Online Methods

Ligand sets. We extracted ligand sets from databases that annotate molecules by therapeutic or biological category. E.g., the 2006.1 MDL Drug Data Report (MDDR) contains 518 molecules annotated as α₁ adrenergic receptor blockers, which we grouped into a single “α₁ adrenergic blocker” set.

As ligand reference sources, we used the three databases shown in Supplementary Table 8. The first was a subset of the 2006.1 MDDR, prepared as previously described. The second was the 2006.2 World of Molecular Bioactivity (WOMBAT) database, processed as above. We collapsed WOMBAT targets across species and organized them into inhibitory, activating, and unspecified-binding classes. All ligands with affinities worse than 1 μM to their targets were removed. This left 1,133 classes built from 191,943 ligands with median and mean of 37 and 169 ligands per target class. The third database was StARlite, which we also processed as above. We extracted StARlite annotations at the two highest confidence levels (5 and 7), discarded those with affinities worse than 1 μM, and organized them into target classes. This yielded 1,158 classes built from 111,329 ligands with median and mean of 43 and 186 ligands per target class.

For drugs and bioactive molecules, we used the 2004 MDL Comprehensive Medicinal Chemistry database (CMC) of 7,517 compounds. Drugs were processed identically to the ligands above. Filtering by vendor availability (as reported in MDL 2006.3 Available Chemical Directory (ACD), MDL 2006.1 Screening Compounds Directory, and ZINC) yielded 3,665 unique purchasable drugs.

The 1,216 drugs used to link protein targets in the drug-target networks (Figure 1) were downloaded from the 2008 EPA Distributed Structure-Searchable Toxicity (DSSTox) Database at http://www.epa.gov/NCCT/dsstox/, and prepared as above.
**Ligand activity predictions.** We compared each drug individually against each set of ligands. Molecules were represented by two topological descriptors: 2048-bit Daylight and 1024-bit folded ECFP_4 fingerprints. We used the Similarity Ensemble Approach (SEA) using each descriptor separately and chose top-scoring hits (i.e., those with small E-values) from each such “screen” independently.

The initial SEA screens of 3,665 CMC drugs against 246 MDDR targets yielded 901,590 drug-target comparisons, and we subjected these to retrospective literature analysis and prospective empirical testing. However, we later extended SEA screens to WOMBAT and StARlite databases, comprising 4,152,445 and 4,244,070 drug-target comparisons, respectively. We have not mined these expanded SEA screens for retrospective validations; instead we used them only to conduct prospective tests. **Supplementary Table 8** records the screen (i.e., database) from which each prediction in Table 1 and Table 2 is derived.

To compare SEA predictions against those of naïve Bayesian classifiers (**Supplementary Table 1**), we implemented a Laplacian-corrected naïve Bayesian classifier with Avidon weighting, as previously described.

**Drug-target and target-target networks.** The drug-target networks in Figure 1 are bipartite; along any given path, nodes alternate between protein targets and the drugs that link them. Targets are from WOMBAT and drugs are from EPA DSSTox. Red edges denote SEA predictions with E-values ≤ 10^{-10}. Predictions already reported in WOMBAT at K_i ≤ 1 μM are shown as additional gray edges. All networks were generated in Cytoscape 2.6.1.

**Figure 3** is a bipartite graph linking drugs from Table 1 and Table 2 with protein targets. Gray edges link drugs to known targets from manual literature and database search. Gray edges denote binding at ≤ 1 μM, except when no K_i value was available; in
these 3 cases (Xenazine-VMAT2, Prantal-M₃, and Fabahistin-H₁), the link was included for completeness.

**WOMBAT out-group analysis.** We mapped 204 MDDR activity classes to matching WOMBAT targets in two phases. In the first, we mapped 87 MDDR activity classes using EC numbers from the Schuffenhauer ontology to those present in WOMBAT. We second mapped a further 118 non-enzyme MDDR activity classes by supervised sub-phrase matching (**Supplementary Table 9, Supplementary Table 10**). While this mapping is not guaranteed to be exhaustive, it is correct to the best of our knowledge.

We then extracted all molecules marked “drug” in WOMBAT (746 unique). Using SEA, we compared them against the mapped MDDR classes only, and discarded all trivial hits (i.e., those where the drug was already annotated in that MDDR class as a ligand). We asked how many of these were, in retrospect, substantiated by the existing WOMBAT annotations, at affinities ≤ 1 μM.

**Sequence similarity comparison.** We associated each drug in **Figure 3** with the human FASTA sequences of its known and its new protein targets, using http://www.uniprot.org. We ran these sequences via PSI-BLAST (BLAST version 2.2.14, default parameters) against a subset of MDDR targets that we prepared as previously described. For each SEA prediction in **Figure 3**, we reported the best direct PSI-BLAST match (along with its E-value and ranking) of the new protein target to any of that drug’s previously known protein targets (**Supplementary Table 6**). Our goal was to address the question, “Starting with the best choice from a drug’s known protein targets, can we recapitulate each SEA prediction solely by sequence similarity?”

**Experimental testing.** Radioligand binding and functional assays were performed as previously described. Detailed experimental protocols are available on the NIMH PDSP website at http://pdsp.med.unc.edu/UNC-CH%20Protocol%20Book.pdf.
**Mice.** All experiments were approved by the Institutional Animal Care and Use Committee at the University of North Carolina, Chapel Hill. Mice were housed under standard conditions – 12 hour light/dark cycle and food and water ad libitum.

**Head Twitch.** Littermate pairs of 5-HT\textsubscript{2A} wildtype and knockout mice were pretreated for two hours with 75 mg/kg pargyline, i.p., prepared in sterile saline (.9% NaCl) (P8013, Sigma-Aldrich, St. Louis, MO). Mice were then injected with sterile saline or 1.0 mg/kg DMT, i.p., prepared in sterile saline and moved to a new cage. Head twitch behavior, which consists of a rapid, rotational flick of the head about the axis of the neck, was counted over 15 minutes. We have determined that trained observers count the same number of head twitches whether blinded or unblinded to genotype (data not shown). We confirmed that this was the case with three littermate pairs, and the rest of the studies were performed by one unblinded observer.\textsuperscript{34}

**Online Methods References**

51 James, C., Weininger, D. & Delany, J. *Daylight theory manual*. (Daylight Chemical Information Systems Inc, 1992-2005).

52 Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

53 Roth, B. L. et al. Salvinorin A: a potent naturally occurring nonnitrogenous kappa opioid selective agonist. *Proc Natl Acad Sci USA* **99**, 11934-11939 (2002).