Integrative pharmacogenomics to infer large-scale drug taxonomy

Nehme El-Hachem[1,2,‡], Deena M.A. Gendoo[3,4,‡], Laleh Soltan Ghoraie[3,4], Zhaleh Safikhani[3,4], Petr Smirnov[3], Ruth Isserlin[5], Gary D. Bader[5,6,7], Anna Goldenberg[7,8], Benjamin Haibe-Kains[3,478,$]

[1] Integrative Computational Systems Biology, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada
[2] Department of Biomedical Sciences. Université de Montréal, Montreal, Quebec, Canada
[3] Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada
[4] Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada
[5] The Donnelly Centre, Toronto, Ontario, Canada
[6]The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada
[7] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada
[8] Hospital for Sick Children, Toronto, Ontario, Canada

[‡] Co-first authors
[$] Corresponding author

ABSTRACT

Identification of drug targets and mechanism of action (MoA), particularly for new and uncharacterized drugs, is important for the optimization of drug efficacy. Current approaches towards determining drug MoA largely rely on prior information such as side effects, therapeutic indication and chemo-informatics. However, such information is not transferable or applicable for newly identified small molecules. Despite continuous release of large-scale pharmacogenomic datasets, these valuable data remain underused to classify drugs. Accordingly, a systematic and unbiased approach towards MoA prediction is imperative to efficiently classify new compounds and infer their potential targets of MoA. Here, we propose a method that only relies on basic drug characteristics, including drug structural information, drug perturbation and drug sensitivity profiles, which have not been previously combined towards predicting drug targets and MoA. We harnessed the full potential of pharmacogenomics data using our Similarity Network Fusion approach to implement Drug Network Fusion (DNF), a scalable, integrative drug taxonomy. We demonstrate that DNF is effective towards prediction of drug targets and anatomical therapeutic chemical (ATC) classification). Our method enables robust inference of drug MoAs for new and existing compounds, using integrative computational pharmacogenomics.

INTRODUCTION

Continuous growth and ongoing deployment of large-scale pharmacogenomic datasets has opened new avenues of research for the prediction of biochemical interactions of small drug molecules with their respective targets, also referred to as drug mechanisms of action (MoA). Several computational strategies have relied on chemical structure similarity to infer drug-target interactions [1–3], based on the assumption that structurally-similar drugs share similar targets, and ultimately, similar pharmacological and biological activity [4]. However, sole reliance on chemical structure information fails to consider drug-induced genomic and phenotypic perturbations, which directly connect with biological pathways and molecular disease mechanisms [5,6]. Recent approaches have thereby integrated drug-induced transcriptional profiles from Connectivity Map (CMAP) [7] into their algorithms, creating new ways for identification of drug-drug similarities and MoA solely based on gene expression profiles [8]. Other methods have integrated prior knowledge such as adverse effects annotations [9,10] and recent approaches showed that integrating multiple layers of information had improved ATC prediction for FDA-approved drugs [11]. While these initiatives have undoubtedly paved great strides towards characterizing drug MoA, determining the consistency of such efforts towards prediction of new, uncharacterized small molecules remains a challenge.

The advent of high-throughput molecular profiling to identify patterns of small-molecule sensitivities across cell lines promises to shed additional insight into drug MoA. This type of drug bioactivity information remains largely unexploited in drug classification algorithms, despite its ongoing development over the past decade. The pioneering initiative of the NCI60 panel provided an assembly of tumour cell lines that have been treated against a diverse panel of over 100,000 small molecules [12,13]. The NCI60 dataset was the first large-scale resource enabling identification of lineage-selective small molecule sensitivities [14]. However, the relatively small number of 59 cancer cell lines of the NCI60 panel restricted the relevance of these data for prediction of drug MoA. The Cancer Therapeutics Response Portal (CTRP) has recently addressed this limitation by providing a resource of sensitivity measurements for extensively characterized cancer cell lines tested against a set of nearly 300 small molecules [15,16]. The latest CTRP release, coined CTRPv2, presents the largest quantitative *in vitro* sensitivity dataset available to date, spanning 860 cancer cell lines screened against a set of 481 small molecule compounds [16]. Individual assessment of these *in vitro* sensitivity datasets have highlighted their use towards determining mechanism of growth inhibition, and inference of MoA of compounds from natural products. It remains to be demonstrated, however, whether integration of these drug sensitivity data with other drug-related data, such as drug structures and drug-induced transcriptional signatures, can be used to systematically infer drug MoA.

Comprehensive molecular characterization of drug MoA for newly identified compounds requires high-throughput datasets that encapsulate a widespread range of drugs across multiple cancer cell lines. The aforementioned CTRPv2 sensitivity data perfectly qualifies for these requirements. However, such a dataset is unmatched by corresponding drug perturbation signatures from CMAP, which only characterizes 1309 drugs across 5 cancer cell lines. The CMAP project has recently been superseded by the L1000 dataset from the NIH Library of Integrated Network-based Cellular Signatures (LINCS) consortium [17], which has expanded

upon the conceptual framework of CMAP and contains over 1.4 million gene expression profiles spanning 20,413 chemical perturbations. Accordingly, the L1000 dataset provides an unprecedented compendium of both structural and transcriptomic drug data. A recent integrative study of the LINCS data showed that structural similarity are significantly associated with similar transcriptional changes, supporting the complementarity of these drug-related data [6].

To improve inference of drug MoA for new compounds, we leveraged our recent Similarity Network Fusion algorithm [18] to efficiently integrate drug structure, sensitivity, and perturbation data towards developing a large-scale molecular drug taxonomy, called Drug Network Fusion (DNF) **(Figure 1)**. DNF significantly outperformed taxonomies based on single data types at classifying drugs based on drug targets and therapeutic annotations. Our explorative analysis sheds light on how data integration approach can substantially improve characterization of MoA, both for existing drugs, but more specifically, for new compounds that lack deep pharmacological and biochemical characterization. Our results support DNF as a valuable resource to the cancer research community by providing new hypotheses on the compound MoA and potential insights for drug repurposing.

MATERIAL AND METHODS

A schematic overview of the analysis design is presented in **Figure 2**.

Processing of drug-related data and identification of drug similarity

Drug structure annotations: Canonical SMILES strings for the small molecules were extracted from PubChem [19], a database of more than 60 millions unique structures. Tanimoto similarity measures [20] between drugs were calculated by first parsing annotated SMILES strings for existing drugs through the *parse.smiles* function of the *rcdk* package (version 3.3.2). Extended connectivity fingerprints (hash-based fingerprints, default length 1,024) across all drugs was subsequently calculated using the *rcdk::get.fingerprints* function [21].

Drug perturbation signatures: We obtained transcriptional profiles of cancer lines treated with drugs from the L1000 dataset recently released by the Broad Institute [22], which contains over 1.4 million gene expression profiles of 1000 'landmark' genes across 20,413 drugs. We used our *PharmacoGx* package (version 1.1.4) [23] to compute signatures for the effect of drug concentration on the transcriptional state of a cell, using a linear regression model adjusted for treatment duration, cell line identity, and batch to identify the genes whose expression is significantly perturbed by drug treatment:

$$G = \beta_0 + \beta_i C_i + \beta_t T + \beta_d D + \beta_b B$$

where

G = *molecular feature expression (gene)*

C_i = *concentration of the compound applied*
T = *cell line identity*
D = *experiment duration*
B = *experimental batch*
βs = *regression coefficients.*

The strength of the feature response is quantified by βi. G and C are scaled variables (standard deviation equals to 1) to estimate standardized coefficients from the linear model. The transcriptional changes induced by drugs on cancer cell lines are subsequently referred to throughout the text as *drug perturbation signatures*. Similarity between estimated standardized coefficients of drug perturbation signatures was computed using the Pearson correlation coefficient, with the assumption that drugs similarly perturbing the same set of genes might have similar mechanisms of action.

Drug sensitivity signatures: We obtained summarized dose-response curves from the published drug sensitivity data of the NCI60 [14] and CTRPv2 [16] datasets integrated in the *PharmacoGx* package. We relied on the calculated Z-score and area under the curve (AUC) metrics for NCI60 and CTRPv2, respectively. Drug similarity was defined as the Pearson correlation of drug sensitivity profiles.

Development of a drug network fusion (DNF) taxonomy

We used our Similarity Network Fusion algorithm [18] to identify drugs that have similar mechanisms of actions by integrating three data types representing drug structure, drug perturbation, and drug sensitivity profiles. Drug structure and drug perturbation taxonomies were based on drug-drug similarity matrices computed from the PubChem SMILES and the the L1000 dataset, respectively. The drug sensitivity taxonomy was composed of the drug-drug similarity matrix of the sensitivity signatures extracted from either the NCI60 or CTRPv2 datasets. For each dataset, an affinity matrix was first calculated using the *affinityMatrix* function as described in the *SNFtool* package (version 2.2), using default parameters. We combined the three affinity matrices of the structure, perturbation, and sensitivity taxonomies into a Drug Network Fusion (DNF) matrix using the *SNFtool::SNF* function **(Figure 2)**. Two separate DNF matrices were generated dependant on the sensitivity layer used (either CTRPv2 or NCI60). The developed DNF taxonomies, as well as the single data type taxonomies, were subsequently tested against benchmark datasets to validate their drug mode of action (MoA).

Assessment of drug mode of action across drug taxonomies

Drug-target associations. Known target associations for drugs pertaining to the NCI-60 dataset were downloaded from CHEMBL (file version 15-3-46-00) [24]. Drug-target associations for drugs of the CTRPv2 dataset were obtained from the CTRPv2 website

(http://www.broadinstitute.org/ctrp.v2/?page=#ctd2Target). Drugs with annotated targets were filtered to retain only targets with at least two drugs.

Anatomical therapeutic classification system (ATC). ATC annotations [25] for the drugs common to the NCI60 and CTRPv2 datasets were downloaded from ChEMBL (file version 15-3-18-59) [24]. These ATC codes were filtered to retain only those categories with at least one pair of drugs sharing a pharmacological indication. The drugs with known ATC annotations from the NCI60 and CTRPv2 datasets were subsequently used as a validation benchmark against singular drug taxonomies and the DNF taxonomy.

Evaluation of drug mechanism of action across taxonomies

We assessed the predictive value of our developed taxonomies against drug-target and ATC benchmark datasets to determine the extent to which single data type taxonomies and the DNF taxonomy recapitulate known drug MoA **(Figure 3)**. We adapted the method from Cheng et al [26] to compare benchmarked datasets against singular drug taxonomies (Drug Perturbation, Drug Structure, or Drug Sensitivity) as well as the integrated DNF taxonomy. This method is further detailed below for the benchmark datasets used in our study. First, we created adjacency matrices that indicate whether each pair of drugs share a target molecule or ATC annotation. The drug-target and ATC adjacency matrices were then converted into a vector of similarities between every possible pair of drugs where the value '1' was assigned in the vector if the paired drugs were observed the same target/ATC set, and '0' otherwise. Similarly, the affinity matrices of singular drug taxonomies as well as the DNF taxonomy matrix were converted into vectors of drug pairs, with the similarity value of the drug pairs retained from their original corresponding matrix. Binary vectors of the benchmarks were compared to the four continuous vectors of the drug taxonomies by computing the receiver-operating curves (ROC) using the *ROCR* package (version 1.0.7) [27], and the area under the curve (AUC) using the *concordance.index* function of the *survcomp* package (version 1.18.0) [28]. The AUC estimates the probability that, for two pairs of drugs, drugs that are part of the same drug set (same therapeutic targets or ATC functional annotations) have higher similarity than drugs that do not belong to the same drug set. AUC calculations for each of the four taxonomies were statistically compared against each other using the *survcomp::compare.cindex* function.

Detection of drug communities and visualization

Clusters of drug communities were determined from the DNF taxonomy using the affinity propagation algorithm [29,30] from the *apcluster* package (version 1.4.2). The apcluster algorithm generates non-redundant drug communities, with each community represented by an exemplar drug. An elevated *q* value parameter, which determines the quantiles of similarities to be used as input preference to generate small or large number of clusters, was set at q=0.9 within the *apcluster* function to produce a large number of communities. Networks of exemplar drugs were rendered in *Cytoscape* (version 3.3.0) [31]. Drug structures were rendered using the

chemViz plugin version 1.0.3 for cytoscape [32]. A minimal spanning tree of the exemplar drugs was determined using Kruskal's algorithm as part of the *CySpanningTree* plugin version 1.1 [33] for cytoscape.

Research Reproducibility

All the code and data links required to reproduce this analysis is publicly available on https://github.com/bhklab/drugSNF. The procedure to setup the software environment and run our analysis pipeline is also provided. This work complies with the guidelines proposed by Robert Gentleman [34] in terms of code availability and reproducibility of results.

RESULTS

We developed a large-scale molecular taxonomy, Drug Network Fusion (DNF), by integrating drug structure, drug sensitivity, and perturbation signatures using our recently developed Similarity Network Fusion algorithm [18]. Drug structure (SMILES representations) were extracted from the PubChem database, containing 60 million compounds. Drug perturbation signatures, representing drug-induced gene expression changes, were extracted from the recent LINCS L1000 dataset. Drug sensitivity signatures representing cell line viability across cancer cell lines were extracted from the CTRP portal, which contains pharmacological profiles of several hundred cell lines **(Supplementary Figure 1)**. We have tested the robustness of our approach by also generating a DNF taxonomy using the NCI60 sensitivity dataset, which contains pharmacological profiles for only 60 cell lines but spans thousands of drugs **(Supplementary Figure 1)**. Collectively, both tests serve to span a large spectrum of sensitivity signatures across both drug compounds and cancer cell lines. Using CTRPv2, our DNF taxonomy is composed of 239 drugs for which all of drug structure, drug perturbation, and drug sensitivity information could be fused. Using NCI60, a total of 238 common drugs were used. Notably, the overlap between the drugs of the NCI60 and CTRP datasets is small **(**64 drugs; **Supplementary Figure 1)**, which underscores the complementarity of these two datasets.

To demonstrate the benefit of our integrative approach, we assessed the predictive value of the DNF taxonomy for drug targets and functional classification and compared it to only using structure, sensitivity, or perturbation data alone. In addition, we used affinity propagation clustering (APC) on the DNF taxonomy to determine communities of drugs which share a similar MoA.

Performance of drug taxonomies against known drug targets

Determining drug-target interactions is important in the drug development process. The identification of new targets opens new avenues for drug repurposing efforts, and suggests new pathways and mechanisms by which drugs can operate in cells. Drug targets were identified for 193 drugs in CTRPv2 and these drugs were filtered to retain target categories with more than one drug, resulting in a set of 141 drugs available for benchmarking. Similarly, drug targets were

identified for 101 drugs in NCI60, from which 73 drugs shared a target with at least another drug.

We assessed the predictive value of our single-data layer and integrative drug taxonomies against drug targets and ATC functional classification. We performed a ROC analysis to quantify how well our drug taxonomies align with established drug target **(Figure 3)**. By statistically comparing the resulting AUC values, we were able to determine whether our integrative drug taxonomy outperformed taxonomies based on a singular data analyses **(Figure 3, Table 1)**. We tested how our integrated taxonomy using the CTRPv2 drug sensitivity taxonomy compares against single-layers for drug-target designations from ChEMBL **(Figure 4A)**. Of the three single-layer taxonomies validated against annotated drug targets from CTRPv2, the drug sensitivity layer outperformed the structure and perturbation taxonomies (AUC of 0.83, 0.71 and 0.64 for sensitivity, structural and perturbation data layers, respectively) **(Figure 4A)**. Importantly, DNF yielded the best predictive value (AUC of 0.89, **Figure 4A**), and was significantly higher than any single-layer taxonomy (one-sided t test p-value < 1E-16, **Table 1**).

We replicated our integrative taxonomy approach using the set of drug sensitivity signatures obtained from the NCI60 dataset where a much smaller panel of cell lines has been screened (60 vs. 860 cell lines for NCI60 and CTRPv2, respectively). This integrative taxonomy **(Supplementary Figure 2)** was generated and validated against the drug-target benchmark from ChEMBL databases since no drug-targets annotation were provided from the NCI60 site. Our evaluation of single-layer taxonomies demonstrates that drug similarities based on sensitivity signatures were the most efficient in predicting drug-target associations (AUC of 0.69; **Supplementary Figure 2A**) compared to structure and perturbation (AUC of 0.61 and 0.49, respectively; **Supplementary Figure 2A**). DNF was significantly more predictive of drug-target associations compared to single-layer taxonomies from structure and perturbation but not sensitivity (AUC of 0.70 and one-sided superiority test p-values < 0.05, **Supplementary Figure 2A, Table 1**).

Performance of drug taxonomies against known functional classes

Predicting the anatomical classification (ATC) of a drug provides existing and new insights about its pharmacological mechanism, and ultimately presents new potential indications for previously uncharacterized drugs. ATC codes were identified for 59 and 122 drugs pertaining to CTRPv2 and NCI60, respectively. These codes were filtered to retain only those categories with at least one pair of drugs sharing a pharmacological indication. A total of 43 and 88 drugs with known ATC annotations from the CTRPv2 and NCI60 datasets, respectively, were subsequently used for performance assessment.

We conducted a second validation of our taxonomies against ATC drug classification **(Figure 4B)**. Drug sensitivity was not the most predictive layer for ATC classification and exhibited comparable predictive power as drug perturbation **(Figure 4B)**. The structure-based taxonomy **(Figure 4B)** was the most predictive amongst single-layer taxonomies (AUC of 0.72, 0.57 and 0.54 for structure, sensitivity, and perturbation layers, respectively). The integrative drug taxonomy significantly outperformed single-layer taxonomies (AUC of 0.77 with one-sided t

test p-value < 0.05; **Figure 4B, Table 1**). Interestingly, DNF outperforms single-layer taxonomies when tested for functional classification based on ATC (AUC of 0.87 with one-sided t test p-values < 0.05; **Supplementary Figure 2B, Table 1**). Similarly, as observed with CTRPv2, structural similarity remains the best performing single-layer taxonomy when tested against ATC classification **(Supplementary Figure 2)**.

Identification of Drug Communities Using DNF Taxonomy

To assess the biological relevance of integrative drug taxonomy in discovering drugs with similar MoA, we applied the affinity cluster propagation algorithm [30] to identify clusters of highly similar drugs referred to as *drug communities* **(Figure 5, Supplementary Figure 3)**. These communities can be represented by their most representative drug and the similarities between communities represented a network where each node is labeled by the exemplar drug. Our initial analysis of the DNF taxonomy based on CTRPv2 sensitivity identified 53 communities **(Table 2)**. Of these, we identified 39 drug communities **(Table 3)**, which have at least two drugs with a known mechanism of action.

Overall, our integrative taxonomy developed using the CTRPv2 has produced a substantial and consistent classification of drugs for a variety of functional classes **(Table 4)**. Briefly our classifications recapitulate most of the protein target-drug associations represented in CTRPv2: Receptor tyrosine kinases and non-receptor tyrosine kinases (including EGFR, VEGFR, ALK, ABL1, SRC, RAF, MEK, IGFR-1) inhibitors, PI3K/mTOR family inhibitors, proapoptotic (including the p53 tumor suppressor) and anti-apoptotic (including the MDM2 and BCL-2 oncogenes) inhibitors, epigenetic regulators (HDACs) inhibitors, glycosyltransferase NAMPT inhibitors, cell cycle kinases inhibitors (CDKs, PLK, ATM), DNA replication (topoisomerases), repair and synthesis (TYMS) inhibitors, HMG CoA and proteasome inhibitors **(Table 2 and Supplementary Table 3)**.

We replicated our integrative taxonomy using the NCI60 sensitivity dataset **(Supplementary Figure 1)**, and identified 51 communities **(Supplementary Table 1)**, of which 20 communities **(Supplementary Table 2)** showed at least two drugs with a known mechanism of action. We are aware that an important number of drugs has unannotated target and ATC codes, as most of the drugs in this study are experimental or uncharacterized chemicals in NCI60, however for reproducibility and validation concerns we did not manually annotated our compound collections **(Table 4)**.

DISCUSSION

Identification of MoA for newly uncharacterized compounds is a key challenge towards characterizing on-targets responsible of pharmacological effect and off-targets associated with unexpected physiological effects. Shortcomings of current approaches include a degree of reliance on pharmacological, biochemical, and functional annotations that pertain to existing, well-characterized drugs, and which may not be applicable towards prediction of a new small compounds **(Figure 6)** [11][35]. Compounding this issue is the absence of a high-throughput,

integrative classification that merges complementary and basic drug characteristics, such as chemical structure, *in vitro* drug sensitivity and transcriptional perturbation signatures. Such shortcomings have not only hindered efficient classification of new drugs, but also pose an obstacle towards proper evaluation of the current methods for drug taxonomy inference **(Figure 6)**. Our analysis addresses these issues by conducting, to our knowledge, the first large-scale integration of drug structure, sensitivity and perturbation signatures towards prediction of drug MoA. We demonstrate how the DNF taxonomy represents a new resource that can be mined to uncover relationships between small molecule compounds and new mechanisms of action.

We have capitalized upon our integrative Similarity Network Fusion method [18] to construct a high-throughput drug similarity network (DNF), based on the fusion of drug structure, sensitivity, and perturbation data. The construction of drug-similarity networks and their subsequent fusion allows us to fully harness the complementary nature of several drug datasets, and generate an informative clustering of drugs across multiple data types. We have previously demonstrated how Similarity Network Fusion substantially outperforms single data type analysis [18] in an analysis of genomic data across several cancers and we demonstrated here that this holds true in the context of drug taxonomy inference. Testing how well different drug taxonomies correctly predict drug targets **(Figure 4A)** and anatomical (ATC) drug classifications **(Figure 4B)**, indicates that DNF constitutes a marked improvement towards drug classification, compared to single data type analyses using either drug sensitivity, structure, or perturbation information alone. This observation is sustained even with the use of a different type and scale of *in vitro* sensitivity data **(Supplementary Figure 1)** to generate the DNF matrix **(Supplementary Figure 2)**. Accordingly, our integrative approach succeeds in combining several drug data types into a single comprehensive network that represents the full spectrum of the underlying data.

Relying on drug-related data that only encompasses drug sensitivity, structure, and perturbation profiles ultimately presents a flexible approach towards comprehensive drug classification. We have removed any reliance on existing pharmacological, biochemical, or functional annotations that pertain to existing drugs, such as drug-target classifications or knowledge of the anatomical and organ system targeted by the drug compounds. Accordingly, our DNF method only requires basic drug information, including drug structures, sensitivity, and perturbation profiles, to determine drug MoA. These types of data, compared to other mechanistic annotations including ATC or drug target information, are much easier to generate for newly uncharacterized compounds, which ultimately facilitates proper characterization of new compounds. This promises to provide a more extensive characterization of the compound across multiple manifolds of drug associations, and ultimately allows us to test our DNF drug associations against both drug-target and anatomical therapeutic classifications (ATC).

Comparing our integrative DNF taxonomy with single data layers revealed the importance of drug sensitivity information towards improving prediction performance of drug-target associations **(Figure 4A).** Such findings support the relevance of bioactivity assays to predict drug targets, and underscore the comprehensive nature of the CTRPv2 dataset (860 cell lines screened with 16 drug concentrations, tested in duplicate) [16]**.** Similarity, we have observed a priority for drug structure information towards prediction of ATC drug classification **(Figure 4B).** Our approach thus exemplifies how DNF and singular taxonomies are compared

against a number of drug benchmarks, and highlights the interplay between different types of data for generating relevant drug classifications.

The DNF taxonomy highlights many cases of drug clusters with known mechanism of action, capturing context-specific features associated to drug sensitivity and genomic profiles in cancer cells. These cases, to some extent, serve as experimental validation of our method. We classified correctly all BRAF (V600E mutation) inhibitors, which include drugs already tested in metastatic melanoma (community C18:dabrafenib, GDC0879, PLX4720) and mitogen-activated protein kinase/ERK kinase (MEK) inhibitors (C41: namely trametinib and selumetinib). BRAF regulates the highly conserved MAPK/ERK signaling pathway, and BRAF mutational status has been proposed as a biomarker of sensitivity towards selumetinib and other MEK inhibitors [36,37]. This explains the tight connection of these two communities (**Figure 5**).

The DNF taxonomy also represents a new and comprehensive resource that can be mined to uncover new relationships between drugs and mechanisms of action. We identified a community of HMG Co-A reductase inhibitors (statins) composed of fluvastatin, lovastatin, and simvastatin (C30; **Figure 5**). These are a class of cholesterol-lowering drugs, and which have been found to reduce cardiovascular disease. Interestingly, parthenolide is the only drug clustering with this community, and has been experimentally observed to inhibit the NF-Kb inflammatory pathway in atherosclerosis and in colon cancer [38,39], thereby exhibiting similar behavior to statin compounds. By inhibiting similar targets and modulating similar pathways as statins, these findings suggest that parthenolide may present a statin-like MoA. We also classified correctly drugs with unannotated mechanisms/targets in CTRPv2 such as ifosfamide, cyclophosphamide and procarbazine (C17; **Figure 5**) which are known alkylating agents (ATC code: L01A). Furthermore, this was also true for docetaxel and paclitaxel (C21; **Figure 5**), two taxanes drugs with unannotated target in CTRPv2 (ATC code: L01CD).

Our integrative drug taxonomy is also able to identify targets for drugs with poorly understood mechanisms and to infer new mechanism for other drugs. Community C15, for example, contains tigecycline and Col-3 (**Figure 5**); both are derivatives of the antibiotic tetracycline [40]. Tigecycline is an approved drug, however its target is not characterized in humans. Col-3 showed antitumor activities by inhibiting matrix metalloproteinase [40]. Interestingly, tosedostat (CHR-2797), a metalloenzyme inhibitor with antiproliferative potential, is also a member of this community [41]. Another drug in this community, phloretin, is a natural compound with uncharacterized targets and has been recently shown to deregulate matrix metalloproteinases at both gene and protein levels [42]. Our results suggest that matrix metalloproteinases would be the preferred target for drugs in this community, supporting the need for further experimental investigation. DNF also consolidated previous findings for drugs that may serve as tubulin polymerization disruptors, and which have not been previously classified as such. We identified a community of 3 drugs (C49) in which LY2183240, and YK-4-279 have been recently identified to decrease alpha-tubulin levels [16]. TIVANTINIB a c-MET tyrosine kinase inhibitor also blocked microtubule polymerization [43]. Interestingly, this community is tightly connected to known microtubule perturbators (community C21; **Figure 5**).

Our results also concur with the study of Rees et al. [44] regarding cluster of the BCL-2 inhibitors ABT-737 and navitoclax (community C33; **Figure 5**), where the authors reported that a high expression of BCL-2 confers sensitivity to these two drugs. This was not the case for

another BCL-2 inhibitor, obatoclax. They proposed that a metabolic modification of obatoclax in cells impacts its interaction with BCL-2 proteins, therefore reducing its potency. We showed indeed that obatoclax did not cluster with the other two BCL-2 inhibitors (ABT-737 and navitoclax). Such an example demonstrates how the structural and sensitivity profiles of these two BCL-2 inhibitors are largely coherent in contrast to obatoclax, which previously showed off-target effects compared to ABT-737 [45]. This provides a good evidence to consider sensitivity profiles when developing new potent and specific BCL-2 inhibitors.

Our results suggest the existence of "super communities", that are a grouping of several communities sharing similar MoA, or contributing to a larger, systems-based MoA. An example is provided by the tightly connected communities C3, C21, C23, C43. One of these communities (C3: Alvocidib, PHA-793887 and staurosporine) includes well-characterized inhibitors of cyclin dependant kinases (CDKs) that are known to be major regulators of the cell cycle. BMS-345541 for example, which also clusters with drugs in C3, is an ATP non-competitive allosteric inhibitor of CDK [46]. Those compounds are positioned close in the community network to topoisomerase I and II inhibitors (C43: SN-38, topotecan, etoposide, teniposide), microtubule dynamics perturbators (C21: paclitaxel, docetaxel, vincristine, parbendazole) and polo-like kinase inhibitors (C23: GSK461364, GW843682X). Iorio *et al.*, reported that the similarity between CDK inhibitors and the other DNA-damaging agents is mediated through a p21 induction, which explains the interconnection and rationale of similar transcriptional and sensitivity effects [8] of these regulators of cell cycle progression.

Our study suggests that drug sensitivity data is an important asset for computational methods that predict drug mechanism of action. To test the stability of the fusion algorithm with respect to the scale of the drug sensitivity profiles, we also applied our methodology on the NCI60 dataset, which comprises a much smaller panel of cell lines (60 vs. 860 for NCI60 and CTRPv2, respectively). The NCI60 panel compensates for its small cell line panel by the large number of screened drugs (>40,000 drugs tested on the full panel; **Supplementary Figure 1**). Testing DNF using the NCI60 sensitivity information reveals that our integrative taxonomy continues to supersede single-layer drug taxonomies across varying benchmarks **(Supplementary Figure 2)**. Interestingly, some of the identified communities using NCI60, such as the tight connection between BRAF/MEK inhibitor drugs (C42; **Supplementary Figure 3**), had also been identified in our original analysis using CTRPv2 sensitivity profiles. This demonstrates a high degree of specificity of drug-target associations across cell lines and experimental platforms, which is crucial in biomarker identification and translational research.

The DNF taxonomy encompassing the NCI60 dataset has also identified a number of well-characterized drug communities **(Supplementary Tables 1-3)**. These include the community composed of EGFR inhibitors (C20; **Supplementary Figure 3**). Our results for community C14 (cardiac glycosides) also concur with the study of Khan et al [5] (**Supplementary Figure 3**). These compounds inhibit Na+/K+ pumps in cells. Using a 3D chemical descriptor approach combined with genomic features, Khan *et al* had also identified bisacodyl, a laxative drug, as sharing a similar mechanism with cardiac glycosides, despite its structural dissimilarity to that class of compounds [5]. Notably, our integrative taxonomy recapitulates these findings, which demonstrates that combination of structural and genomic drug information is a promising strategy towards elucidating drug mechanisms.

Our DNF based on NCI60 sensitivity information enabled identification of new drugs with uncharacterized MoA that we believe warrant further experimental investigation. We found that communities C2, C5, C32, and C51 were closely connected (**Supplementary Figure 3**). These communities contain a number of compounds which showed antitumor activity by generating reactive oxygen species (e.g. C2: elesclomol, fenretinide; C5: ethacrynic acid, curcumin; C32: bortezomib, menadione; C51: celastrol, withaferin A, parthenolide, thapsigargin). Interestingly, ethacrynic acid, an FDA approved drug indicated for hypertension, clustered with curcumin, a component of turmeric. Ethacrynic acid inhibits glutathione S-transferase (GSTP1) and induced mitochondrial dependant apoptosis through generation of reactive oxygen species (ROS) and induction of caspases [47]. Curcumin showed antitumor activity by production of ROS and promotion of apoptotic signaling. Thus, we suggest that GSTP1 could be a potential target of the widely-used natural compound curcumin.

In conclusion, we have developed Drug Network Fusion (DNF), an integrative taxonomy inference approach leveraging the largest quantitative compendiums of structural information, pharmacological phenotypes and transcriptional perturbation profiles to date. We used DNF to conduct a cross-comparative assessment between our integrative taxonomy, and single-layer drug taxonomies based on either drug structure, perturbation, or sensitivity signatures. Our exploratory analysis indicates the superiority of DNF towards drug classification, and also highlights singular data types that are pivotal towards prediction of drug categories in terms of anatomical classification as well as drug-target relationships. Overall, the DNF taxonomy has produced a consistent classification of drugs for multiple functional classes in both CTRPv2 and NCI60 **(Table 4).** The comprehensive picture of drug-drug relationships produced by DNF has also succeeded in identifying new and potentially interesting drug MoA. The integrative DNF taxonomy has the potential to serve as a solid framework for future studies involving inference of MoA of new, uncharacterized compounds, which represents a major challenge in drug development for precision medicine.

ACKNOWLEDGEMENTS

The authors would like to thank Jacques Archambault for sharing his expertise in pharmacology and financially supporting Nehme El-Hachem.

FUNDING

This study was conducted with the support of the Canadian Cancer Research Society and the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. Z.S. was supported by the Cancer Research Society (Canada). B.H.K. was supported by the Gattuso-Slaight Personalized Cancer Medicine Fund at Princess Margaret Cancer Centre.

REFERENCES

1. Sheridan RP, Kearsley SK. Why do we need so many chemical similarity search methods? Drug Discov Today. 2002;7: 903–911.
2. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462: 175–181.
3. Dunkel M, Günther S, Ahmed J, Wittig B, Preissner R. SuperPred: drug classification and target prediction. Nucleic Acids Res. 2008;36: W55–9.
4. Martin YC, Kofron JL, Traphagen LM. Do structurally similar molecules have similar biological activity? J Med Chem. 2002;45: 4350–4358.
5. Khan SA, Virtanen S, Kallioniemi OP, Wennerberg K, Poso A, Kaski S. Identification of structural features in chemicals associated with cancer drug response: a systematic data-driven analysis. Bioinformatics. 2014;30: i497–504.
6. Chen B, Greenside P, Paik H, Sirota M, Hadley D, Butte AJ. Relating Chemical Structure to Cellular Response: An Integrative Analysis of Gene Expression, Bioactivity, and Structural Data Across 11,000 Compounds. CPT Pharmacometrics Syst Pharmacol. 2015;4: 576–584.
7. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313: 1929–1935.
8. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci U S A. 2010;107: 14621–14626.
9. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008;321: 263–266.
10. Kuhn M, Al Banchaabouchi M, Campillos M, Jensen LJ, Gross C, Gavin A-C, et al. Systematic identification of proteins that elicit drug side effects. Mol Syst Biol. 2013;9: 663.
11. Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, et al. Drug repositioning: a machine-learning approach through data integration. J Cheminform. 2013;5: 30.
12. Driscoll JS. The preclinical new drug research program of the National Cancer Institute. Cancer Treat Rep. 1984;68: 63–76.
13. Luna A, Rajapakse VN, Sousa FG, Gao J, Schultz N, Varma S, et al. rcellminer: exploring molecular profiles and drug response of the NCI-60 cell lines in R. Bioinformatics. 2015; doi:10.1093/bioinformatics/btv701
14. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006;6: 813–823.
15. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell. 2013;154: 1151–1161.
16. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. Cancer Discov. 2015;5: 1210–1223.
17. NIH, Broad Institute. The LINCS Connectivity Map Project. In: The LINCS Connectivity Map Project [Internet]. 2013 [cited 2014]. Available: http://lincscloud.org/
18. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11: 333–337.
19. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and

Compound databases. Nucleic Acids Res. 2016;44: D1202–13.
20. Tanimoto TT. An Elementary Mathematical Theory of Classification and Prediction. International Business Machines Corporation; 1958.
21. Guha R, Others. Chemical informatics functionality in R. J Stat Softw. 2007;18: 1–16.
22. Duan Q, Flynn C, Niepel M, Hafner M, Muhlich JL, Fernandez NF, et al. LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. Nucleic Acids Res. 2014;42: W449–60.
23. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, et al. PharmacoGx: An R package for analysis of large pharmacogenomic datasets. Bioinformatics. 2015; doi:10.1093/bioinformatics/btv723
24. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res. 2014;42: D1083–90.
25. Nahler G. Anatomical therapeutic chemical classification system (ATC). In: Nahler G, editor. Dictionary of Pharmaceutical Medicine. Springer Vienna; 2009. pp. 8–8.
26. Cheng J, Xie Q, Kumar V, Hurle M, Freudenberg JM, Yang L, et al. Evaluation of analytical methods for connectivity map data. Pac Symp Biocomput. 2013; 5–16.
27. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21: 3940–3941.
28. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27: 3206–3208.
29. Bodenhofer U, Kothmeier A, Hochreiter S. APCluster: an R package for affinity propagation clustering. Bioinformatics. 2011;27: 2463–2464.
30. Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007;315: 972–976.
31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13: 2498–2504.
32. chemViz2: Cheminformatics App for Cytoscape [Internet]. [cited 1 Mar 2016]. Available: http://www.cgl.ucsf.edu/cytoscape/chemViz2/index.shtml
33. Shaik F, Bezawada S, Goveas N. CySpanningTree: Minimal Spanning Tree computation in Cytoscape. F1000Res. 2015;4. doi:10.12688/f1000research.6797.1
34. Gentleman R, Lang DT. Statistical analyses and reproducible research. J Comput Graph Stat. Taylor & Francis; 2012; Available: http://amstat.tandfonline.com/doi/abs/10.1198/106186007X178663
35. Ma'ayan A, Rouillard AD, Clark NR, Wang Z, Duan Q, Kou Y. Lean Big Data integration in systems biology and systems pharmacology. Trends Pharmacol Sci. 2014;35: 450–460.
36. Friday BB, Yu C, Dy GK, Smith PD, Wang L, Thibodeau SN, et al. BRAF V600E disrupts AZD6244-induced abrogation of negative feedback pathways between extracellular signal-regulated kinase and Raf proteins. Cancer Res. 2008;68: 6145–6153.
37. Solit DB, Garraway LA, Pratilas CA, Sawai A, Getz G, Basso A, et al. BRAF mutation predicts sensitivity to MEK inhibition. Nature. 2006;439: 358–362.
38. Riganti C, Doublier S, Costamagna C, Aldieri E, Pescarmona G, Ghigo D, et al. Activation of nuclear factor-kappa B pathway by simvastatin and RhoA silencing increases doxorubicin cytotoxicity in human colon cancer HT29 cells. Mol Pharmacol. 2008;74: 476–484.
39. López-Franco O, Hernández-Vargas P, Ortiz-Muñoz G, Sanjuán G, Suzuki Y, Ortega L, et al. Parthenolide modulates the NF-kappaB-mediated inflammatory responses in

experimental atherosclerosis. Arterioscler Thromb Vasc Biol. 2006;26: 1864–1870.

40. Syed S, Takimoto C, Hidalgo M, Rizzo J, Kuhn JG, Hammond LA, et al. A phase I and pharmacokinetic study of Col-3 (Metastat), an oral tetracycline derivative with potent matrix metalloproteinase and antitumor properties. Clin Cancer Res. 2004;10: 6512–6521.
41. Krige D, Needham LA, Bawden LJ, Flores N, Farmer H, Miles LEC, et al. CHR-2797: an antiproliferative aminopeptidase inhibitor that leads to amino acid deprivation in human leukemic cells. Cancer Res. 2008;68: 6669–6679.
42. Ma L, Wang R, Nan Y, Li W, Wang Q, Jin F. Phloretin exhibits an anticancer effect and enhances the anticancer ability of cisplatin on non-small cell lung cancer cell lines by regulating expression of apoptotic pathways and matrix metalloproteinases. Int J Oncol. 2016;48: 843–853.
43. Katayama R, Aoyama A, Yamori T, Qi J, Oh-hara T, Song Y, et al. Cytotoxic activity of tivantinib (ARQ 197) is not due solely to c-MET inhibition. Cancer Res. 2013;73: 3087–3096.
44. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat Chem Biol. 2016;12: 109–116.
45. Vogler M, Weber K, Dinsdale D, Schmitz I, Schulze-Osthoff K, Dyer MJS, et al. Different forms of cell death induced by putative BCL2 inhibitors. Cell Death Differ. 2009;16: 1030–1039.
46. Bogoyevitch MA, Fairlie DP. A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. Drug Discov Today. 2007;12: 622–633.
47. Wang R, Liu C, Xia L, Zhao G, Gabrilove J, Waxman S, et al. Ethacrynic Acid and a Derivative Enhance Apoptosis in Arsenic Trioxide--Treated Myeloid Leukemia and Lymphoma Cells: The Role of Glutathione S-Transferase P1-1. Clin Cancer Res. AACR; 2012;18: 6690–6701.

FIGURES

Figure 1: Schematic representation of the SNF method and its use towards integration of different types of drug information. Datasets representing drug similarity, drug sensitivity, and drug perturbation profiles are first converted into drug-drug similarity matrices. Similarity matrices are fully integrated within the SNF method to generate a large-scale, multi-tier, Drug Fusion Network (DNF) taxonomy of drug-drug relationships.

Figure 2: Overview of the study design. Drug sensitivity profiles from the NCI60 and the CTRPv2 datasets, along with drug perturbation and drug structure data from the L1000 dataset, are first parsed into drug-drug similarity matrices that represent single-dataset drug taxonomies. Two DNF taxonomies are generated using the drug perturbation and drug structure taxonomies and the drug sensitivity taxonomy from either the NCI60 or CTRPv2 datasets. DNF taxonomies and single-dataset taxonomies are tested against benchmarked datasets containing ATC drug classification and drug-target information, to validate their efficacy in predicting drug MoA. Additional clustering is conducted on DNF taxonomies to identify drug communities sharing a MoA.

Figure 3: Schematic representation of the validation of the DNF and single data type analyses against drug benchmarks. Drug taxonomies are converted into a continuous vector of drug-drug pairs. Benchmark datasets are converted into binary vectors, whereby a given drug-drug pair is assigned a value of '1' if the drugs share a common drug target or ATC classification, and '0' otherwise. Vectors are compared using the concordance index and the area under the curve (AUC) is calculated from the receiver-operating curves (ROCs).

Figure 4: Validation of the DNF taxonomy (using CTRPv2 sensitivity data) and single dataset taxonomies against the ATC and Drug-target benchmarks. ROC curves are shown for each of the taxonomies generated with the CTRPv2 sensitivity dataset, tested against ATC annotations and drug-target information from Chembl or internal benchmarks. A diagonal (red) representing the null case (AUC=0.5) is drawn for clarity. A) ROC curve against drug-targets B) ROC curve against ATC drug classifications.

Figure 5: Network representation of 51 exemplar drugs that are representative of the drug communities identified by the DNF taxonomy (using CTRPv2 sensitivity data). Each node represents the exemplar drugs, and node sizes reflect the size of the drug community represented by the exemplar node. Nodes are colored to reflect shared MoA as determined using the drug-target benchmark used for Figure 4. Communities sharing similar MoA and proximity in the network are highlighted, with the community number indicated next to each community. Drug communities pertaining to the super-community are labelled in red.

Figure 6: Schematic of the adaptability of DNF towards prediction of new experimental compounds.

TABLES

Table 1. Statistical comparison of the DNF taxonomy against single datasets taxonomies, using one-sided superiority tests. Comparisons were conducted for both DNFs generated using the CTRPv2 or the NCI60 datasets. Reported scores pertain to comparisons conducted using both drug benchmarks (Drug-target information as well as ATC).

Table 2. List of identified communities using the APC cluster algorithm against the DNF (generated using CTRPv2). Exemplar drugs for each community are identified, along with the number of drugs in that community. The list of drugs pertaining to each community is indicated. Drug populations are coloured to indicate communities that have in green means that they have at least 2 drugs with a known mechanism of action intersecting with the GMT file (total 139 drugs in case of ctrpv2, green), and those communities where drugs are unlabeled or unclassified (orange).

Table 3. Refined list of identified communities using the APC cluster algorithm against the DNF (generated using CTRPv2), selected for communities that have at least two drugs with a known mechanism of action. Exemplar drugs for each community are identified, along with the number of drugs in that community. The list of drugs pertaining to each community is indicated.

Table 4. Summary of Functional Drug Classes Identified Using DNF

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Overlap of drug annotations across the L1000 and the NCI60 and CTRPv2 sensitivity datasets.

Supplementary Figure 2: Validation of Single-dataset and DNF taxonomies against drug benchmark datasets, based on the replicated DNF using NCI60 sensitivity data. ROC curves are shown for each of the taxonomies, tested against ATC annotations and drug-target information from Chembl or internal benchmarks. A diagonal (red) representing the null case (AUC=0.5) is drawn for clarity. A) ROC curve for NCI60 against drug-targets B) ROC curve for NCI60 against ATC

Supplementary Figure 3: Community of 53 Exemplar drugs of the DNF taxonomy using the NCI60 sensitivity datasets. Communities sharing similar MoA and proximity in the network are highlighted, with the community number indicated.

SUPPLEMENTARY TABLES

Supplementary Table 1. List of identified communities using the APC cluster algorithm against the DNF (generated using NCI60). Exemplar drugs for each community are identified, along with the number of drugs in that community. The list of drugs pertaining to each community is indicated. Drug populations are coloured to indicate communities that have in green means that they have at least 2 drugs with a known mechanism of action intersecting with the GMT file, and those communities where drugs are unlabeled or unclassified (orange).

Supplementary Table 2. Refined list of identified communities using the APC cluster algorithm against the DNF (generated using NCI60), selected for communities that have at least two drugs with a known mechanism of action. Exemplar drugs for each community are identified, along with the number of drugs in that community. The list of drugs pertaining to each community is indicated.

Supplementary Table 3. Summary of communities generated from CTRPv2/L1000 integrative layers showing positive controls cases (at least 2 drugs sharing a mechanism of action from the same community).

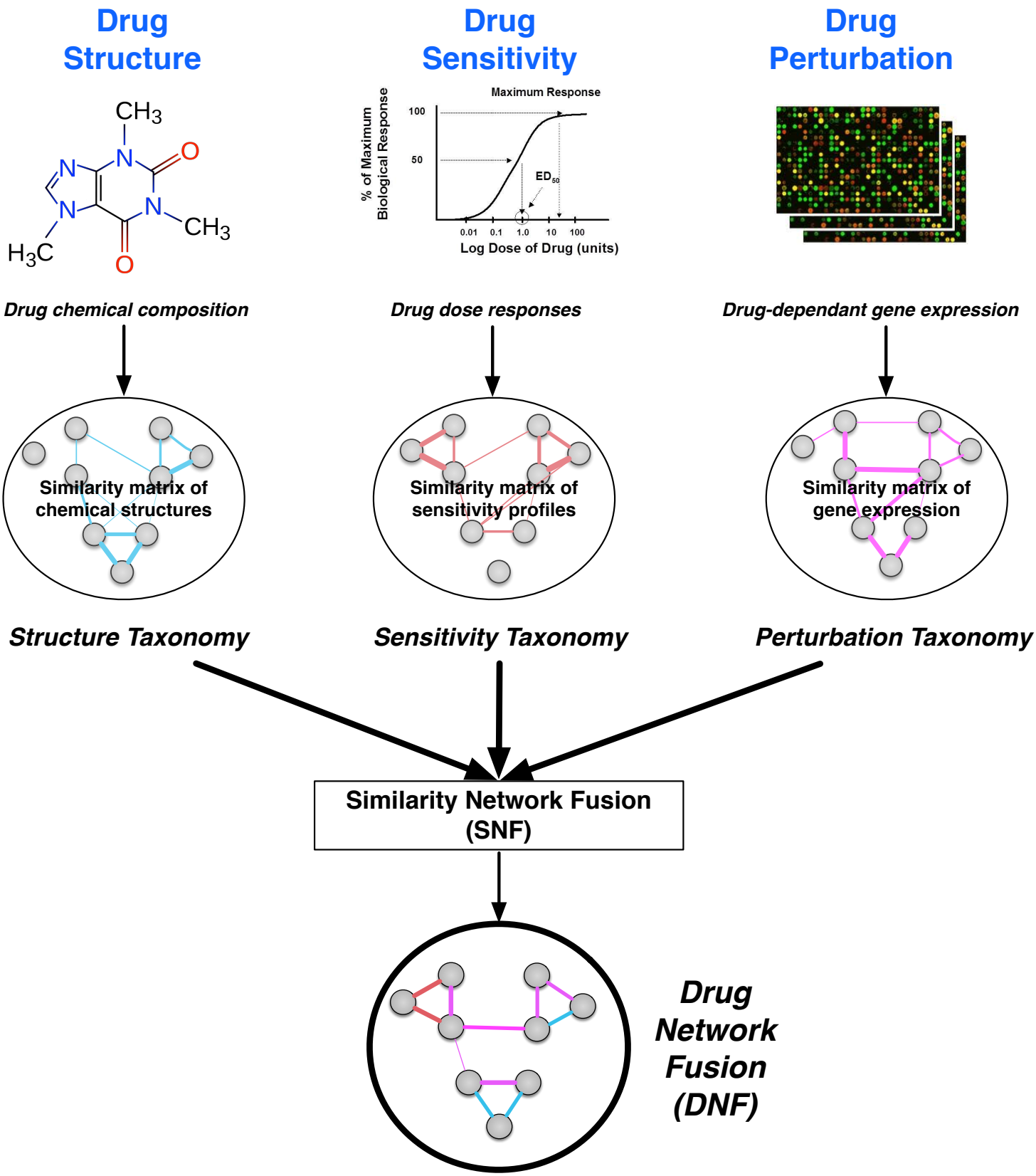


FIGURE 1

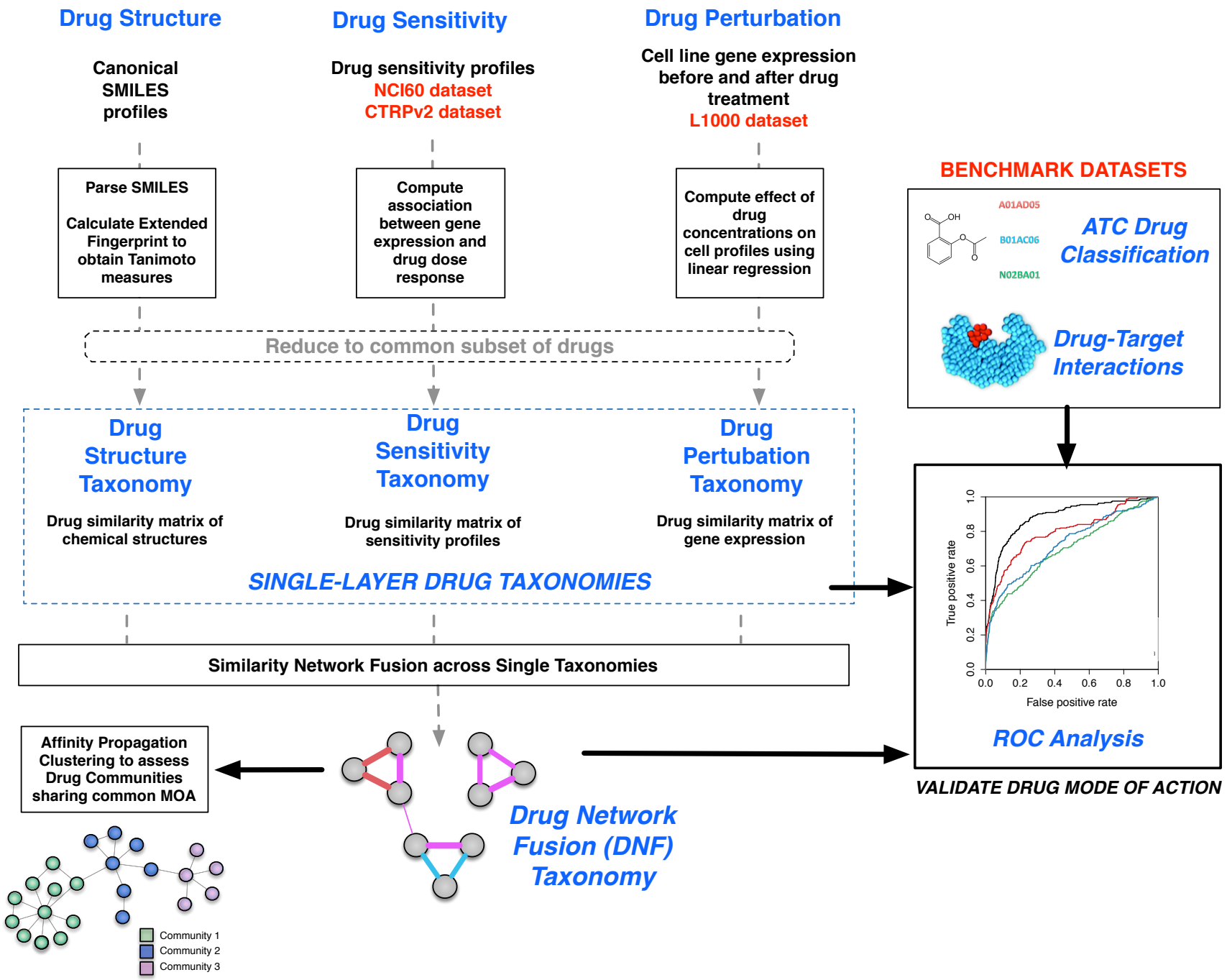


FIGURE 2

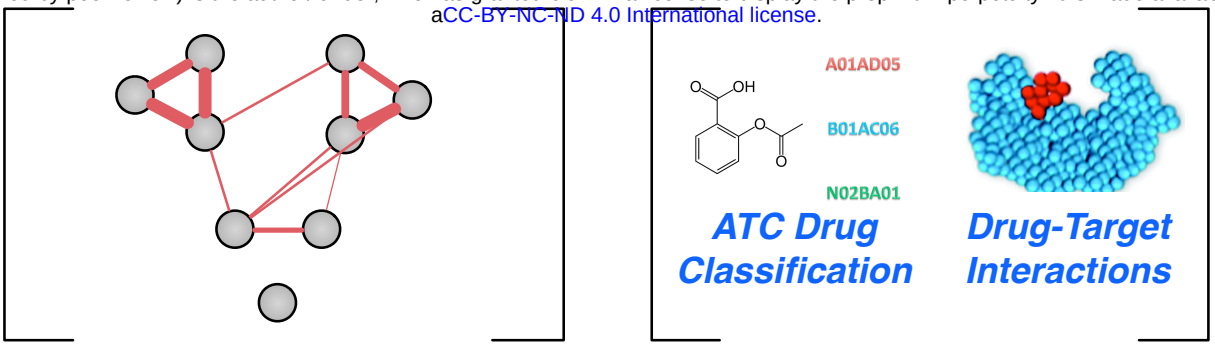
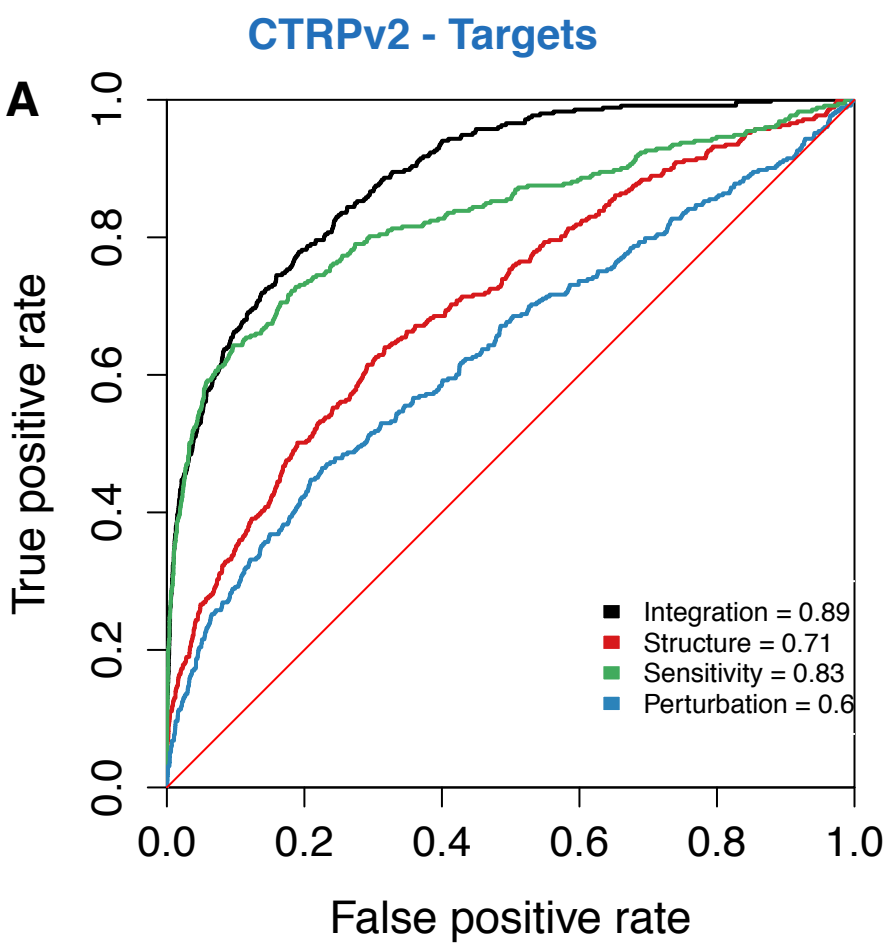


FIGURE 3



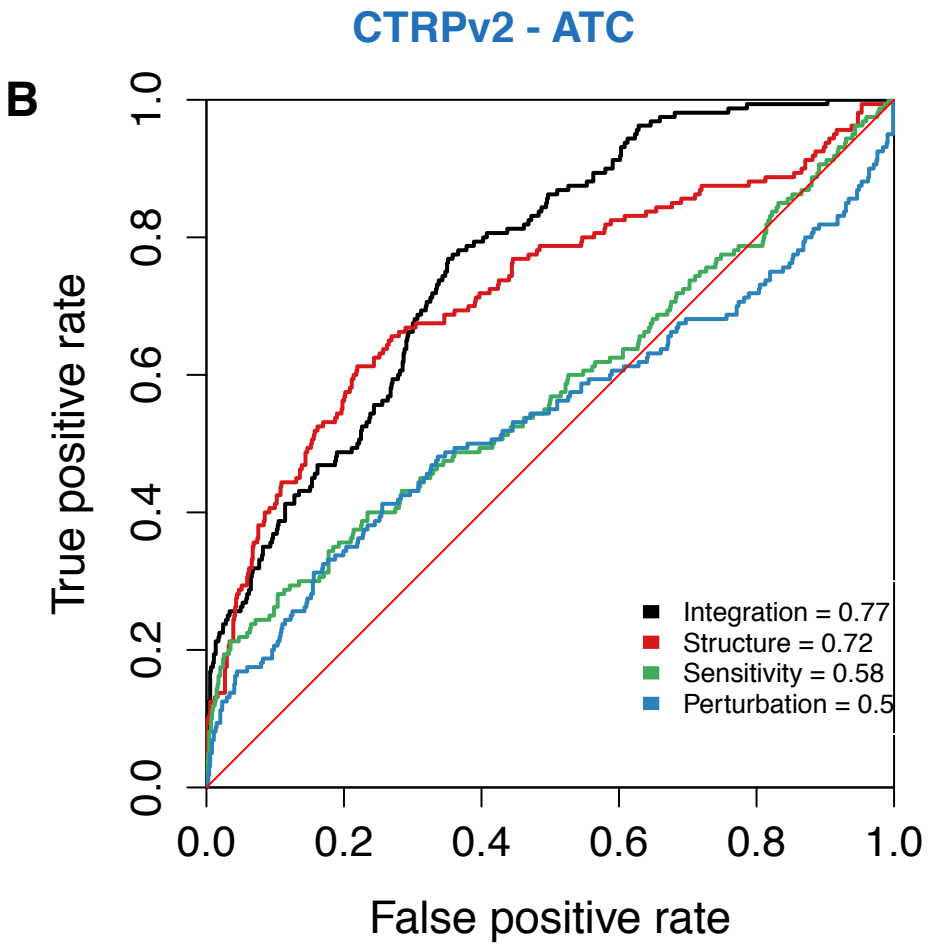


FIGURE 4

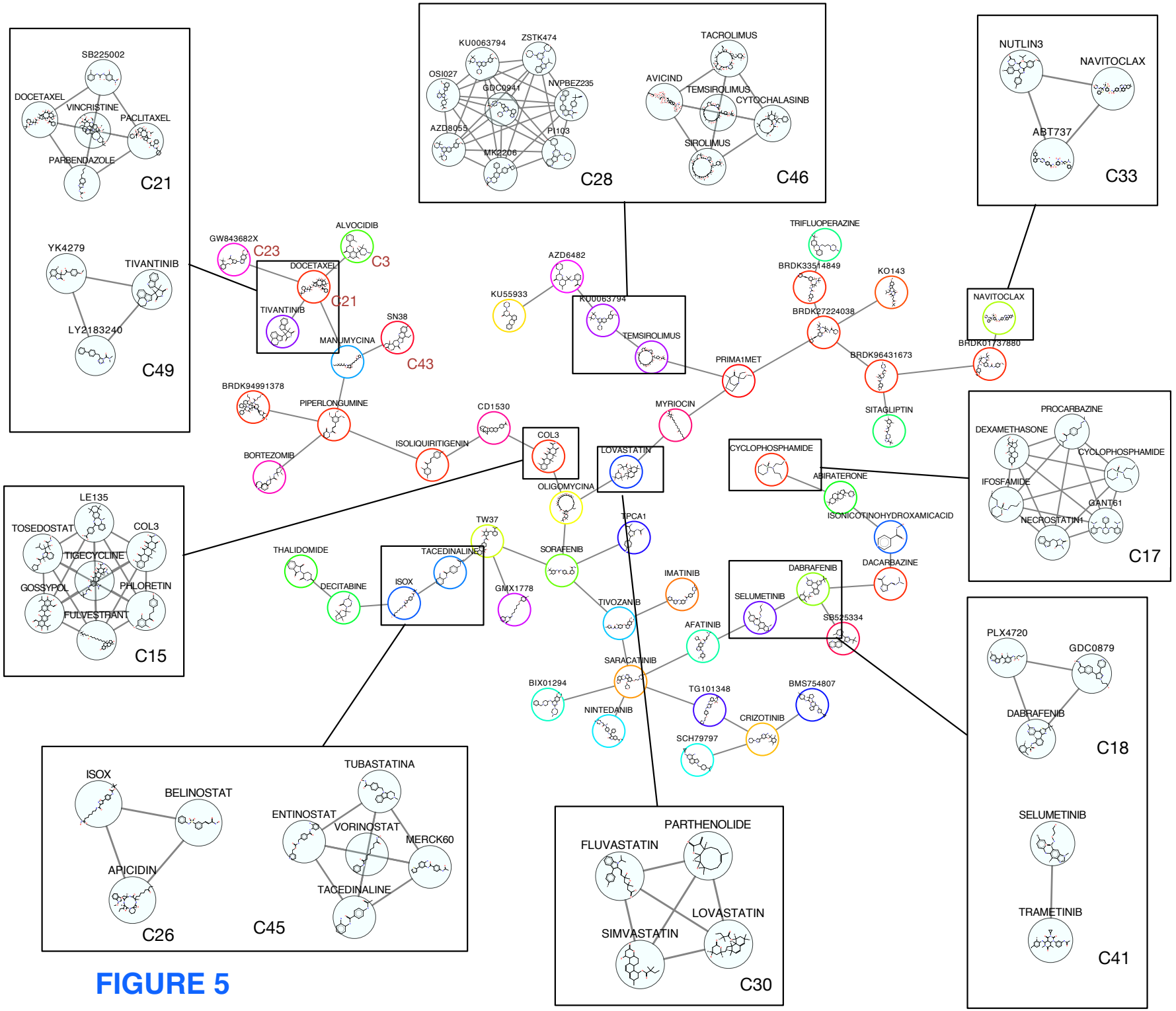


FIGURE 5

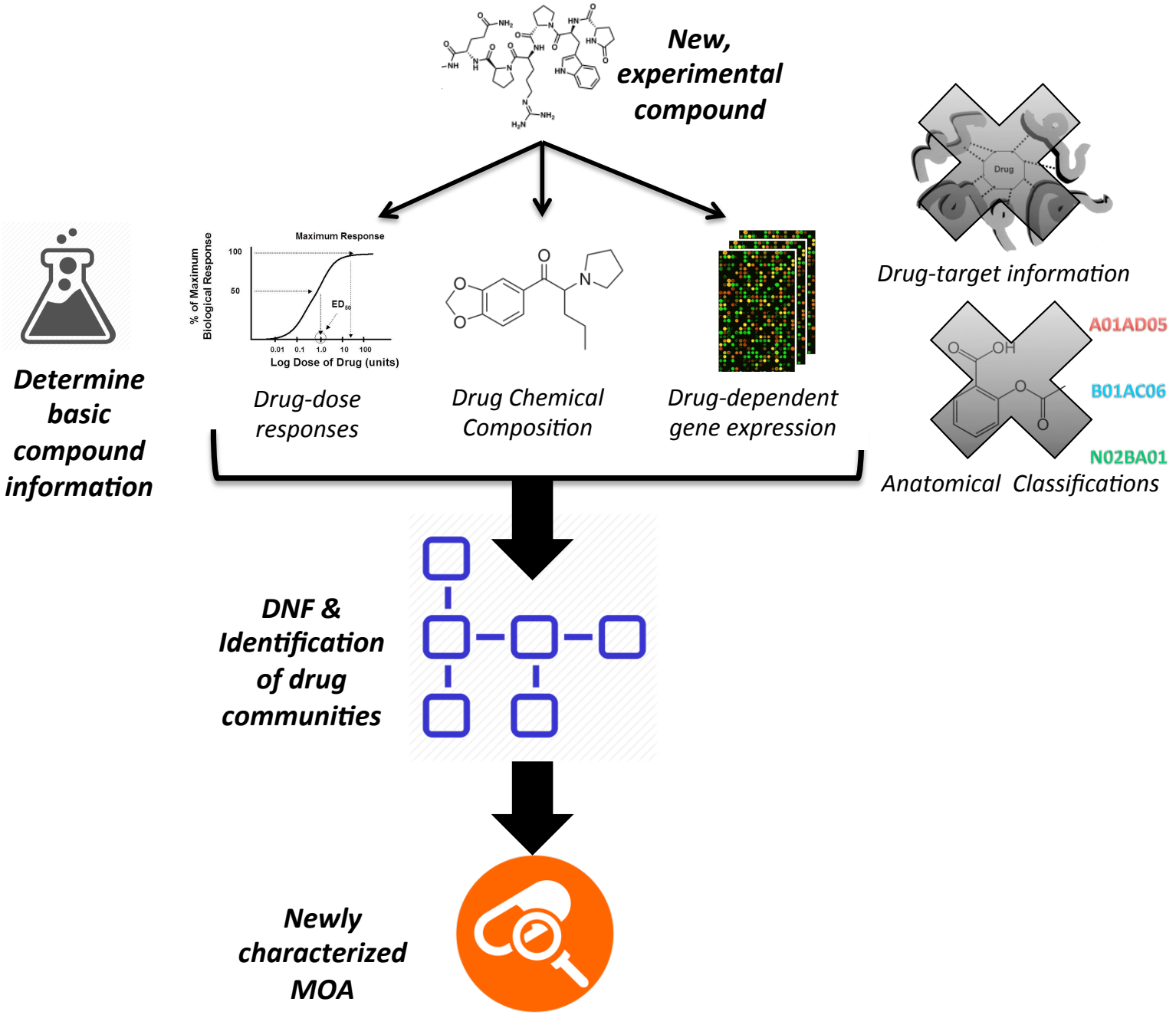


FIGURE 6

TABLE 1

CTRPv2 (targets)

Taxonomy	C-Index	pval (DNF vs. Single)
DNF	0.8872687	
structure	0.7076697	1.01E-42
sensitivity	0.8274693	3.75E-06
perturbation	0.6352601	1.25E-58

NCI60 (targets)

Taxonomy	C-Index	pval (DNF vs. Single)
DNF	0.698599	
structure	0.6077617	8.44E-08
sensitivity	0.6862717	2.35E-01
perturbation	0.489481	5.13E-25

CTRPv2 (ATCs)

Taxonomy	C-Index	pval (DNF vs. Single)
DNF	0.765503	
structure	0.7210044	2.64E-02
sensitivity	0.5751178	4.77E-12
perturbation	0.5374916	3.27E-16

NCI60 (ATCs)

Taxonomy	C-Index	pval (DNF vs. Single)
DNF	0.8700235	
structure	0.8024505	8.22E-06
sensitivity	0.7005064	2.94E-23
perturbation	0.6237462	1.11E-33

TABLE 2

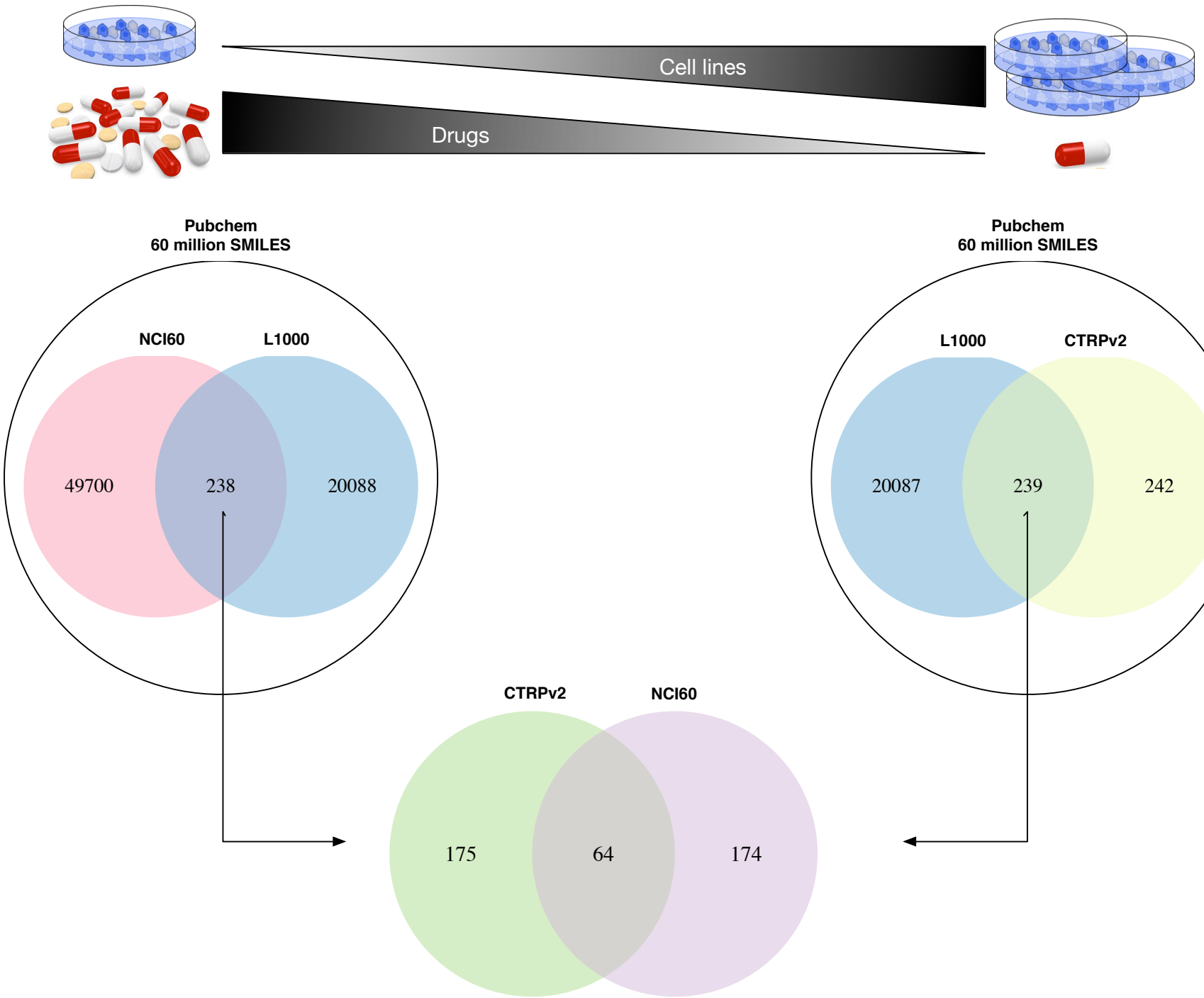
Exemplar	Target/mechanism	Community	number.of.drugs	Drugs								
ABIRATERONE	?	1	5	ABIRATERON	BRIVANIB	ERISMODEGI	EX527	TAMOXIFEN				
AFATINIB	Receptor tyrosine kinase EGFR	2	9	AFATINIB	CANERTINIB	CYANOQUIN	ERLOTINIB	GEFITINIB	IBRUTINIB	LAPATINIB	NERATINIB	WZ4002
ALVOCIDIB	CDKs	3	9	ALVOCIDIB	BMS345541	DOXORUBICI	HLI373	MYRICETIN	NARCICLASIN	PHA793887	STAUROSPOI	TRIPTOLIDE
AZD6482	PI3Ks	4	3	AZD6482	GSK1059615	TGX221						
BIX01294	EHMT2	5	4	BIX01294	BOSUTINIB	NSC23766	UNC0321					
BMS754807	IGF1R	6	4	AZD1480	BMS536924	BMS754807	LINSITINIB					
BORTEZOMIB	Proteasome	7	3	BORTEZOMIE	MG132	MLN2238						
BRDK01737880	Aurora kinase	8	4	ALISERTIB	BARASERTIB	BRDK017378	BRDK55116708					
BRDK27224038	?	9	5	BRDK139994	BRDK272240	BRDK279866	BRDK507999	VELIPARIB				
BRDK33514849	?	10	4	BRDK148442	BRDK335148	BRDK634312	SILDENAFIL					
BRDK71935468	?	11	2	BRDK719354	BRDK94991378							
BRDK88742110	?	12	2	BRDK887421	ISONICOTINO	HYDROXAMICACID						
BRDK96431673	?	13	4	BRDK024921	BRDK492906	BRDK865357	BRDK96431673					
CD1530	RAR family	14	5	AC55649	AM580	CD1530	CD437	SALERMIDE				
COL3	?	15	7	COL3	FULVESTRAN	GOSSYPOL	LE135	PHLORETIN	TIGECYCLINE	TOSEDOSTAT		
CRIZOTINIB	ALK	16	4	AT7867	CRIZOTINIB	NVPTAE684	RUXOLITINIB					
CYCLOPHOSPHAMIDE	DNA	17	6	CYCLOPHOSF	DEXAMETHA	GANT61	IFOSFAMIDE	NECROSTATI	PROCARBAZINE			
DABRAFENIB	BRAF mutant	18	3	DABRAFENIB	GDC0879	PLX4720						
DACARBAZINE	?	19	4	CIMETIDINE	DACARBAZIN	FLUOROURA	TEMOZOLOMIDE					
DECITABINE	DNMT1	20	4	AZACITIDINE	DECITABINE	QS11	ZEBULARINE					
DOCETAXEL	microtubule dynamics	21	5	DOCETAXEL	PACLITAXEL	PARBENDAZ	SB225002	VINCRISTINE				
GMX1778	NAMPT	22	3	CAY10618	GMX1778	TIPIFARNIBP2						
GW843682X	PLK1	23	4	GSK461364	GW843682X	MK1775	PRL3INHIBITORI					
IMATINIB	ABL1	24	5	AXITINIB	CHIR99021	IMATINIB	MASITINIB	NILOTINIB				
ISOLIQUIRITIGENIN	?	25	4	ISOLIQUIRITI	ITRACONAZC	PIFITHRINALI	RITA					
ISOX	HDACs	26	3	APICIDIN	BELINOSTAT	ISOX						
KO143	?	27	5	GW405833	KO143	PURMORPHA	RG108	SID26681509				
KU0063794	AKT/PI3K/mTOR axis	28	8	AZD8055	GDC0941	KU0063794	MK2206	NVPBEZ235	OSI027	PI103	ZSTK474	
KU55933	ATM	29	3	AZD7762	KU55933	KU60019						
LOVASTATIN	HMGCR	30	4	FLUVASTATIN	LOVASTATIN	PARTHENOLI	SIMVASTATIN					
MANUMYCINA	?	31	3	CCT036477	MANUMYCIN	TANESPIMYCIN						
MYRIOCIN	?	32	4	BETULINICAC	ETOMOXIR	MYRIOCIN	TRETINOIN					
NAVITOCLAX	BCL2	33	3	ABT737	NAVITOCLAX	NUTLIN3						
NINTEDANIB	?	34	3	GSK3INHIBIT	NINTEDANIB	SU11274						
OLIGOMYCINA	?	35	7	AZD7545	BREFELDINA	HYPERFORIN	OLIGOMYCIN	OUABAIN	PF750	VALDECOXIB		
PIPERLONGUMINE	Reactive oxygen species inducer	36	7	CERULENIN	CUCURBITAC	CURCUMIN	NSC632839	PIFITHRINML	PIPERLONGU	PX12		
PRIMA1MET	mutant p53	37	6	DARINAPARS	FUMONISINE	PRIMA1	PRIMA1MET	SRT1720	VER155008			
SARACATINIB	multikinases	38	3	DASATINIB	SARACATINIE	TANDUTINIB						
SB431542	TGFBR1	39	2	SB431542	SB525334							
SCH79797	?	40	4	IMPORTAZOL	METHOTREX	SCH79797	YM155					
SELUMETINIB	MEK	41	2	SELUMETINIE	TRAMETINIB							
SITAGLIPTIN	?	42	6	BLEBBISTATII	FGIN127	SGX523	SITAGLIPTIN	SJ172550	TG100115			
SN38	Topoisomerase/DNA repair,synt	43	8	CHLORAMBU	CLOFARABIN	ETOPOSIDE	GEMCITABIN	OBATOCLAX	SN38	TENIPOSIDE	TOPOTECAN	
SORAFENIB	?	44	5	BIBR1532	CI976	ERASTIN	OSI930	SORAFENIB				
TACEDINALINE	HDACs	45	5	ENTINOSTAT	MERCK60	TACEDINALIN	TUBASTATIN	VORINOSTAT				
TEMSIROLIMUS	mTOR	46	5	AVICIND	CYTOCHALAS	SIROLIMUS	TACROLIMUS	TEMSIROLIMUS				
TG101348	?	47	3	BI2536	KU0060648	TG101348						
THALIDOMIDE	?	48	3	AGK2	OLAPARIB	THALIDOMIDE						
TIVANTINIB	?	49	3	LY2183240	TIVANTINIB	YK4279						
TIVOZANIB	VEGFRs	50	8	CEDIRANIB	FORETINIB	KI8751	LINIFANIB	MGCD265	PAZOPANIB	QUIZARTINIB	TIVOZANIB	
TPCA1	?	51	3	PYRAZOLANT	SUNITINIB	TPCA1						
TRIFLUOPERAZINE	dopamine receptor antagonist	52	6	BAXCHANNE	CAY10594	PF543	PROCHLORPI	SERDEMETAI	TRIFLUOPERAZINE			
TW37	?	53	5	CICLOPIROX	MST312	NICLOSAMID	PAC1	TW37				

TABLE 3

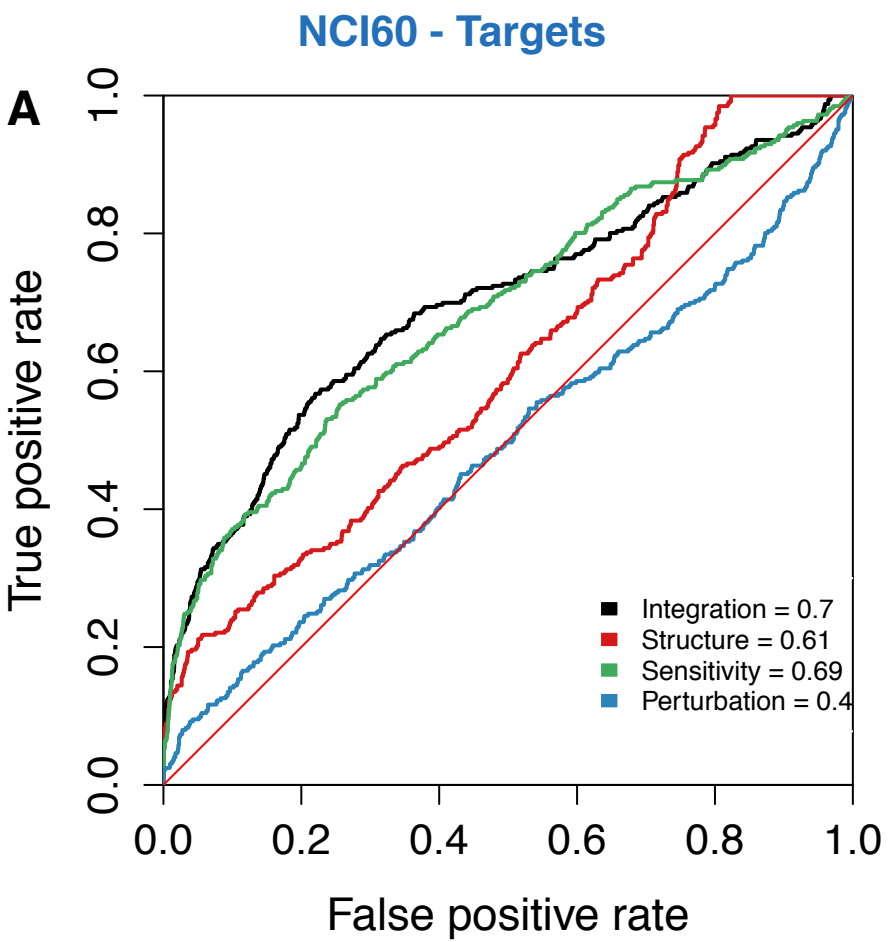
Positive control target in CTRPv2	Drugs (communities with at least 2 drugs with a known target in CTRPv2)								
	ABIRATERON	BRIVANIB	ERISMODEGI	EX527	TAMOXIFEN				
Receptor tyrosine kinase EGFR	AFATINIB	CANERTINIB	CYANOQUIN	ERLOTINIB	GEFITINIB	IBRUTINIB	LAPATINIB	NERATINIB	WZ4002
CDKs	ALVOCIDIB	BMS345541	DOXORUBICI	HLI373	MYRICETIN	NARCICLASIN	PHA793887	STAUROSPOI	TRIPTOLIDE
PI3Ks	AZD6482	GSK1059615	TGX221						
EHMT2	BIX01294	BOSUTINIB	NSC23766	UNC0321					
IGF1R	AZD1480	BMS536924	BMS754807	LINSITINIB					
Proteasome	BORTEZOMIE	MG132	MLN2238						
Aurora kinase	ALISERTIB	BARASERTIB	BRDK017378	BRDK55116708					
RAR family	AC55649	AM580	CD1530	CD437	SALERMIDE				
	COL3	FULVESTRAN	GOSSYPOL	LE135	PHLORETIN	TIGECYCLINE	TOSEDOSTAT		
ALK	AT7867	CRIZOTINIB	NVPTAE684	RUXOLITINIB					
BRAF mutant	DABRAFENIB	GDC0879	PLX4720						
DNMT1	AZACITIDINE	DECITABINE	QS11	ZEBULARINE					
NAMPT	CAY10618	GMX1778	TIPIFARNIBP2						
PLK1	GSK461364	GW843682X	MK1775	PRL3INHIBITORI					
ABL1	AXITINIB	CHIR99021	IMATINIB	MASITINIB	NILOTINIB				
	BRDK887421	ISONICOTINO	HYDROXAMIC	ACID					
HDACs	APICIDIN	BELINOSTAT	ISOX						
AKT/PI3K/mTOR axis	AZD8055	GDC0941	KU0063794	MK2206	NVPBEZ235	OSI027	PI103	ZSTK474	
ATM	AZD7762	KU55933	KU60019						
HMGCR	FLUVASTATIN	LOVASTATIN	PARTHENOLI	SIMVASTATIN					
BCL2	ABT737	NAVITOCLAX	NUTLIN3						
	GSK3INHIBIT	NINTEDANIB	SU11274						
mutant p53	DARINAPARS	FUMONISINE	PRIMA1	PRIMA1MET	SRT1720	VER155008			
multikinases	DASATINIB	SARACATINIE	TANDUTINIB						
TGFBR1	SB431542	SB525334							
MEK	SELUMETINIE	TRAMETINIB							
	BLEBBISTATII	FGIN127	SGX523	SITAGLIPTIN	SJ172550	TG100115			
Topoisomerase/DNA repair,synthesis	CHLORAMBU	CLOFARABIN	ETOPOSIDE	GEMCITABIN	OBATOCLAX	SN38	TENIPOSIDE	TOPOTECAN	
	BIBR1532	CI976	ERASTIN	OSI930	SORAFENIB				
HDACs	ENTINOSTAT	MERCK60	TACEDINALIN	TUBASTATIN	VORINOSTAT				
mTOR	AVICIND	CYTOCHALAS	SIROLIMUS	TACROLIMUS	TEMSIROLIMUS				
	BI2536	KU0060648	TG101348						
	AGK2	OLAPARIB	THALIDOMIDE						
	LY2183240	TIVANTINIB	YK4279						
VEGFRs	CEDIRANIB	FORETINIB	KI8751	LINIFANIB	MGCD265	PAZOPANIB	QUIZARTINIB	TIVOZANIB	
	PYRAZOLANT	SUNITINIB	TPCA1						
dopamine receptor antagonist	BAXCHANNE	CAY10594	PF543	PROCHLORPI	SERDEMETAI	TRIFLUOPERAZINE			
	CICLOPIROX	MST312	NICLOSAMID	PAC1	TW37				

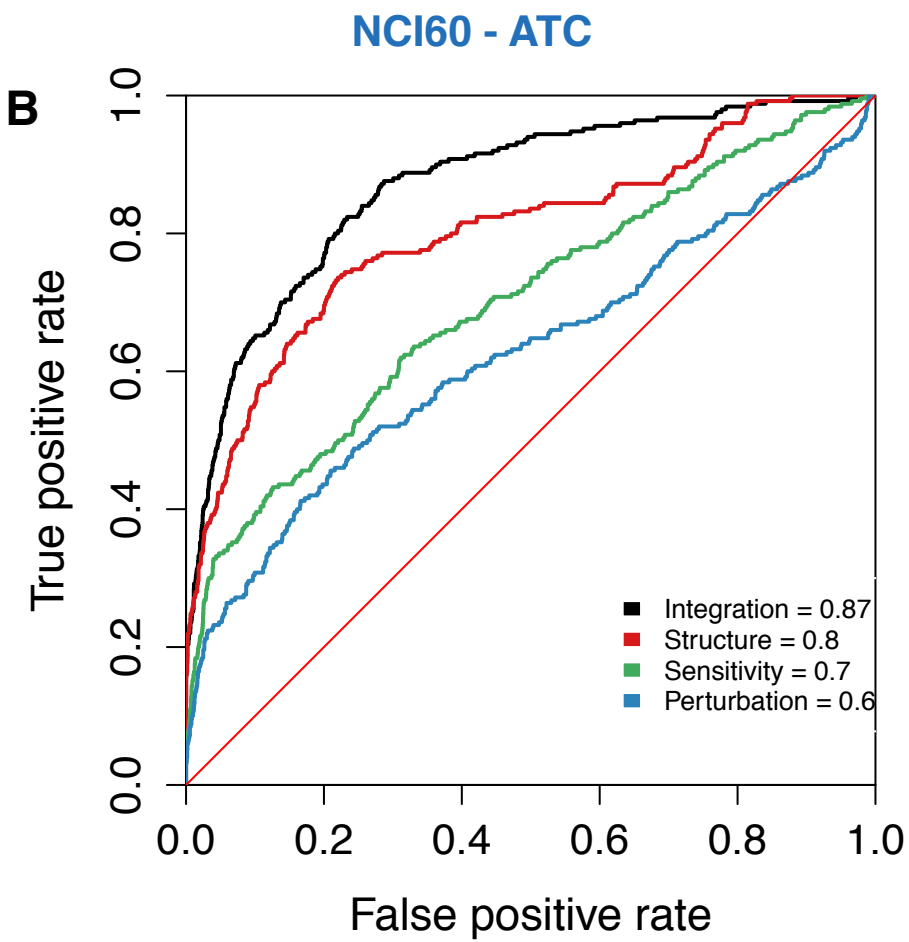
Table 4: Summary of Common Functional Drug Classes Identified Using DNF (CTRPv2 and NCI60)

- Statins (HMG-CoA reductase inhibitors)
- mTOR/PI3Ks axis inhibitors
- EGFR (epidermal growth factor receptor) inhibitors
- MEK/BRAF (Mitogen-activated protein kinase kinase) inhibitors
- BCR-ABL (fusion protein in chronic myelogenous leukemia) inhibitors
- Tubulin polymerisation/depolymerisation regulators
- DNA topoisomerase inhibitors
- Histone deacetylase inhibitors
- DNA synthesis inhibitors
- Intercalating and DNA damaging agents

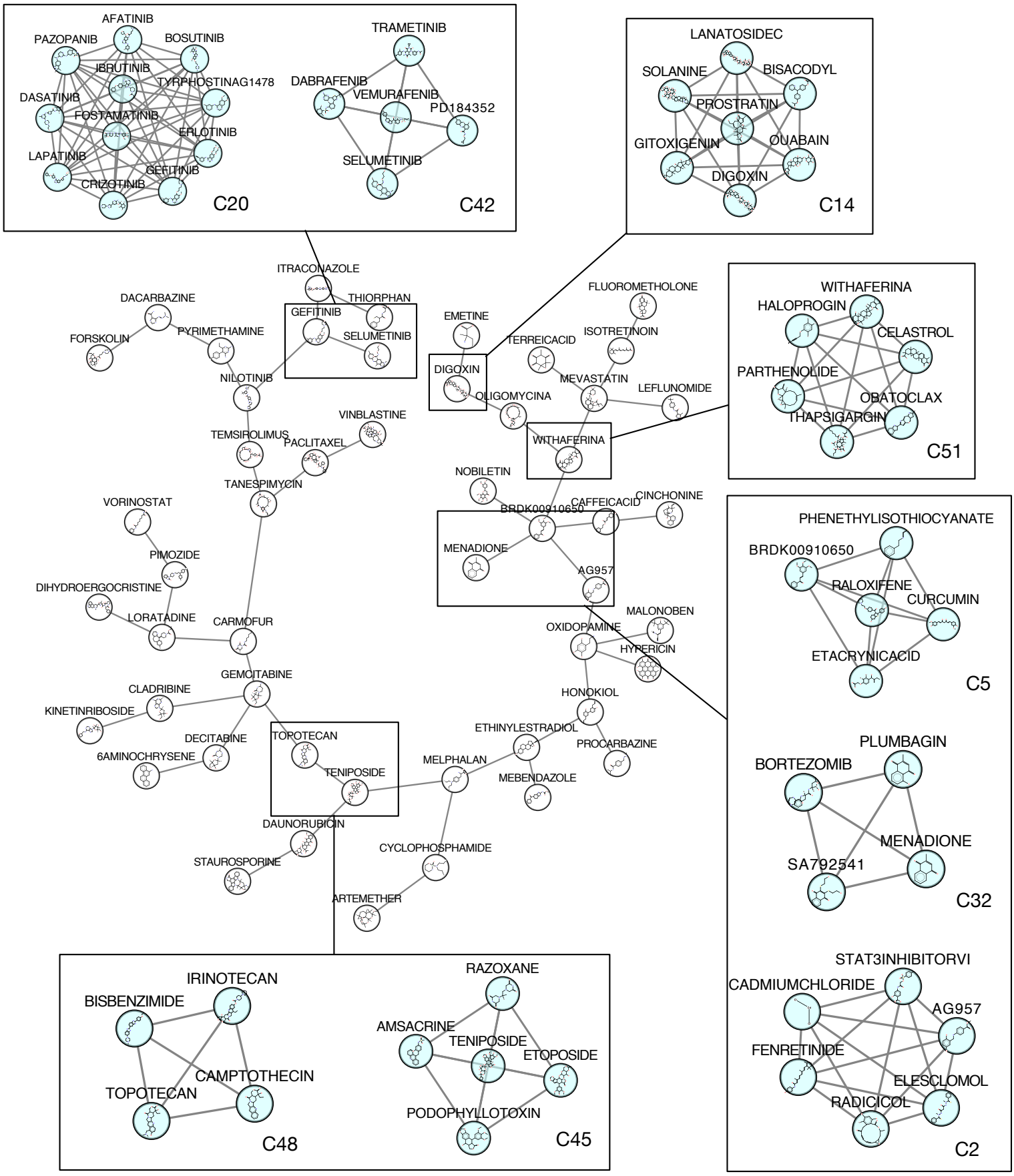


SUPPLEMENTARY FIGURE 1





SUPPLEMENTARY FIGURE 2



SUPPLEMENTARY FIGURE 3

SUPPLEMENTARY TABLE 1

Exemplar	Target/mechanism	population	number.of.d	Drugs
6AMINOCHRYSENE	?	1	4	6AMINOCHR BENZO(A)PYI CLIOQUINOL ELLIPTICINE
AG957	Oxidative stress (ROS inducers)	2	6	AG957 CADMIUMCH ELESCLOMOL FENRETINIDE RADICICOL STAT3INHIBITORVI
ALVESPIMYCIN	HSP90	3	2	ALVESPIMYC TANESPIMYCIN
ARTEMETHER	antimalarial	4	3	ARTEMETHEI ARTESUNATE YOHIMBINE
BRDK00910650	Oxidative stress (ROS inducers)	5	5	BRDK009106 CURCUMIN ETACRYNICA PHENETHYLI RALOXIFENE
CAFFEICACID	?	6	3	CAFFEICACID CAFFEICACID LAVENDUSTINA
CARMOFUR	TYMS	7	3	CARMOFUR FLUOROURA TEGAFUR
CINCHONINE	?	8	4	CINCHONINE GEDUNIN SULFAQUINC SULOCTIDIL
CLADRIBINE	DNA synthesis	9	3	CLADRIBINE CLOFARABIN FLUDARABINE
CYCLOPHOSPHAMIDE	DNA (alkylating)	10	4	BERBAMINE CYCLOPHOSF FLUCONAZOI IFOSFAMIDE
DACARBAZINE	?	11	4	ALLOPURINO DACARBAZIN MITOTANE WORTMANNIN
DAUNORUBICIN	DNA intercalation/repair	12	10	ACTINOMYCI ALVOCIDIB CHROMOMY DAUNORUBI DOXORUBICI EPIRUBICIN IDARUBICIN KENPAULLOI MITOXANTR(ROTENONE
DECITABINE	DNA methylation	13	5	AMONAFIDE AZACITIDINE DECITABINE PHENYLBUTY TUNICAMYCIN
DIGOXIN	ATPase, Na+/K+	14	7	BISACODYL DIGOXIN GITOXIGENI LANATOSIDE OUABAIN PROSTRATIN SOLANINE
DIHYDROERGOCRISTINE	?	15	4	CYCLOPIAZOI DIHYDROERG LASALOCID NELFINAVIR
EMETINE	protein synthesis inhibitors	16	4	ACETYLCYSTE CYCLOHEXIM EMETINE HOMOHARRINGTONINE
ETHINYLESTRADIOL	?	17	3	ETHINYLESTR FULVESTRAN NOSCAPINE
FLUOROMETHOLONE	Glucocorticoid receptor	18	10	DEXAMETHA FLUMETHASC FLUOCINOLO FLUOROMET FLUPHENAZII IDEBENONE ISOFLUPRED(METHYLPREI NICOTINAMI TRIAMCINOLONE
FORSKOLIN	?	19	3	FORSKOLIN PYRAZOLANT SUNITINIB
GEFITINIB	Receptor kinase/EGFR	20	11	AFATINIB BOSUTINIB CRIZOTINIB DASATINIB ERLOTINIB FOSTAMATIN GEFITINIB IBRUTINIB LAPATINIB PAZOPANIB TYRPHOSTINAG1478
GEMCITABINE	TYMS	21	5	CYTARABINE FLOXURIDINE GEMCITABIN IDOXURIDINE RALTITREXED
HONOKIOL	?	22	4	CLOTRIMAZC HONOKIOL PROBUCOL RYANODINE
HYPERICIN	?	23	4	EMODIN GOSSYPOL HYPERICIN PIPAMPERONE
ISOTRETINOIN	?	24	4	IMIQUIMOD ISOTRETINOI MIFEPRISTOI PROGESTERONE
ITRACONAZOLE	Lanosterol 14-alpha demethylas	25	3	ABIRATERON ITRACONAZC TERCONAZOLE
KINETINRIBOSIDE	?	26	4	KINETINRIBO PUROMYCIN ROSCOVITINE TRICIRIBINE
LEFLUNOMIDE	?	27	3	CELECOXIB LEFLUNOMIC NICLOSAMIDE
LORATADINE	?	28	5	AXITINIB CLOFAZIMINE LORATADINE RITONAVIR VALPROICACID
MALONOBEN	?	29	4	HEXACHLOR(MALONOBEN TAMOXIFEN TYRPHOSTINA9
MEBENDAZOLE	tubulin	30	9	CHELIDONINI FENBENDAZC FLUBENDAZC LOBENDAZOI MEBENDAZO NOCODAZOL OXFENDAZOI OXIBENDAZC PARBENDAZOLE
MELPHALAN	DNA	31	6	CHLORAMBU CISPLATIN DIBENZOYLM MELPHALAN THIOTEPA TRYPTOPHAN
MENADIONE	Oxidative stress (ROS inducers)	32	4	BORTEZOMIE MENADIONE PLUMBAGIN SA792541
MEVASTATIN	HMGCR	33	4	FLUVASTATI LOVASTATIN MEVASTATIN SIMVASTATIN
NILOTINIB	BCR-ABL fusion	34	4	IMATINIB NILOTINIB OSI027 PONATINIB
NOBILETIN	?	35	3	NOBILETIN PD98059 RHAMNETIN
OLIGOMYCINA	OxPhos blockers	36	5	AMPHOTERI(METHYLENEI NONOXYNOL OLIGOMYCIN OLIGOMYCINC
OXIDOPAMINE	?	37	4	OXIDOPAMI PENTAMIDIN PURPUROGA TERTBUTYLHYDROQUINONE
PACLITAXEL	microtubules stabilizers	38	3	BACCATINIII DOCETAXEL PACLITAXEL
PIMOZIDE	antipsychotic	39	4	ALLANTOIN CALYCANTHII PIMOZIDE SPIPERONE
PROCARBAZINE	?	40	7	HEXAMETHY MONASTROL NITAZOXANII OFLOXACIN PROCARBAZI SULFATHIAZ(TREMULACIN
PYRIMETHAMINE	DHFR	41	5	DIPYRIDAMC METHOTREX. NIFUROXAZII PEMETREXEC PYRIMETHAMINE
SELUMETINIB	BRAF/MEK	42	5	DABRAFENIB PD184352 SELUMETINI TRAMETINIB VEMURAFENIB
STAUROSPORINE	protein kinases	43	3	LESTAURTINI MIDOSTAURI STAUROSPORINE
TEMSIROLIMUS	mTOR	44	5	CYTOCHALAS EVEROLIMUS LY294002 OLAPARIB TEMSIROLIMUS
TENIPOSIDE	Topoisomerase II	45	5	AMSACRINE ETOPOSIDE PODOPHYLL(RAZOXANE TENIPOSIDE
TERREICACID	?	46	5	ASCORBICAC NBROMOACI PMSF PRIMA1 TERREICACID
THIORPHAN	?	47	4	CYCLOPAMIN HALOPERIDO NAVITOCLAX THIORPHAN
TOPOTECAN	Topoisomerase I	48	4	BISBENZIMID CAMPTOTHE IRINOTECAN TOPOTECAN
VINBLASTINE	tubulin	49	4	COLCHICINE VINBLASTINE VINCRISTINE VINORELBINE
VORINOSTAT	HDACs	50	5	BELINOSTAT ENTINOSTAT PYROXAMIDI THMI94 VORINOSTAT
WITHAFERINA	Oxidative stress (ROS inducers)	51	6	CELASTROL HALOPROGI OBATOCLAX PARTHENOLI THAPSIGARG WITHAFERINA

SUPPLEMENTARY TABLE 2

	population	number.of.di	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1	1	3	CLADRIBINE	CLOFARABIN	FLUDARABIN	NA	NA	NA	NA	NA	NA	NA	NA
2	2	4	BERBAMINE	CYCLOPHOSF	FLUCONAZOI	IFOSFAMIDE	NA	NA	NA	NA	NA	NA	NA
3	3	10	ACTINOMYCI	ALVOCIDIB	CHROMOMY	DAUNORUBI	DOXORUBICI	EPIRUBICIN	IDARUBICIN	KENPAULLOI	MITOXANTR(	ROTENONE	NA
4	4	5	AMONAFIDE	AZACITIDINE	DECITABINE	PHENYLBUTY	TUNICAMYCI	NA	NA	NA	NA	NA	NA
5	5	10	DEXAMETHA	FLUMETHAS(	FLUOCINOLO	FLUOROMET	FLUPHENAZI	IDEBENONE	ISOFLUPRED(	METHYLPRE	NICOTINAMI	TRIAMCINOL	NA
6	6	11	AFATINIB	BOSUTINIB	CRIZOTINIB	DASATINIB	ERLOTINIB	FOSTAMATIN	GEFITINIB	IBRUTINIB	LAPATINIB	PAZOPANIB	TYRPHOSTINAG1478
7	7	5	CYTARABINE	FLOXURIDINE	GEMCITABIN	IDOXURIDINE	RALTITREXEC	NA	NA	NA	NA	NA	NA
8	8	4	IMIQUIMOD	ISOTRETINOI	MIFEPRISTOI	PROGESTERC	NA	NA	NA	NA	NA	NA	NA
9	9	3	ABIRATERON	ITRACONAZC	TERCONAZOI	NA	NA	NA	NA	NA	NA	NA	NA
10	10	5	AXITINIB	CLOFAZIMIN	LORATADINE	RITONAVIR	VALPROICAC	NA	NA	NA	NA	NA	NA
11	11	6	CHLORAMBU	CISPLATIN	DIBENZOYLV	MELPHALAN	THIOTEPA	TRYPTOPHAN	NA	NA	NA	NA	NA
12	12	4	FLUVASTATI	LOVASTATIN	MEVASTATIN	SIMVASTATI	NA	NA	NA	NA	NA	NA	NA
13	13	4	IMATINIB	NILOTINIB	OSI027	PONATINIB	NA	NA	NA	NA	NA	NA	NA
14	14	3	BACCATINIII	DOCETAXEL	PACLITAXEL	NA	NA	NA	NA	NA	NA	NA	NA
15	15	5	DIPYRIDAMO	METHOTREX	NIFUROXAZI	PEMETREXEC	PYRIMETHAN	NA	NA	NA	NA	NA	NA
16	16	5	DABRAFENIB	PD184352	SELUMETINI	TRAMETINIB	VEMURAFEN	NA	NA	NA	NA	NA	NA
17	17	5	CYTOCHALAS	EVEROLIMUS	LY294002	OLAPARIB	TEMSIROLIM	NA	NA	NA	NA	NA	NA
18	18	5	AMSACRINE	ETOPOSIDE	PODOPHYLL(	RAZOXANE	TENIPOSIDE	NA	NA	NA	NA	NA	NA
19	19	4	COLCHICINE	VINBLASTINE	VINCRISTINE	VINORELBINE	NA	NA	NA	NA	NA	NA	NA
20	20	5	BELINOSTAT	ENTINOSTAT	PYROXAMIDE	THMI94	VORINOSTAT	NA	NA	NA	NA	NA	NA

SUPPLEMENTARY TABLE 3

SUPP Table 3. Summary of communities generated from CTRPv2/L1000 integrative layers (positive controls, at least 2 drugs sharing same mechanism)

Exemplar Drug*	Traget/mechanism (positive control)	Community	Number of drugs	Source of drug-target/ATC
* Communities are summerized by an exemplar drug, representative of each community. A detailed description of all 53 communities is found in table 2				
AFATINIB	Receptor tyrosine kinase EC	2	9	CTRPv2
ALVOCIDIB	CDKs	3	9	CTRPv2
AZD6482	PI3Ks	4	3	CTRPv2
BIX01294	EHMT2	5	4	CTRPv2
BMS754807	IGF1R	6	4	CTRPv2
BORTEZOMIB	Proteasome	7	3	CTRPv2
BRDK01737880	Aurora kinase	8	4	CTRPv2
CD1530	RAR family	14	5	CTRPv2
CRIZOTINIB	ALK	16	4	CTRPv2
DABRAFENIB	BRAF mutant	18	3	CTRPv2
DECITABINE	DNMT1	20	4	CTRPv2
DOCETAXEL	microtubule dynamics	21	5	CTRPv2
GMX1778	NAMPT	22	3	CTRPv2
GW843682X	PLK1	23	4	CTRPv2
IMATINIB	ABL1	24	5	CTRPv2
ISOX	HDACs	26	3	CTRPv2
KU0063794	AKT/PI3K/mTOR axis	28	8	CTRPv2
KU55933	ATM	29	3	CTRPv2
LOVASTATIN	HMGCR	30	4	CTRPv2
NAVITOCLAX	BCL2	33	3	CTRPv2
PRIMA1MET	mutant p53	37	6	CTRPv2
SARACATINIB	multikinases	38	3	CTRPv2
SB431542	TGFBR1	39	2	CTRPv2
SELUMETINIB	MEK	41	2	CTRPv2
SN38	Topoisomerase/DNA repair	43	8	CTRPv2
TACEDINALINE	HDACs	45	5	CTRPv2
TEMSIROLIMUS	mTOR	46	5	CTRPv2
TIVOZANIB	VEGFRs	50	8	CTRPv2
TRIFLUOPERAZINE	dopamine receptor antagor	52	6	CTRPv2