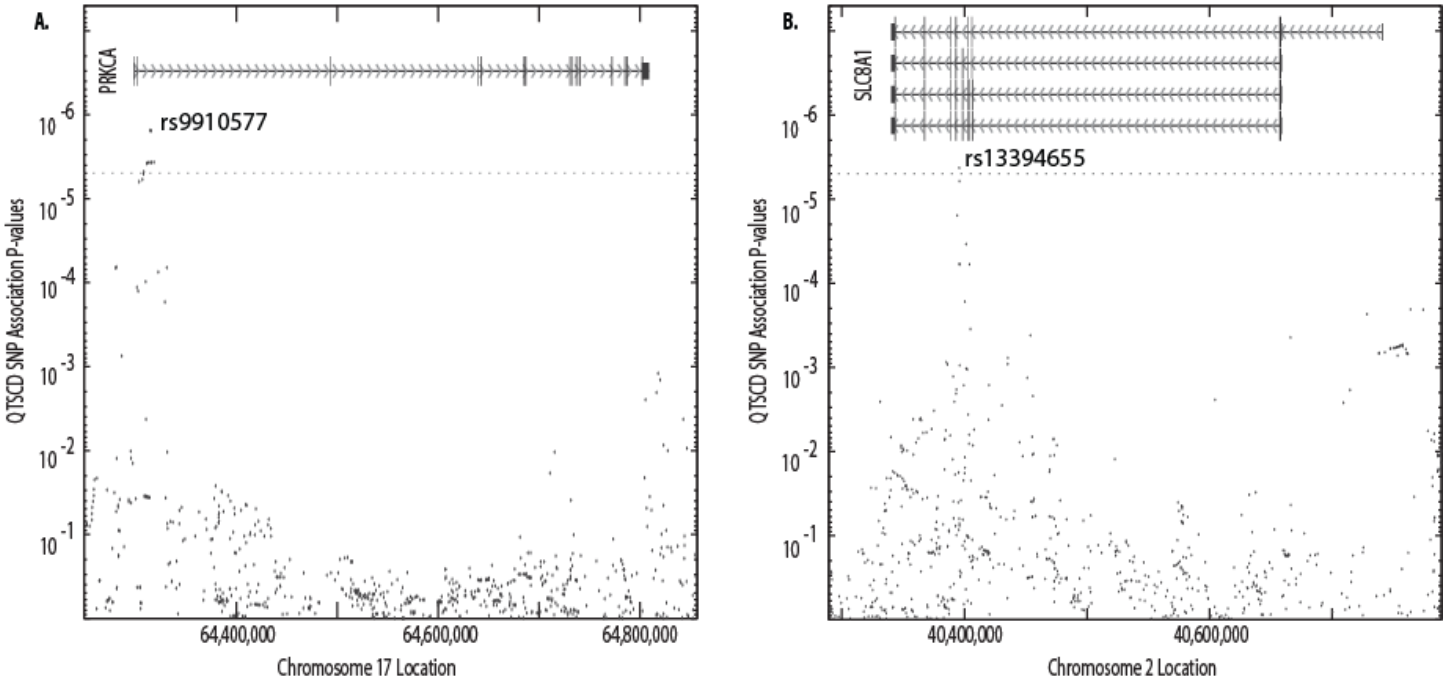

Fig. S7. The use of the LQTS neighborhood rankings to identify SNPs in the QTSCD GWAS. Manhattan plot demonstrates SNPs in the genomic regions around (A) *PRKCA*, encoding protein kinase C α and (B) *SLC8A1*, encoding the cardiac sodium calcium exchanger. These genes contain SNPs, rs9910577 in *PRKCA* and rs13394655 in *SLC8A1*, which have p-values less than 5×10^{-6} but did not attain genome-wide significance in the original QTSCD study. They may be considered significant in the context of the LQTS neighborhood because these genes also rank highly (1st and 94th, respectively) in the LQTS neighborhood.

Fig. S8

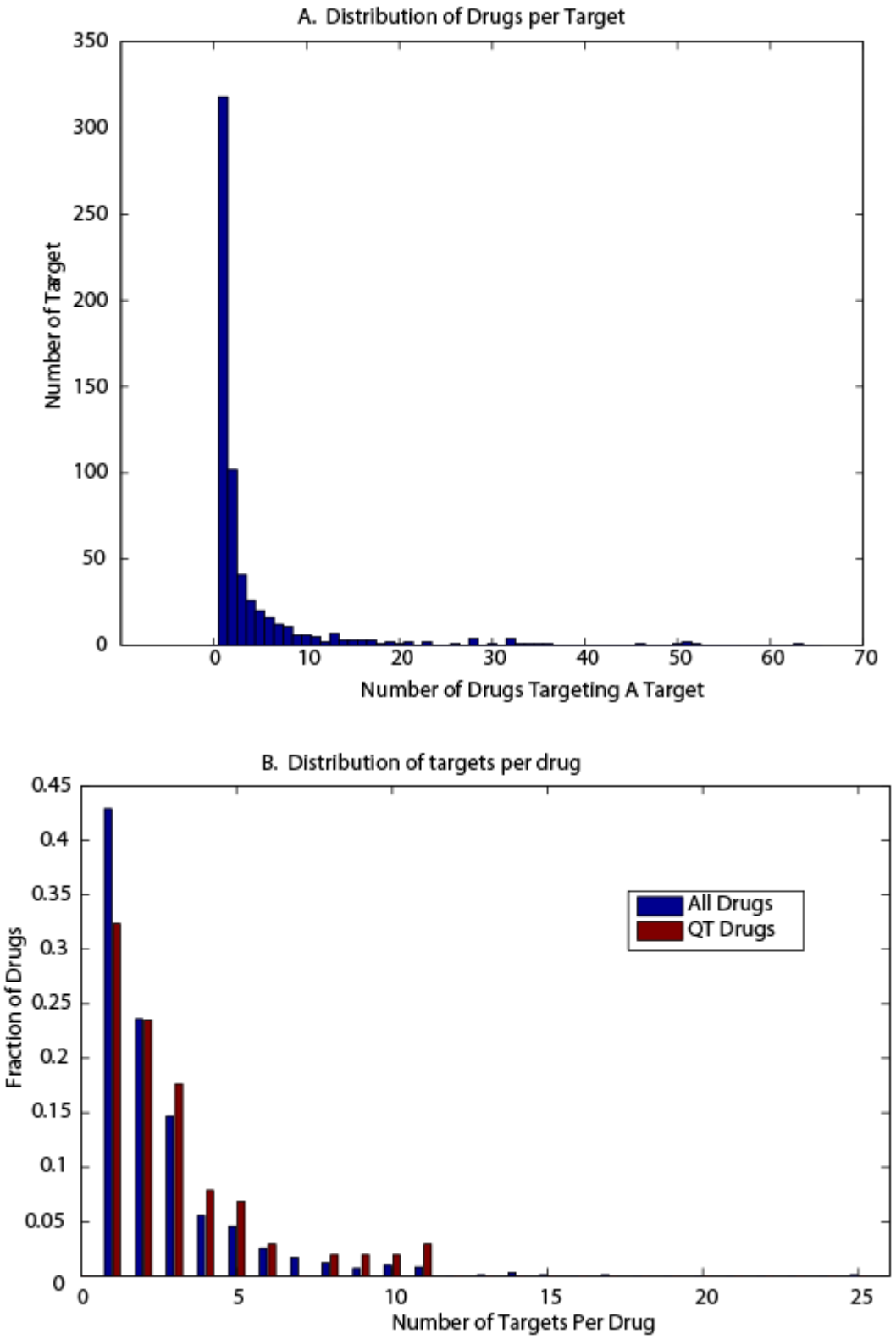

Fig. S8. The distribution of drug targets. (A) Most of the drug targets in DrugBank are targeted by only one or two drugs with relatively few the targets of 10 or more drugs. (**B**) Most of the drugs in DrugBank have only one or two targets and relatively few have greater than five targets. The QT-prolonging drugs have a comparatively larger fraction with greater than five targets.

Fig. S9

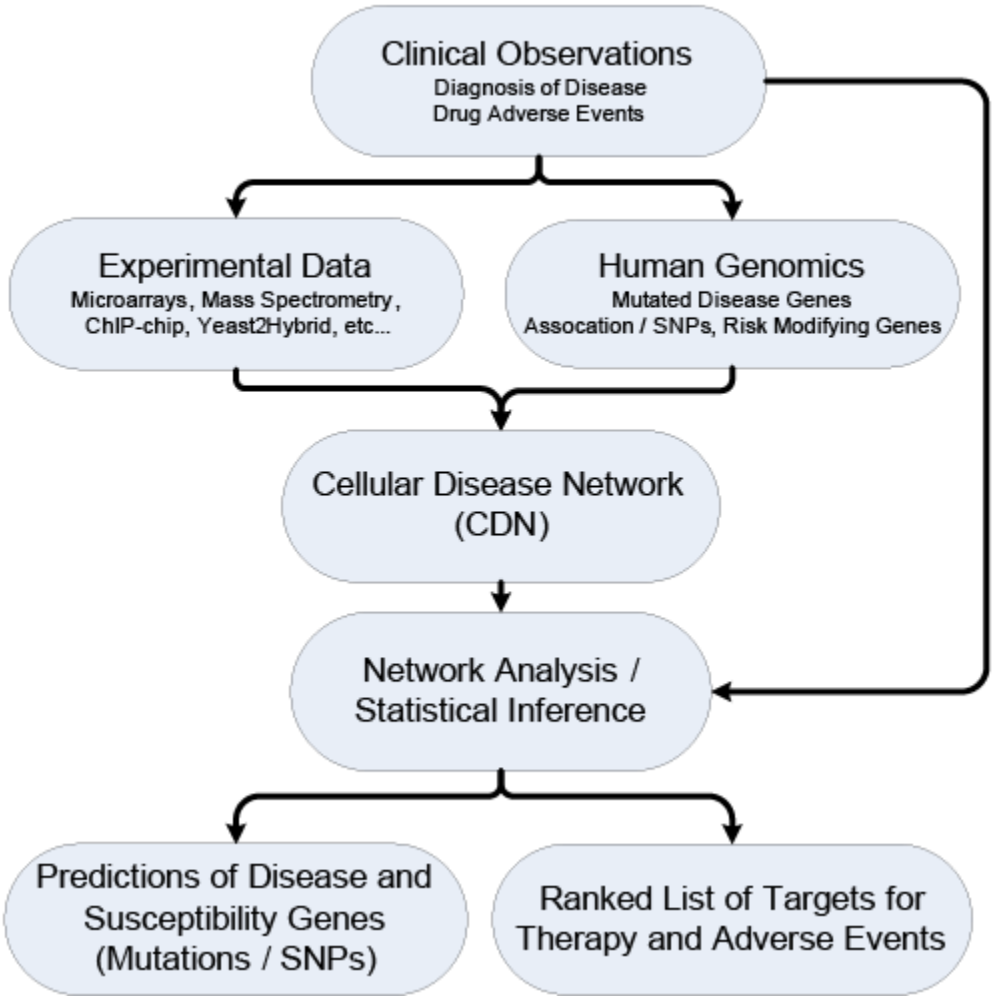

Fig. S9. Methodology for integrating systems pharmacology and personalized medicine. Systems pharmacology studies integrate clinical and experimental datasets to provide new clinical and basic science insights. Clinical observations can lead to discovery of disease-associated genes in human genomic studies. Integrating this with experimental data, such as protein interaction information, enables the creation of a cellular disease network model. Use of this model can assist in analysis of additional clinical information, such as drug adverse effect surveillance datasets, to produce predictions of new disease susceptibility genes and drug targets related to disease or adverse events.

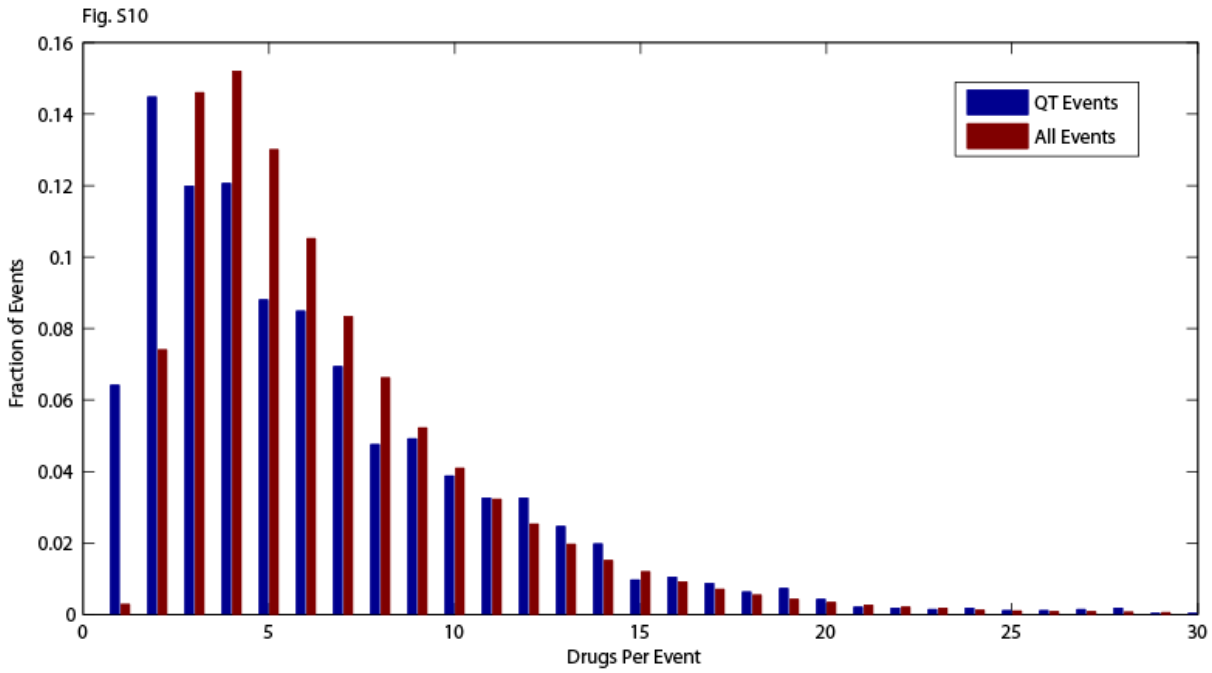

Fig. S10. The distribution of number of drugs associated with each adverse event report in AERS. A greater proportion of QT events are reported with only one or two drugs.

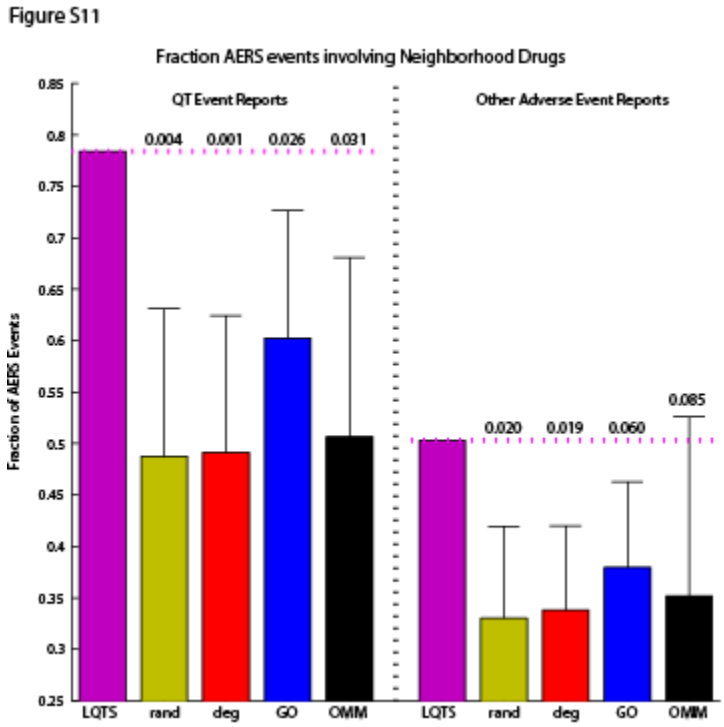

Fig. S11. Fraction of AERS events involving neighborhood drugs. The predicted drugs may explain close to 80% of the previously unexplainable QT events in AERS, which is significantly more than the control neighborhoods and much greater than the fraction of other adverse events that could potentially be attributed to these drugs.