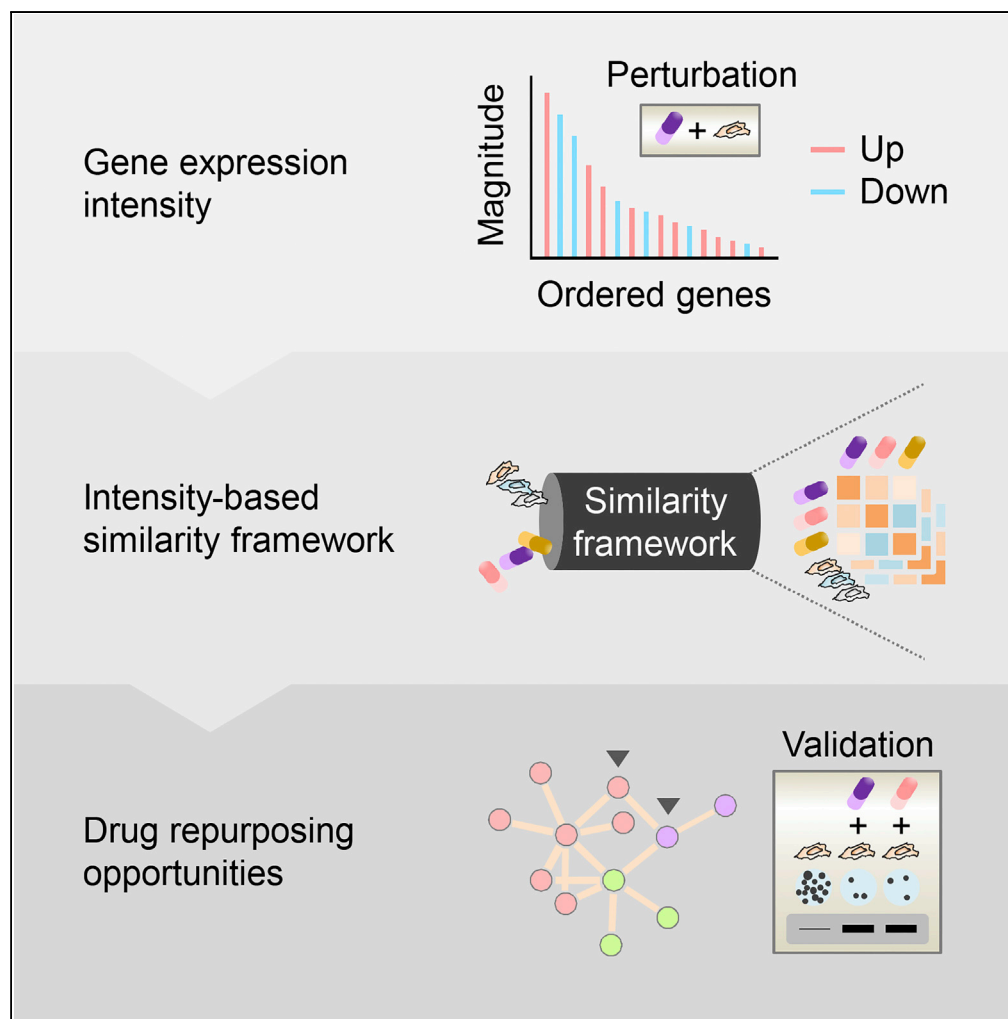**Article**

A Large-Scale Gene Expression Intensity-Based Similarity Metric for Drug Repositioning

Chen-Tsung Huang, Chiao-Hui Hsieh, Yen-Jen Oyang, Hsuan-Cheng Huang, Hsueh-Fen Juan

hsuancheng@ym.edu.tw (H.-C.H.)
yukijuan@ntu.edu.tw (H.-F.J.)

HIGHLIGHTS

Intensity-based similarity metric surpasses other standard metrics in drug clustering

This metric was applied to compare thousands of compounds for drug repurposing

Two drugs are experimentally confirmed as potential topoisomerase inhibitors

Article

A Large-Scale Gene Expression Intensity-Based Similarity Metric for Drug Repositioning

Chen-Tsung Huang,[1] Chiao-Hui Hsieh,[2] Yen-Jen Oyang,[1] Hsuan-Cheng Huang,[4,*] and Hsueh-Fen Juan[1,2,3,5,*]

SUMMARY

Biological systems often respond to a specific environmental or genetic perturbation without pervasive gene expression changes. Such robustness to perturbations, however, is not reflected on the current computational strategies that utilize gene expression similarity metrics for drug discovery and repositioning. Here we propose a new expression-intensity-based similarity metric that consistently achieved better performance than other state-of-the-art similarity metrics with respect to the gold-standard clustering of drugs with known mechanisms of action. The new metric directly emphasizes the genes exhibiting the greatest changes in expression in response to a perturbation. Using the new framework to systematically compare 3,332 chemical and 3,934 genetic perturbations across 10 cell types representing diverse cellular signatures, we identified thousands of recurrent and cell type-specific connections. We also experimentally validated two drugs identified by the analysis as potential topoisomerase inhibitors. The new framework is a valuable resource for hypothesis generation, functional testing, and drug repositioning.

INTRODUCTION

The phenomenon of drug promiscuity (coined as polypharmacology), whereby a single drug can bind to multiple targets (Hopkins, 2007, 2008; Reddy and Zhang, 2013), has revealed new therapeutic indications for marketed drugs (i.e., drug repositioning or repurposing). With recent technological advances, drug re-purposing opportunities have been actively explored through computational strategies that use various similarity measures to relate pharmaceutical and experimental compounds, such as chemical similarity (Paolini et al., 2006; Keiser et al., 2009) and side effect similarity (Campillos et al., 2008). Since polypharmacology interactions that a drug might undergo within a cell should faithfully translate into cellular phenotypic responses (Feng et al., 2009; Moffat et al., 2014), similarity approaches that compare gene expression profiles after chemical perturbations are showing considerable promise, from an initial effort to predict the mechanism of action (MoA) of a drug (Hughes et al., 2000). The Connectivity Map (CMap), as a milestone in this field, incorporates gene expression signatures and a rank-based pattern-matching strategy based on the Kolmogorov-Smirnov (KS) (Massey, 1951) statistic embedded in the Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) technique to connect small molecules and diseases (Lamb et al., 2006). CMap can be tailored to interrogate MoAs from a drug network (Iorio et al., 2010), predict new therapeutic indications using drug-disease relationships (Sirota et al., 2011), or map the effects of drugs on the immune system (Kidd et al., 2016). Although the multiplexity of gene expression profiling is seemingly useful, the mathematical and statistical properties of these high-dimensional data still remain poorly understood (Clarke et al., 2008). In high-dimensional spaces, data points are highly correlated, often with spurious relationships (Caldarelli et al., 2004), and distances of a data point to its nearest and farthest neighbors are becoming almost equidistant (Beyer et al., 1999). In addition, when a cell encounters a genetic or environmental stress, changes in gene expression are usually confined to a small number of genes, establishing the robustness of phenotypic traits (Felix and Barkoulas, 2015). However, these aspects of high dimensionality and biological robustness are mostly overlooked in state-of-the-art gene expression similarity approaches. Although CMap and its extension methods have applied data reduction techniques to a chemical perturbagen by generating a gene signature from differential expression analyses (across cell types) (Lamb et al., 2006; Sirota et al., 2011) or by extracting an arbitrary number of genes from a ranked list (merged across cell types) (Iorio et al., 2010; Kidd et al., 2016), the resulting metrics between perturbagens represent "aggregate" similarities in a way that disables cell type-specific discoveries. Moreover, the extent to which the genes selected from these methods represent a chemical perturbagen is currently unclear, owing to the lack of systematic evaluation and comparison of these similarity metrics.

[1]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan

[2]Institute of Molecular and Cellular Biology, National Taiwan University, Taipei 10617, Taiwan

[3]Department of Life Science, National Taiwan University, Taipei 10617, Taiwan

[4]Institute of Biomedical Informatics, Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei 11221, Taiwan

[5]Lead Contact

*Correspondence: hsuancheng@ym.edu.tw (H.-C.H.), yukijuan@ntu.edu.tw (H.-F.J.) https://doi.org/10.1016/j.isci.2018.08.017

Here, we present a new expression-based similarity framework that exploits the "perturbation intensity" (i.e., the difference in gene expression between perturbed and unperturbed conditions) of individual profiles, rather than aggregate information, to enable cell type-specific discoveries and to identify recurring relationships through a bottom-up approach. The new intensity-based similarity metric was designed not only to reflect the biological robustness to perturbations but also to reduce data dimensionality by simply using the genes showing the greatest changes in expression levels in response to perturbations. This framework was applied to correlate more than 7,000 chemical and genetic perturbagens across 10 different cell types obtained from the Library of Integrated Network-Based Cellular Signatures (LINCS) project (Subramanian et al., 2017). Importantly, clustering analyses using this intensity-based similarity metric produced superior results more often than those using the KS statistic (the state-of-the-art similarity metric [Lamb et al., 2006; Iorio et al., 2010; Sirota et al., 2011; Kidd et al., 2016]), GSEA method (the similarity metric defined using the standard enrichment scores in the original report [Subramanian et al., 2005]), Pearson correlation, or Euclidean distance (two commonly used similarity metrics [D'Haeseleer, 2005]) when grouping a selected set of drugs with diverse MoAs.

We further employed the intensity-based similarity metric to compare perturbation pairs with similar action restricted to one or multiple cell types and effectively recapitulated many well-known drugs as recurrently interconnected groups according to their MoAs. Novel therapeutic indications or previously unknown connectivity for marketed drugs or experimental compounds can be derived using the new metric and the results experimentally tested. For example, using our new approach, pyrvinium pamoate (an anthelmintic) and etacrynic acid (a loop diuretic) were found to connect to a cluster of topoisomerase inhibitors and to irinotecan (a topoisomerase inhibitor), respectively. The topoisomerase-inhibitory action of these two drugs was confirmed *in vitro*, indicating their potential for repurposing as a broad-spectrum (pyrvinium pamoate) or specific (etacrynic acid) anticancer drug.

Our metric represents a significant improvement over state-of-the-art and commonly used gene expression metrics to capture similar perturbations corroborated through the objective evaluation and comparison of clustering performance.

RESULTS

Workflow of the Intensity-Based Similarity Framework

We leveraged the perturbation intensity of gene expression datasets obtained from LINCS to define a tunable intensity-based similarity metric for scoring large-scale perturbation-induced changes and analyzing the similarities among perturbations (Figure 1). We considered three types of perturbations, namely, exposure to chemical drugs for 6 hr (abbreviated as d6), exposure to chemical drugs for 24 hr (d24), and exposure to short hairpin RNAs for 96 hr (sh96) in 10 selected cell types, because most experiments from LINCS were of these perturbation types and performed in these cell types. The intensity-based similarity metric relates any two given perturbations at the level of genes exhibiting the greatest changes in expression (regardless of the direction of regulation) in response to a perturbation, which is slightly different from the KS metric implemented using the same number of genes at both extremes of a ranked list (Iorio et al., 2010; Kidd et al., 2016). Using a rank-based pattern-matching strategy, our algorithm evaluates the degree of overlap between a gene set of the query perturbation and the genes at the extreme of a ranked list of perturbation intensity of the reference perturbation. The ranks of the perturbation intensity are transformed into a sequence of scores such that, with the leading term being 1, each succeeding term is obtained by multiplying its preceding term by a fixed, non-zero number (<1; as the geometric progression). For each of the two cases in which genes in the gene set match in the same or opposite direction to those in the perturbation intensity, the extent of matching can be quantified as the sum of matched entries in the geometric progression. Finally, the intensity-based similarity metric is derived by taking the difference between these two types of matching scores ("the same" minus "the opposite") over the theoretical maximum score (i.e., the geometric series; see Transparent Methods).

Two independent parameters are introduced in the intensity-based similarity metric to select the number of genes as a query gene set size (parameter b) and to govern the weights of ranks of the perturbation intensity (parameter σ, as the common ratio in the geometric progression). To determine an optimal parameter set, we used a gold-standard clustering (the ground truth) of a selected panel of chemical perturbagens with established primary MoAs manually curated from LINCS (Table S1; for the MoAs of all chemical perturbagens used in this study, see Figure S1 and Table S2). For each chemical perturbation
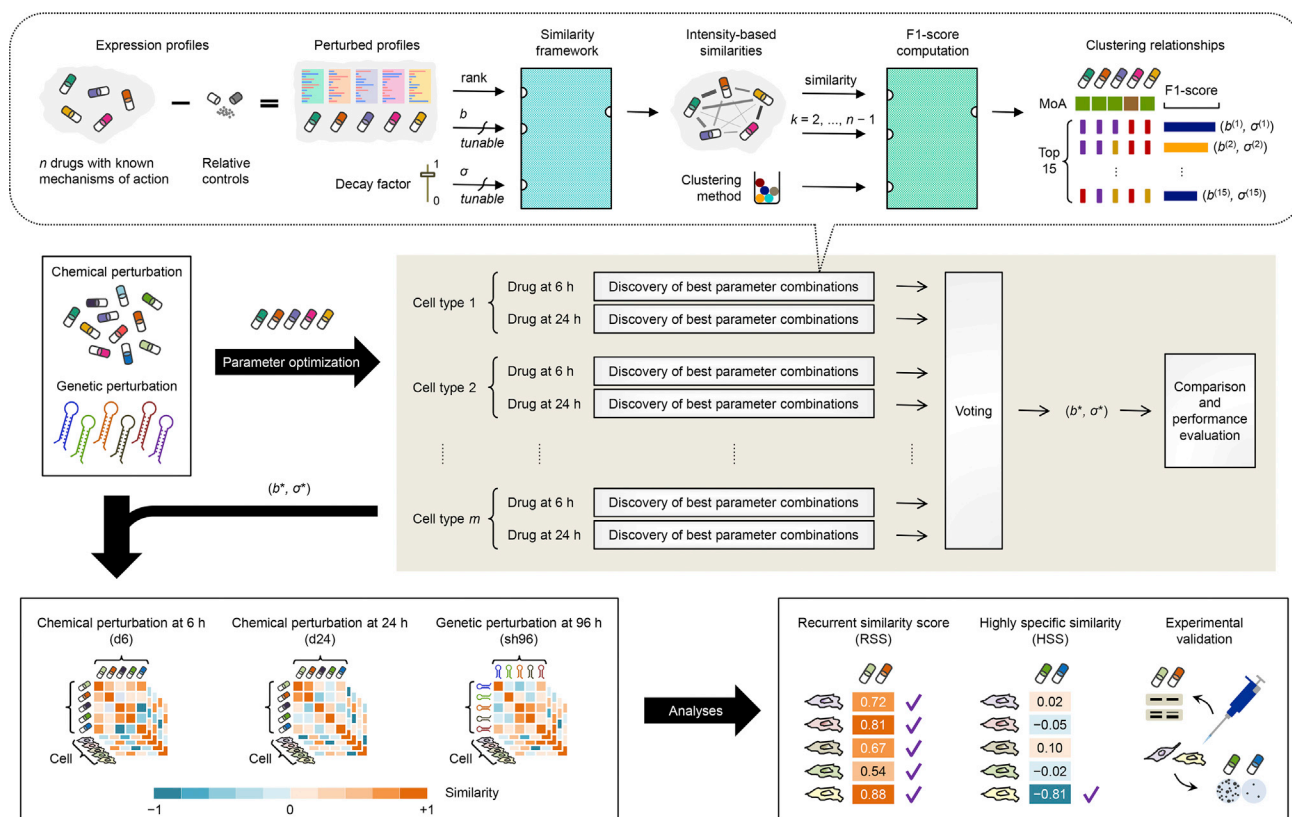
Figure 1. Workflow of the Intensity-Based Similarity Framework

We accessed gene expression profiles corresponding to thousands of chemical and genetic perturbations across multiple cell lines from LINCS. To determine optimal parameters for the intensity-based similarity metric (query gene set size b and decay factor σ), we selected 74 chemical compounds corresponding to 10 mechanisms of action (MoAs) as the gold-standard clustering and then used majority voting to choose the most frequently occurring parameter combination that yielded the best clustering performance from their treatments at 6 or 24 hr across 10 cell types. For each metric paired with one clustering method, an optimal clustering was determined across a full range of number of clusters k with the best external cluster validity index F1 score. These intensity-based metrics and resultant clusterings were further compared with other state-of-the-art and commonly used metrics. We then used the optimal parameter set to compute pairwise intensity-based similarities for all available chemical and genetic perturbations and performed in-depth analyses to uncover perturbation pairs recurrently or exclusively similar among multiple cell types. Experimental validation of discoveries was performed for drugs showing immediate repurposing opportunities.

type in a cell type, which we referred to as a "context", we assessed the clustering performance across a range of parameter sets based on the external cluster validity index F1 score (the harmonic mean between precision and recall, reaching the best value at 1 and the worst at 0) (Wiwie et al., 2015). Figure 2A shows an example clustering result for the context representing A375 cells exposed to chemical perturbations for 6 hr. Using a majority voting scheme, we selected the most frequently occurring parameter set among the top five context-specific intensity-based clusterings across all contexts as the optimal parameter set ($b^* = 100$ and $\sigma^* = 0.99$; Figure 2B). When we varied the size of the top-scoring clusterings used for voting, this optimal parameter set consistently received the most votes (Figure S2). However, we note that the parameter sets yielding the best clustering results were context specific (Figure S2A), reflecting the biological robustness to different extents across cell types (Felix and Barkoulas, 2015). The parameter sets with an emphasis on much genetic information often received the least votes ($b \geq 1,000$ and $\sigma \geq 0.998$; Figures S2A–S2D), consistent with high-dimensional noise in gene expression data (Clarke et al., 2008). Alternatively, when we counted the votes across the bottom-scoring clusterings, the parameter sets that applied far less genetic information usually received the most votes ($b = 10$ or $\sigma = 0.9$; Figure S3), indicating a requirement of a minimum number of genes for distinguishing between chemical perturbations.

Next, we compared our intensity-based similarity metric with two state-of-the-art metrics based on the KS statistic (Lamb et al., 2006; Iorio et al., 2010; Sirota et al., 2011; Kidd et al., 2016) or GSEA method
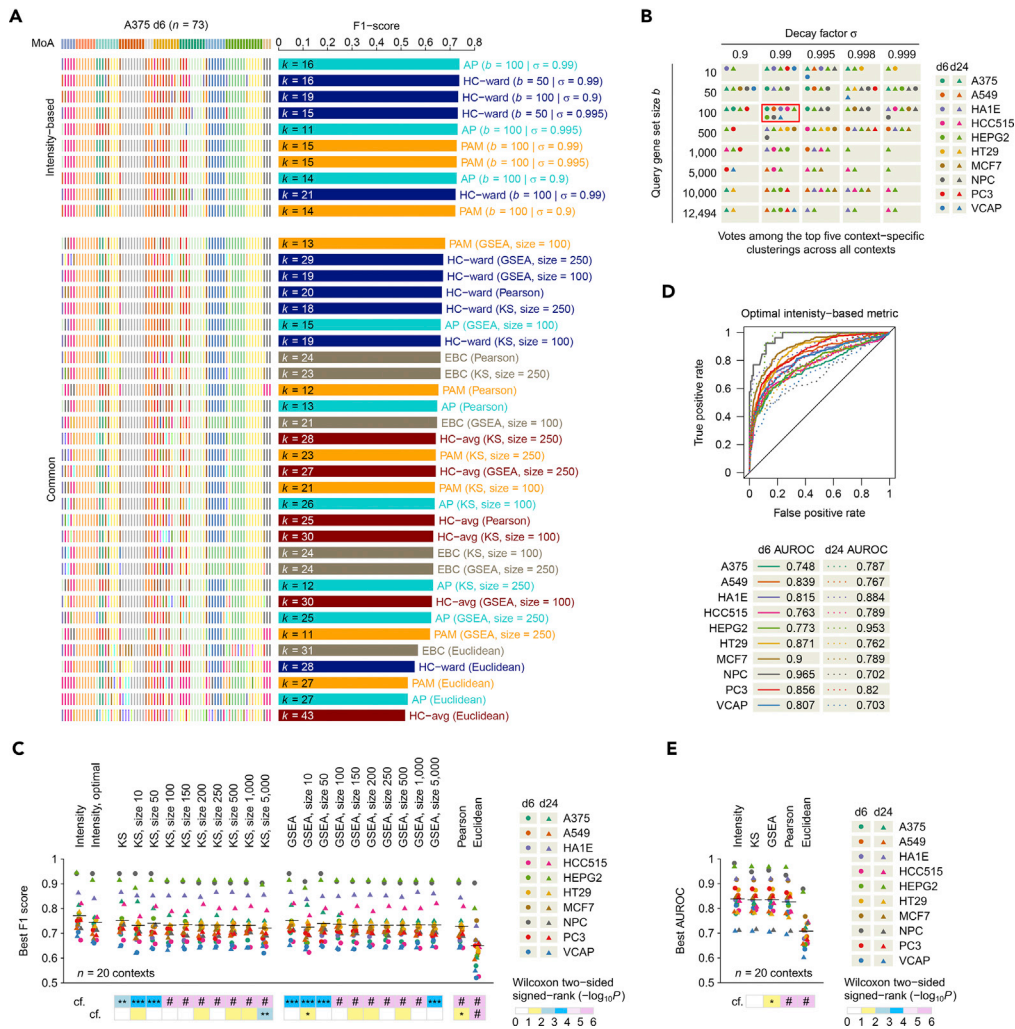
Figure 2. Parameter Optimization, Comparison, and Performance Evaluation

(A) Top 10 intensity-based clusterings along with a comparison of the best clusterings produced by similarity metrics using the KS statistic, GSEA, Pearson correlation, or Euclidean distance in the context of d6 perturbation in A375 cells. For presentation purpose, we show the results using the gene set sizes of 100 and 250 for KS and GSEA metrics. HC-avg, hierarchical clustering using average link. HC-ward, hierarchical clustering using Ward link. PAM, partitioning around medoids; AP, affinity propagation; EBC, exemplar-based agglomerative clustering.

(B) Summary of majority voting results in choosing the best parameter combination for the intensity-based similarity metric. Red box indicates the parameter set that received the most votes in this scheme.

(C) Best F1 scores for clusterings across all contexts ($n = 20$) using intensity-based metric across all parameter sets or with the optimal parameter set or using KS, GSEA, Pearson-based, or Euclidean-based metric (Wilcoxon two-sided paired signed-rank test for the intensity-based metric compared with [cf.] the other metrics).

(D) Area under the receiver operating characteristic curve (AUROC) performance using the intensity-based metric with the optimal parameter set with respect to the 74-drug gold-standard clustering for each context.

(E) Best AUROC scores across all contexts ($n = 20$) achieved by the intensity-based metric, or by the KS, GSEA, Pearson-based, or Euclidean-based metric (Wilcoxon two-sided paired signed-rank test for the intensity-based metric compared with [cf.] the other metrics).

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; #$p < 1 \times 10^{-4}$.

(Subramanian et al., 2005) and two commonly used metrics derived from Pearson correlation or Euclidean distance (D'Haeseleer, 2005) (see Transparent Methods). For KS and GSEA metrics, we considered a range of set sizes for each of the top- and bottom-ranked genes used for their matching algorithms. Across all contexts, the intensity-based metric consistently achieved higher F1 scores than the KS ($p < 0.01$, Wilcoxon two-sided paired signed-rank test, $n = 20$), GSEA ($p < 0.001$), Pearson-based ($p < 1 \times 10^{-4}$), or

Euclidean-based metric ($p < 1 \times 10^{-4}$; Figure 2C). Despite the context-specific preference for parameter sets, the optimal intensity-based metric still exhibited superior performance than both Pearson- ($p < 0.05$) and Euclidean-based metrics ($p < 1 \times 10^{-4}$) and than some of the KS or GSEA metrics obtained with a fixed gene set size ($p < 0.05$ or 0.01; Figure 2C). For each context, the top five intensity-based clusterings had significantly higher F1 scores than all other common clusterings in general (Figure S4). However, we note that all intensity-based clusterings within each context produced a broader range of F1 scores than those using the KS, GSEA, or Pearson-based metric, but they still had significantly higher F1 scores than those using the Euclidean-based metric (Figure S4). In addition to the F1 score, we also used each metric with respect to the gold-standard clustering to obtain an area under the receiver operating characteristic (ROC) curve (AUROC) as another performance measure (here, representing the ability of a similarity metric to classify drugs into clusters according to their MoAs, reaching the best value at 1 and the worst at 0, with 0.5 generated from random prediction). Of 20 context-specific maximum AUROC scores, nine were achieved by the intensity-based metric, seven by the KS metric, three by the GSEA metric, one by the Euclidean-based metric, and none by the Pearson-based metric (Figure S5). Consistent with the F1 scores, the intensity-based metric exhibited superior AUROC performance than the GSEA ($p < 0.05$, Wilcoxon two-sided paired signed-rank test, $n = 20$), Pearson-based ($p < 1 \times 10^{-4}$), and Euclidean-based metrics ($p < 1 \times 10^{-4}$) but not the KS metric ($p = 0.368$), which was slightly compromised for the optimal intensity-based metric (Figures 2D and 2E). In addition, we emphasize that the similarity values themselves produced by a given metric do not correlate with the performance of the metric. For example, among the top metrics in Figure 2A, the KS or GSEA metrics (gene set size = 100 or 250) tended to yield overall higher similarity values than the intensity-based metrics ($b = 50$ or 100 and $\sigma = 0.9$, 0.99, or 0.995; Figure S6), but achieved relatively lower F1 scores than those intensity-based metrics (Figure 2A).

We also evaluated the robustness of the intensity-based, KS, and GSEA similarity metrics to the gene set size (i.e., the query gene set size b for the intensity-based metric with a fixed decay factor σ, or the set size for each of the top- and bottom-ranked genes for the KS or GSEA metric). In general, all of these three similarity metrics were insensitive to variation in the gene set size (Figures S7 and S8).

Given that "one-step clustering" of gene expression profiles is usually unstable (Li et al., 2010), we repeated the analysis for parameter tuning and performance evaluation with resampling of 75% of chemical perturbagens. We found that the selected optimal parameter set ($b^* = 100$ and $\sigma^* = 0.99$; Figure 2B) received the most votes in 27% rounds of the resampling analysis (Figure S9A), ascertaining the general suitability of this parameter set for clustering analysis. For F1 score performance, the intensity-based metric consistently performed better than the KS or GSEA metric with $p < 0.05$ (Wilcoxon two-sided paired signed-rank test, $n = 20$) in >80% rounds of the resampling analysis and also than the Pearson- and Euclidean-based metrics with $p < 0.001$ in 100% of the analysis (Figure S9B). Consistent with F1 score performance, the intensity-based metric exhibited superior AUROC performance than the KS metric with $p < 0.05$ in ~40% of the analysis, the GSEA metric with $p < 0.05$ in >60% of the analysis, and the Pearson- and Euclidean-based metrics with $p < 1 \times 10^{-4}$ in ~100% of the analysis (Figure S9C). Notably, the intensity-based metric was able to achieve significantly higher AUROC performance than the KS metric frequently in the resampling analysis ($p < 0.05$; Figure S9C), albeit failing to reach statistical significance in the one-step, gold-standard clustering analysis ($p = 0.368$; Figure 2E). These data corroborated the ability of intensity-based metric to consistently yield better performance than the KS, GSEA, Pearson-based, or Euclidean-based metric for the clustering analysis.

To streamline the systematic analysis, we used the optimal intensity-based metric to compute all pairwise similarities for three perturbation types in 10 cell types (Table S3). For each perturbation type, positive and negative heuristic similarity cutoffs were estimated from the empirical cumulative distribution function (CDF) of perturbation similarity values across cell types, corresponding to CDF values of 0.999 and 0.001, respectively (Figure 3A). We consider these heuristic similarity cutoffs to be very conservative, because the positive heuristic cutoffs were far beyond any matched cutoff defined by Youden's index (Youden, 1950) from the ROC generated by the optimal intensity-based similarity metric for each context (Figure S10). Notably, some perturbation pairs with similarity values above the positive heuristic cutoffs were observed in up to 10 cell types, whereas most perturbation pairs with similarity values below the negative heuristic cutoffs were exclusive to one cell type (Figure 3B). This gives us an impression about the nature of the perturbations that similar responses to perturbations are often observed irrespective of the cell identity, whereas opposite effects between perturbations exhibit cell type specificity. Consistent with this
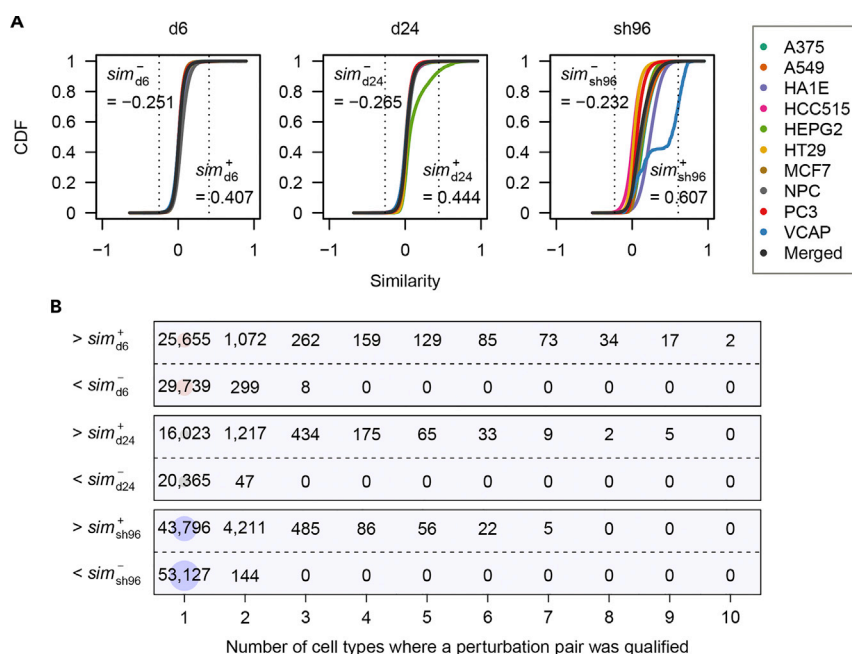
Figure 3. Distribution of Intensity-Based Similarity Measures
(A) Empirical cumulative distribution function (CDF) plots of intensity-based similarities for each perturbation type.
(B) The number of cell types in which a perturbation pair was qualified for each perturbation type.

observation, the expected small-molecule connections were recovered by intensity-based similarities when we queried the LINCS dataset using independently derived perturbation intensities of corresponding small molecules (Figure S11).

Recurrent Perturbation Pairs across Multiple Cell Types

We next sought to identify perturbation pairs that showed recurrent positive or negative similarities across many cell types and could share mechanistically related actions independent of the cell of origin. To this end, we developed a strategy that combines the intensity-based similarities and a rank-based statistical approach (Jacobsen et al., 2013) modified *ad hoc* to evaluate the recurrence of similarities across multiple cell types, generating a recurrent similarity score (RSS) for each perturbation pair considered (see Transparent Methods; Figure S12). Using these RSS scores (false discovery rate < 0.001), together with the heuristic similarity cutoffs (by considering perturbation pairs that were qualified in at least three cell types, Figure 3B), we identified 698 recurrent relationships among 203 d6 perturbagens, 399 recurrent relationships among 197 d24 perturbagens, and 575 recurrent relationships among 346 sh96 perturbagens (Figures S13–S15 and Tables S4, S5, and S6). Most of these recurrent relationships were recovered in the analysis with resampling of 60% of cell types (Figure S16). Owing to the scarcity of perturbation pairs qualified with the negative similarities in more than one cell type, no recurrent relationship with negative association was found for any perturbation type.

To explore the recurrent connections between chemical perturbagens, we combined the d6 and d24 recurrent relationships (i.e., "d6 + d24") as a recurrent drug network (Figure 4). We observed that many widely used compounds and approved drugs were recapitulated as discrete groups according to their MoAs. For example, the histone deacetylase inhibitors vorinostat, panobinostat, and dacinostat were all placed in a single group, as were the cardiac glycosides (digoxin, digitoxin, and ouabain), antimicrotubules (vinblastine, vincristine, and vinorelbine), heat shock protein 90 (HSP90) inhibitors (geldanamycin, tanespimycin, and NVP-AUY922), and lipid-lowering statins (lovastatin, simvastatin, and atorvastatin). Topoisomerase inhibitors were, nevertheless, found in two separate groups (amsacrine, etoposide, and irinotecan in one group, and doxorubicin, daunorubicin, and mitoxantrone in the other). We also note some temporal differences in the drug-drug associations. For some drug classes, recurrent connections did not appear at 6 hr until 24 hr, suggesting delayed drug responses (e.g., statins and
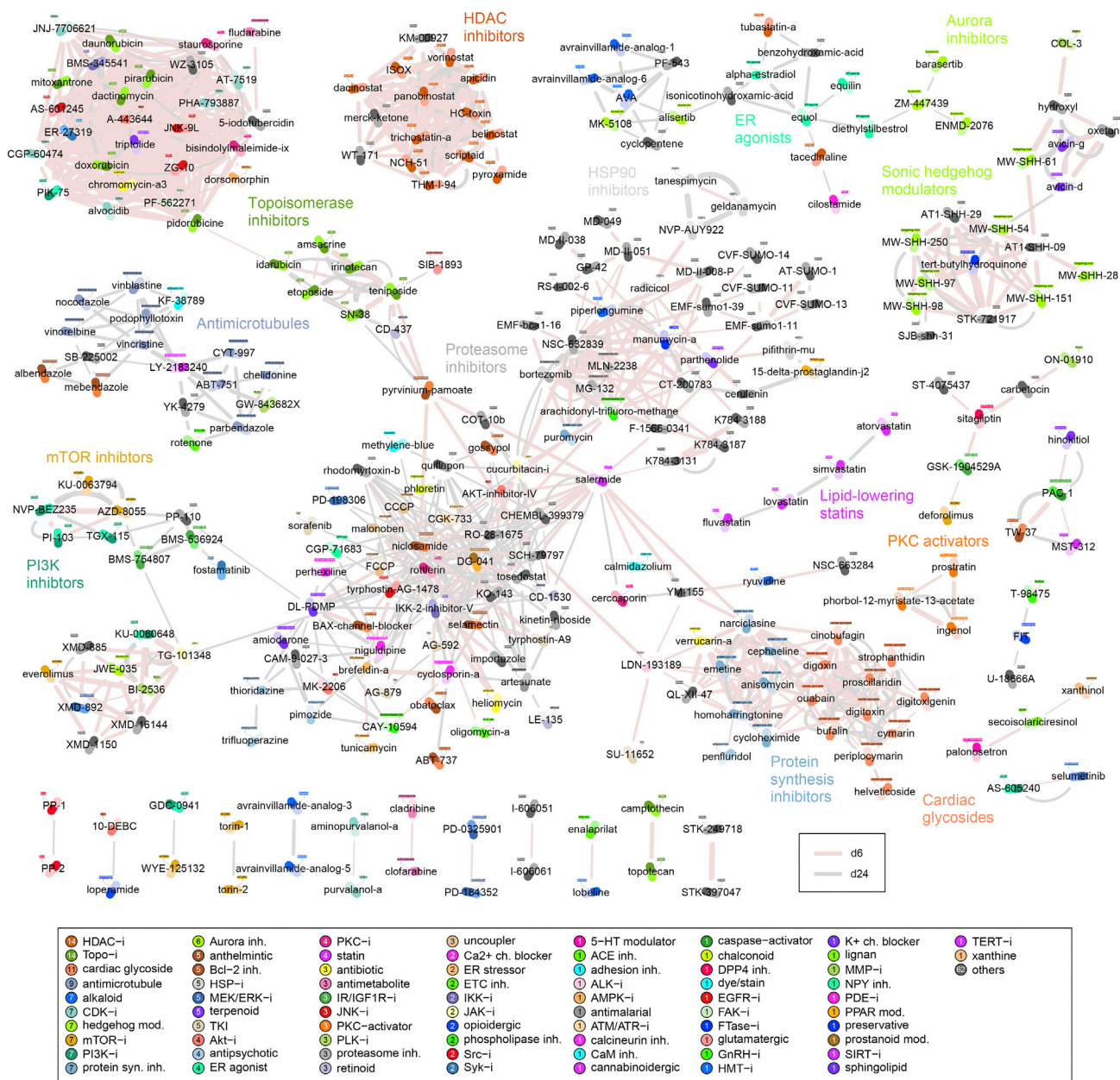
Figure 4. Recurrent Similarity Score (RSS) Analysis

Chemical perturbation pairs recurrently similar across cell types are shown as a drug RSS network (merging d6 and d24 perturbations; see Table S10 for network communities and their enrichment analyses). A primary MoA was assigned to each chemical perturbagen, as represented by color coding on the darker side of each icon and the associated text inset. The box below summarizes the total number of occurrences for each primary MoA in the network.

antimicrotubules; Figures S17A and S17B). In contrast, associations between protein synthesis inhibitors and cardiac glycosides were recurrent at 6 hr but not at 24 hr, indicative of initially similar pathways diverging later, consistent with the ability of cardiac glycosides to inhibit general protein synthesis (Perne et al., 2009) (Figure S17C). However, most connections (73%, 806 of 1,097) were between drugs of different MoAs, opening up many possibilities for drug repositioning. For example, a subnetwork of the potent protein kinase C (PKC) inhibitor staurosporine comprised substantial associations with other kinase inhibitors (e.g., cyclin-dependent kinase inhibitors, c-Jun N-terminal kinase inhibitors, topoisomerase inhibitors; Figure S17D), consistent with its well-reported polypharmacology profile (Collins and Workman, 2006; Reddy and Zhang, 2013). By contrast, pyrvinium pamoate, a US Food and Drug

Administration-approved anthelmintic, was part of a connected component overrepresented by topo-isomerase inhibitors (Figure S17E), suggesting a repurposing opportunity for cancer therapy. Function-ally related genes might also be reflected on the recurring pairs of genetic perturbations; for example, *AURKA* (aurora kinase A) knockdown was recurrently similar to knockdown of other genes involving *AURKB* (Figure S17F), consistent with their coordinated role in mitosis (Goldenson and Crispino, 2015) and reported functional similarity (Fu et al., 2009; Hans et al., 2009).

For each recurrent network (d6, d24, sh96, or d6 + d24), we applied unsupervised Markov clustering (En-right et al., 2002) to identify perturbation communities, and determined their enrichments for each of the Molecular Signature Database (Subramanian et al., 2005) gene set collections (Tables S7, S8, S9, and S10). From analyses using hallmark gene sets, several drug communities were statistically enriched for the tumor necrosis factor α-dependent nuclear factor-kappa B (NF-κB) signaling, hypoxia, p53 pathway, apoptosis, mammalian target of rapamycin complex 1 signaling, and unfolded protein response, whereas most genetic perturbation clusters impinged on the KRAS and NF-κB pathways (Figure S18; hypergeomet-ric p < 0.001). Remarkably, drug communities composed of miscellaneous MoAs also displayed functional enrichments. For example, a cluster of 27 drugs containing the PKC inhibitor staurosporine (cl2) was spe-cifically enriched for a KRAS signaling hallmark (p = 1.01 × 10^{-6}; Figure S18A). By contrast, another cluster of 31 drugs (cl1) showed multiple hallmark enrichments (Figure S18A) and might engage in cytosolic tRNA aminoacylation (p = 9.85 × 10^{-9}; Figure S19A). The drug cluster consisting not only of topoisomerase in-hibitors but also of the anthelmintic pyrvinium pamoate (cl10) was particularly enriched for p53 direct effec-tors (p = 7.03 × 10^{-16} for PID_P53_DOWNSTREAM_PATHWAY; Figure S19A). Interestingly, three genetic perturbation clusters (cl5, cl27, and cl28) revealed the enrichments for ATP-binding cassette (ABC) trans-porters that efflux various small molecules and metabolites from cells (Fletcher et al., 2010) (p = 4.34 × 10^{-64} for KEGG_ABC_TRANSPORTERS; Figure S19B), of which one cluster (cl27) contained a knockdown of *TIMP4* (TIMP metallopeptidase inhibitor 4), the overexpression of which has been shown to increase *ABCB1* and *ABCG2* transcripts (Lizarraga et al., 2016). Of the drug communities, 76% were significantly en-riched for at least one of the immunologic signatures (Figure S20; p < 0.0001), consistent with the profound influence of chemical perturbations on the immune system (Kidd et al., 2016).

Perturbation Pairs with Cell Type Specificity

We also explored the cell type specificity of perturbation similarities using an analogous definition of a metric based on information entropy (Schug et al., 2005) (see Transparent Methods). For each perturbation type (d6, d24, or sh96), we calculated the similarity specificities for all available perturbation pairs. We found that most perturbation pairs exhibited low specificity for a particular cell type (Figure S21), consistent with the nature of perturbations to induce similar effects across many cell types (Figure 3B). Using the cor-responding CDFs of the similarity specificities, we identified 1,148 highly specific similarity (HSS) pairs for d6, 204 HSS pairs for d24, and 269 HSS pairs for sh96 (Tables S11, S12, and S13; a network for d24 is shown in Figure 5). A substantial proportion of HSS pairs was associated with negative similarity scores (32%, 51%, and 83% for d6, d24, and sh96, respectively), consistent with their cell type-exclusive distributions (Fig-ure 3B). For chemical perturbagens, the HSS relationships included almost entirely different MoAs (99% and 96% for d6 and d24, respectively), suggesting that certain molecular contexts, rather than solely MoAs, dominated cellular responses and consequences following these perturbations.

Validation of Discoveries with Repurposing Potential

We performed *in vitro* experiments to validate one recurrent (from RSS) and one cell type-specific (from HSS) relationship with great promise for repurposing in oncology: pyrvinium pamoate (Thorne et al., 2010; Chen et al., 2017) and etacrynic acid (Lu et al., 2009). From the recurrent drug network, pyrvinium pa-moate was recurrently similar to a cluster of topoisomerase inhibitors, suggesting a likely antineoplastic activity against a broad spectrum of cancers. In contrast, the loop diuretic etacrynic acid was exclusively connected to the topoisomerase inhibitor irinotecan in MCF7 cells (similarity = 0.48), but not in the other cell lines screened (similarities ≈0; Table S12), indicating cell type-specific toxicity. Using proliferation and clonogenic assays, we confirmed the widespread cytotoxic effects of pyrvinium pamoate in MCF7 and A549 cells (two human cancer cell lines included in this study) as well as in H1299 and H1975 (two other human non-small-cell lung cancer cell lines not included in this study), and also confirmed the selective killing of etacrynic acid in MCF7 but not in A549 cells (Figures 6A and 6B). Western blot analysis further demon-strated that pyrvinium pamoate and etacrynic acid induced DNA damage responses, such as increased γH2AX (i.e., phosphorylated histone variant H2AX on serine 139) and p53 induction (Figure 6C), in the
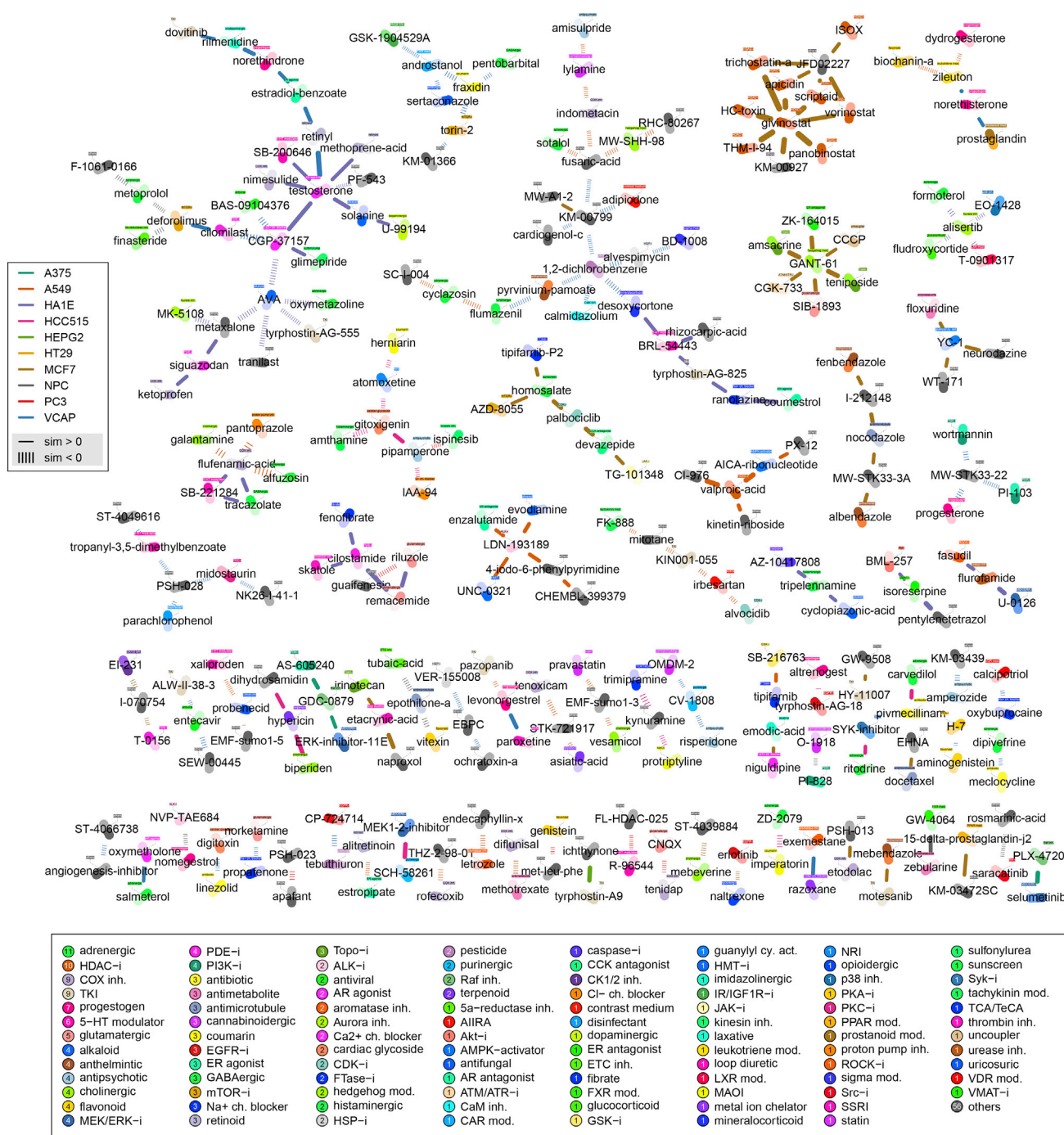
Figure 5. Highly Specific Similarity (HSS) Analysis

Chemical pairs of d24 perturbation with cell type specificity are displayed as a d24 HSS network. A primary MoA is assigned to each chemical perturbagen and color coded on the darker side of each icon and the associated text inset. The box below summarizes the total number of occurrences for each primary MoA in the network.

drug-sensitive cell lines. The levels of TOP2A (DNA topoisomerase II alpha) and its major phosphorylation site in the catalytic domain on serine 1106 (TOP2A Ser1106) (Chikamori et al., 2003) were slightly elevated by these agents to various extents (Figure 6C), consistent with trapping of TOP2A-DNA covalent complexes by known TOP2 poisons (Nitiss, 2009).
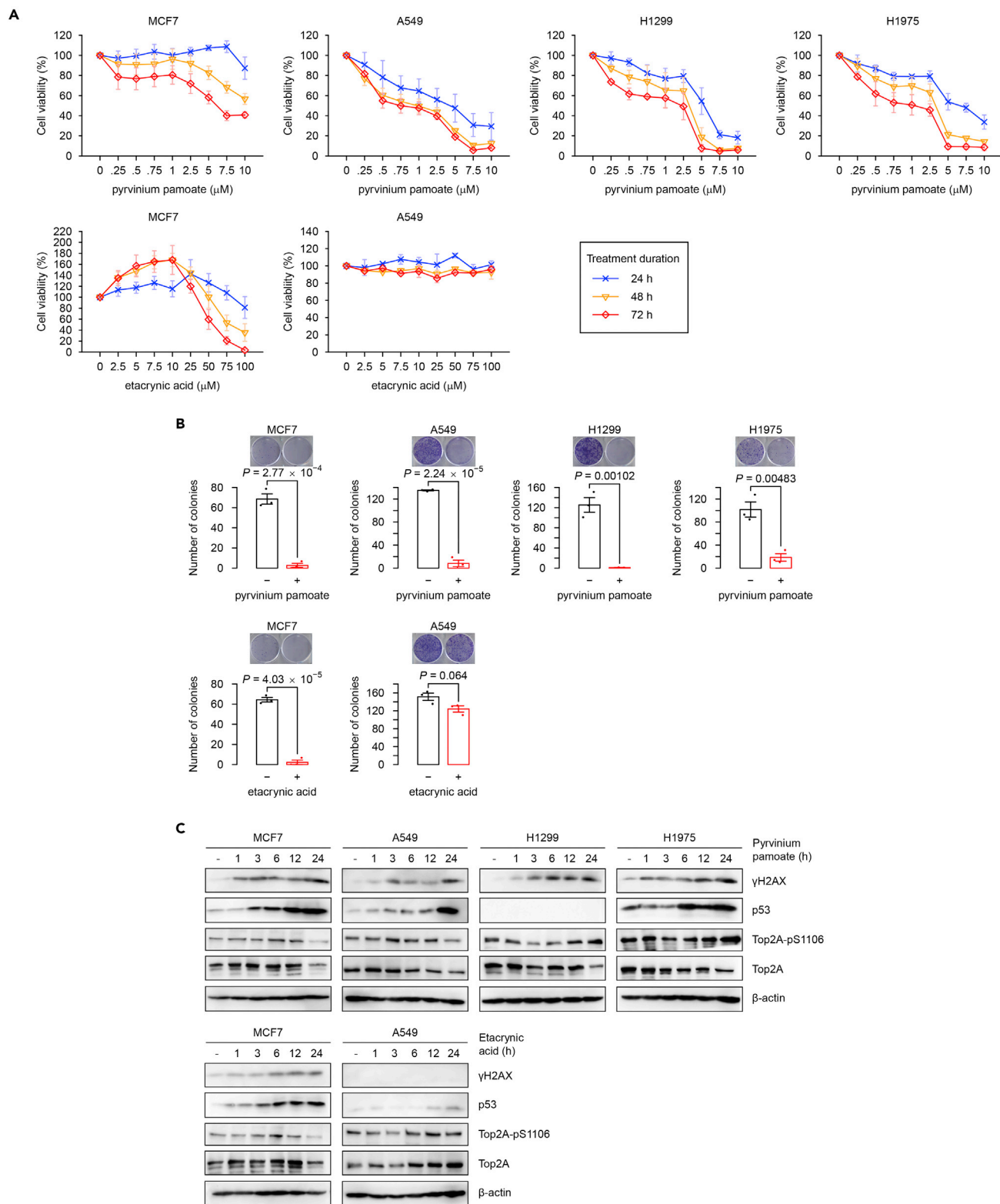
Figure 6. Experimental Validation

We verified that the US Food and Drug Administration-approved anthelmintic drug pyrvinium pamoate and the loop diuretic etacrynic acid showed RSS and HSS connections to topoisomerase inhibitors, respectively.

Figure 6. *Continued*

(A) Cell viability assays using MTS dye for pyrvinium pamoate in MCF7 and A549 cells (in this study) as well as H1299 and H1975 cells (not in this study) and etacrynic acid in MCF7 cells (effective) and A549 cells (ineffective) at indicated dosage and time duration. Error bars, SEM. Three technical replicates were performed for each of the three independent biological replicates.

(B) Clonogenic assays for pyrvinium pamoate (10 μM) or etacrynic acid (100 μM) in cell lines as indicated for 14 days. Error bars, SEM. Three technical replicates were performed for each of the three independent biological replicates (two-sided unpaired Student's t test p values).

(C) Western blot analyses revealed that pyrvinium pamoate (10 μM) and etacrynic acid (100 μM) induced DNA damage responses in drug-sensitive cell lines. H1299 cells harbor homozygous partial deletion of *TP53* and therefore do not express p53 protein. Shown are representative results from one of three independent experiments.

DISCUSSION

We developed an intensity-based similarity metric using the genes exhibiting the greatest changes in expression in response to perturbation. This new metric surpassed the state-of-the-art KS metric and other commonly used metrics in a clustering task of given known drugs, suggesting that expression intensity is a reliable proxy to represent a perturbation. In contrast to those studies that apply the KS metric to an aggregate ranked list across multiple cell types or tissues (Lamb et al., 2006; Iorio et al., 2010; Sirota et al., 2011; Kidd et al., 2016), our framework considers the intensity-based metric at the single-condition level, which enables cell type-specific discoveries and the identification of recurring relationships through a bottom-up approach. In lieu of a heuristic search from a narrow window of possible gene set sizes used for their matching algorithms (Iorio et al., 2010; Kidd et al., 2016), our framework provides a strategy using a predefined ground truth clustering to interrogate the optimal parameters for the intensity-based similarity metric (b and σ) across a broad range of parameter space.

The intensity-based similarity metric, similar to the KS and GSEA metrics, offers an advantage over the Pearson-based or Euclidean-based metric when it is only possible to use a small number of the most influential genes in a given unknown condition, even though there seems to be a minimum number of genes to adequately distinguish between chemical perturbations (≥ 50; Figure S3). Owing to the lack of a way to define a gold standard for genetic perturbations, we did not explicitly tune the intensity-based similarity metric for the genetic perturbation. Despite this, some of the known functionally related genes can still be revealed by using the optimal intensity-based similarity metric derived for the chemical perturbation.

The superior clustering performance achieved by the intensity-based similarity metric indicates that the way we created this metric is more compatible with the pervasive robustness to perturbation in biological systems (Felix and Barkoulas, 2015) and the dimensionality reduction scheme for exploring gene expression data (Clarke et al., 2008). However, compared with KS and GSEA metrics, the intensity-based similarity metric produced a broader range of F1 scores and seemed to be more sensitive to the choice of clustering method (Figure S4). This, together with the poor performance for the intensity-based similarity metric with minimal genetic information (~10 genes; Figure S3) suggests that the KS or GSEA metric might be more appropriate in the cases when only tens of genes are available for a condition.

The temporal differences observed in some drug connections pose a potential limitation to the gene expression similarity. Cellular responses to specific chemical perturbagens might only become apparent (for antimicrotubules and statins) or may be unambiguous (between cardiac glycosides and protein synthesis inhibitors) after prolonged treatment. In these cases, attempts to capture drug MoAs by using the phenotypic effects on gene expression at early time points might be frustrating. Nonetheless, this can be seen as an added benefit of the gene expression similarity for improving our knowledge of the molecular events behind chemical perturbations.

A substantial fraction of the recurrent connections spanned promiscuous small-molecule kinase inhibitors, consistent with the polypharmacology interactions for this drug family (Paolini et al., 2006; Knight et al., 2010). Enrichment analysis further indicated that such polypharmacological communities were significantly enriched for functionally distinct pathways and processes. These findings have potential clinical significance in that even some marketed drugs sharing the same MoAs were associated with disparate biological clusters, which in turn might correlate with therapeutic responses. For example, the majority of drugs that target topoisomerase II (TOP2) were found in two drug communities: one includes most anthracyclines as a promiscuous cluster, whereas the other contains the epipodophyllotoxins such as etoposide and other TOP1 inhibitors (Figure 4). This discrepancy may support the hypothesis that anthracyclines mainly act against targets other than TOP2 (Gewirtz, 1999), which is consistent with the clinical rationale for combining

etoposide and an anthracycline through its "beneficial" polypharmacological interactions (Nitiss, 2009), but probably the anthracycline-induced cardiotoxicity is a consequence of "harmful" off-target activities (Gewirtz, 1999; Nitiss, 2009). However, the extensive use of these current TOP2-targeting drugs is limited due to the increased risk of developing secondary malignancies (Nitiss, 2009); there is an urgent need to identify novel TOP2-targeting drugs with greater potency, but fewer undesirable effects, and a new way to incorporate them into clinical regimens.

Previous studies have shown that pyrvinium pamoate (Thorne et al., 2010; Chen et al., 2017) and etacrynic acid (Lu et al., 2009) exert anticancer activity by targeting the Wnt signaling pathway. Through systematically analyzing gene expression similarities, we emphasize that these drugs show their capabilities to induce DNA damage responses via topoisomerase inhibition. A careful evaluation of the polypharmacology and toxicity profiles of these compounds is required to address the opportunities for clinical development as improved TOP2-targeting agents.

Overall, we believe that our intensity-based similarity framework and the new similarity metric represent a significant step forward in the field of gene expression similarity. We have suggested a generally applicable parameter set for the intensity-based similarity metric for analyzing chemical perturbations, although it can even be fine-tuned in many relevant tasks for context-specific purpose. Moreover, data from our unbiased analysis are a useful resource for many repurposing opportunities that could potentially lead to some fast-track approvals for clinical use in the future.

Limitations of the Study

Although we have demonstrated that the intensity-based, KS, and GSEA similarity metrics are generally robust to variations in the gene set size (Figures S7 and S8), the intensity-based metric was found to achieve relatively poor performance when an insufficiently small amount of genetic information was used ($b = 10$ or $\sigma = 0.9$, among all parameter sets considered; Figure S3). This suggests that the KS or GSEA metric might be more appropriate than the intensity-based metric in cases in which only about 10 genes are allowed. Nevertheless, the use of gene expression similarity for drug repurposing relies on the cellular phenotypic responses to chemical perturbations; therefore, a potential caveat could be that some of the compound mechanisms are not able to be faithfully translated into the transcriptional changes in cells assayed at given time points.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, 21 figures, and 13 tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.08.017.

AUTHOR CONTRIBUTIONS

H.-C.H. and H.-F.J. defined the research theme and supervised the work. C.-T.H. analyzed all data. C.-H.H. performed the validation experiments. Y.-J.O. helped with data analysis. C.-T.H., H.-C.H., and H.-F.J. conceived the research, interpreted the results, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? Proc. 7th Int. Conf. Database Theory 1540, 217–235.

Caldarelli, G., Pastor-Satorras, R., and Vespignani, A. (2004). Structure of cycles and local ordering in complex networks. Eur. Phys. J. B 38, 183–186.

Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., and Bork, P. (2008). Drug target identification using side-effect similarity. Science 321, 263–266.

Chen, B., Ma, L., Paik, H., Sirota, M., Wei, W., Chua, M.S., So, S., and Butte, A.J. (2017). Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat. Commun. 8, 16022.

Chikamori, K., Grabowski, D.R., Kinter, M., Willard, B.B., Yadav, S., Aebersold, R.H., Bukowski, R.M., Hickson, I.D., Andersen, A.H., Ganapathi, R., et al. (2003). Phosphorylation of serine 1106 in the catalytic domain of topoisomerase II alpha regulates enzymatic activity and drug sensitivity. J. Biol. Chem. 278, 12696–12702.

Clarke, R., Ressom, H.W., Wang, A., Xuan, J., Liu, M.C., Gehan, E.A., and Wang, Y. (2008). The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat. Rev. Cancer 8, 37–49.

Collins, I., and Workman, P. (2006). New approaches to molecular cancer therapeutics. Nat. Chem. Biol. 2, 689–700.

D'Haeseleer, P. (2005). How does gene expression clustering work? Nat. Biotechnol. 23, 1499–1501.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Felix, M.A., and Barkoulas, M. (2015). Pervasive robustness in biological systems. Nat. Rev. Genet. 16, 483–496.

Feng, Y., Mitchison, T.J., Bender, A., Young, D.W., and Tallarico, J.A. (2009). Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat. Rev. Drug Discov. 8, 567–578.

Fletcher, J.I., Haber, M., Henderson, M.J., and Norris, M.D. (2010). ABC transporters in cancer: more than just drug efflux pumps. Nat. Rev. Cancer 10, 147–156.

Fu, J., Bian, M., Liu, J., Jiang, Q., and Zhang, C. (2009). A single amino acid change converts Aurora-A into Aurora-B-like kinase in terms of partner specificity and cellular function. Proc. Natl. Acad. Sci. USA 106, 6939–6944.

Gewirtz, D.A. (1999). A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. Biochem. Pharmacol. 57, 727–741.

Goldenson, B., and Crispino, J.D. (2015). The aurora kinases in cell cycle and leukemia. Oncogene 34, 537–545.

Hans, F., Skoufias, D.A., Dimitrov, S., and Margolis, R.L. (2009). Molecular distinctions between Aurora A and B: a single residue change transforms Aurora A into correctly localized and functional Aurora B. Mol. Biol. Cell 20, 3491–3502.

Hopkins, A.L. (2007). Network pharmacology. Nat. Biotechnol. 25, 1110–1111.

Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. 4, 682–690.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Functional discovery via a compendium of expression profiles. Cell 102, 109–126.

Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc. Natl. Acad. Sci. USA 107, 14621–14626.

Jacobsen, A., Silber, J., Harinath, G., Huse, J.T., Schultz, N., and Sander, C. (2013). Analysis of microRNA-target interactions across diverse cancer types. Nat. Struct. Mol. Biol. 20, 1325–1332.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. Nature 462, 175–181.

Kidd, B.A., Wroblewska, A., Boland, M.R., Agudo, J., Merad, M., Tatonetti, N.P., Brown, B.D., and Dudley, J.T. (2016). Mapping the effects of drugs on the immune system. Nat. Biotechnol. 34, 47–54.

Knight, Z.A., Lin, H., and Shokat, K.M. (2010). Targeting the cancer kinome through polypharmacology. Nat. Rev. Cancer 10, 130–137.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935.

Li, J., Lenferink, A.E., Deng, Y., Collins, C., Cui, Q., Purisima, E.O., O'Connor-McCourt, M.D., and Wang, E. (2010). Identification of high-quality cancer prognostic markers and metastasis network modules. Nat. Commun. 1, 34.

Lizarraga, F., Espinosa, M., Ceballos-Cancino, G., Vazquez-Santillan, K., Bahena-Ocampo, I., Schwarz-Cruz, Y.C.A., Vega-Gordillo, M., Garcia Lopez, P., Maldonado, V., and Melendez-Zajgla, J. (2016). Tissue inhibitor of metalloproteinases-4 (TIMP-4) regulates stemness in cervical cancer cells. Mol. Carcinog. 55, 1952–1961.

Lu, D., Liu, J.X., Endo, T., Zhou, H., Yao, S., Willert, K., Schmidt-Wolf, I.G., Kipps, T.J., and Carson, D.A. (2009). Ethacrynic acid exhibits selective toxicity to chronic lymphocytic leukemia cells by inhibition of the Wnt/beta-catenin pathway. PLoS One 4, e8294.

Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46, 68–78.

Moffat, J.G., Rudolph, J., and Bailey, D. (2014). Phenotypic screening in cancer drug discovery - past, present and future. Nat. Rev. Drug Discov. 13, 588–602.

Nitiss, J.L. (2009). Targeting DNA topoisomerase II in cancer chemotherapy. Nat. Rev. Cancer 9, 338–350.

Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S., and Hopkins, A.L. (2006). Global mapping of pharmacological space. Nat. Biotechnol. 24, 805–815.

Perne, A., Muellner, M.K., Steinrueck, M., Craig-Mueller, N., Mayerhofer, J., Schwarzinger, I., Sloane, M., Uras, I.Z., Hoermann, G., Nijman, S.M., et al. (2009). Cardiac glycosides induce cell death in human cells by inhibiting general protein synthesis. PLoS One 4, e8292.

Reddy, A.S., and Zhang, S. (2013). Polypharmacology: drug discovery for the future. Expert Rev. Clin. Pharmacol. 6, 41–47.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. 6, R33.

Sirota, M., Dudley, J.T., Kim, J., Chiang, A.P., Morgan, A.A., Sweet-Cordero, A., Sage, J., and Butte, A.J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci. Transl. Med. 3, 96ra77.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 171, 1437–1452.e17.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550.

Thorne, C.A., Hanson, A.J., Schneider, J., Tahinci, E., Orton, D., Cselenyi, C.S., Jernigan, K.K., Meyers, K.C., Hang, B.I., Waterson, A.G., et al. (2010). Small-molecule inhibition of Wnt signaling through activation of casein kinase 1alpha. Nat. Chem. Biol. 6, 829–836.

Wiwie, C., Baumbach, J., and Rottger, R. (2015). Comparing the performance of biomedical clustering methods. Nat. Methods 12, 1033–1038.

Youden, W.J. (1950). Index for rating diagnostic tests. Cancer 3, 32–35.

Supplemental Information

A Large-Scale Gene Expression

Intensity-Based Similarity

Metric for Drug Repositioning

Chen-Tsung Huang, Chiao-Hui Hsieh, Yen-Jen Oyang, Hsuan-Cheng Huang, and Hsueh-Fen Juan

Supplemental Figures

89 5-HT modulator	4 bisphosphonate	3 ER antagonist	2 IMPDH-i	2 PDK-i	8 sulfonylurea
1 5a-reductase inh.	1 bradykinin mod.	5 ER mixed mod.	4 interleukin mod.	4 PDK1 inh.	4 sunscreen
2 AADC inh.	8 CA inh.	3 ER stressor	8 IR/IGF1R-i	22 pesticide	3 Syk-i
6 AARI	33 Ca2+ ch. blocker	16 ergot alkaloid	12 JAK-i	6 phosphatase inh.	3 tachykinin mod.
11 ACE inh.	2 Ca2+ ch. opener	6 ET-i	6 JNK-i	1 phospholipase act.	14 TCA/TeCA
2 adenylyl cy. act.	3 calcineurin inh.	7 ETC inh.	17 K+ ch. blocker	5 phospholipase inh.	31 terpenoid
2 adenylyl cy. inh.	3 calpain inh.	2 FAK-i	14 K+ ch. opener	24 PI3K-i	4 TERT-i
3 adhesion inh.	8 CaM inh.	12 fatty acid	3 kinesin inh.	4 PKA-i	4 TGF-beta inh.
1 AdoRI	4 CaMK inh.	2 FGFR-i	5 laxative	5 PKC-activator	2 thalidomide
112 adrenergic	25 cannabinoidergic	5 fibrate	10 leukotriene mod.	10 PKC-i	7 thiazide
9 AIIRA	4 CAR mod.	41 flavonoid	3 lignan	4 PLK-i	2 thiazide-like
9 Akt-i	7 carcinogen	6 FTase-i	4 lipoxygenase inh.	20 PPAR mod.	3 thiazolidinedione
3 ALK-i	17 cardiac glycoside	14 FXR mod.	6 loop diuretic	5 preservative	2 thrombin inh.
46 alkaloid	7 caspase-activator	3 g-secretase inh.	4 LXR mod.	21 progestogen	6 thyroid
2 AMPK-activator	3 caspase-i	54 GABAergic	20 MAOI	28 prostanoid mod.	74 TKI
1 AMPK-i	3 cathepsin inh.	3 GGTase-i	7 MDM-i	4 proteasome inh.	18 Topo-i
28 anthelmintic	6 CCK antagonist	2 glinide	17 MEK/ERK-i	8 protein syn. inh.	5 Trk mod.
5 antiamoebic	23 CDK-i	37 glucocorticoid	3 melanin inh.	1 proton pump inh.	3 TRP ch. blocker
161 antibiotic	7 chalconoid	2 glucosidase inh.	9 metal ion chelator	25 purinergic	6 TRP ch. opener
26 antifungal	3 Chk-i	50 glutamatergic	9 mineralocorticoid	9 Raf inh.	3 TSPO mod.
2 antileprotic	3 cholesterol inh.	2 glycinergic	9 MMP-i	7 reducing agent	3 uncoupler
15 antimalarial	101 cholinergic	2 GnRH-i	12 mTOR-i	1 renin inh.	4 urease inh.
21 antimetabolite	6 CK1/2 inh.	8 GSK-i	49 Na+ ch. blocker	18 retinoid	4 uricosuric
19 antimicrotubule	2 Cl- ch. blocker	3 guanylyl cy. act.	4 NAMPT-i	8 ROCK-i	3 V-ATPase inh.
55 antipsychotic	3 COMT-i	2 guanylyl cy. inh.	2 NaSSA	2 SARI	5 VDR mod.
3 antitrypanosomal	11 contrast medium	1 HAT-activator	2 NDRI	8 sigma mod.	5 VEGFR-i
10 antituberculous	14 coumarin	1 HAT-i	6 NF-kB inh.	3 SIRT-activator	10 vit. B
8 antitussive	73 COX inh.	37 HDAC-i	2 NO donor	7 SIRT-i	2 vit. K
34 antiviral	3 DNA-PK inh.	14 hedgehog mod.	8 NO inh.	8 SNDRI	5 vit. K antagonist
14 AR agonist	14 DNA crosslinker	1 HIF act.	4 NPY inh.	4 SNRI	6 VMAT-i
9 AR antagonist	52 dopaminergic	1 HIF inh.	4 NRI	7 sphingolipid	2 Wnt-activator
2 AR mixed mod.	3 DPP4 inh.	68 histaminergic	25 opioidergic	5 Src-i	7 Wnt-i
7 aromatase inh.	5 DRI	6 HMT-i	14 p38 inh.	1 SRI	12 xanthine
4 ATM/ATR-i	10 dye/stain	13 HSP-i	1 p53 inh.	1 SSRE	2 XDH-i
16 Aurora inh.	32 EGFR-i	10 imidazolinergic	10 PARP-i	10 SSRI	663 others
10 Bcl-2 inh.	27 ER agonist	10 IMPase-i	27 PDE-i	9 statin	
2 biguanide			2 PDGFR-i	5 stilbenoid	

Figure S1. Frequencies of primary MoAs of LINCS compounds, related to Figure 1.

See Table S2 for detailed descriptions.

Figure S2. Votes for the best parameter set for the intensity-based metric across context-specific clusterings, related to Figure 2.

Shown are the votes among the best (A), top three (B), top five (C), top 10 (D), top 15 (E), top 20 (F), top 25 (G), or top 30 (H) context-specific intensity-based clusterings across all contexts, in which the parameter set (or sets) that received the most votes was boxed by a red rectangle.
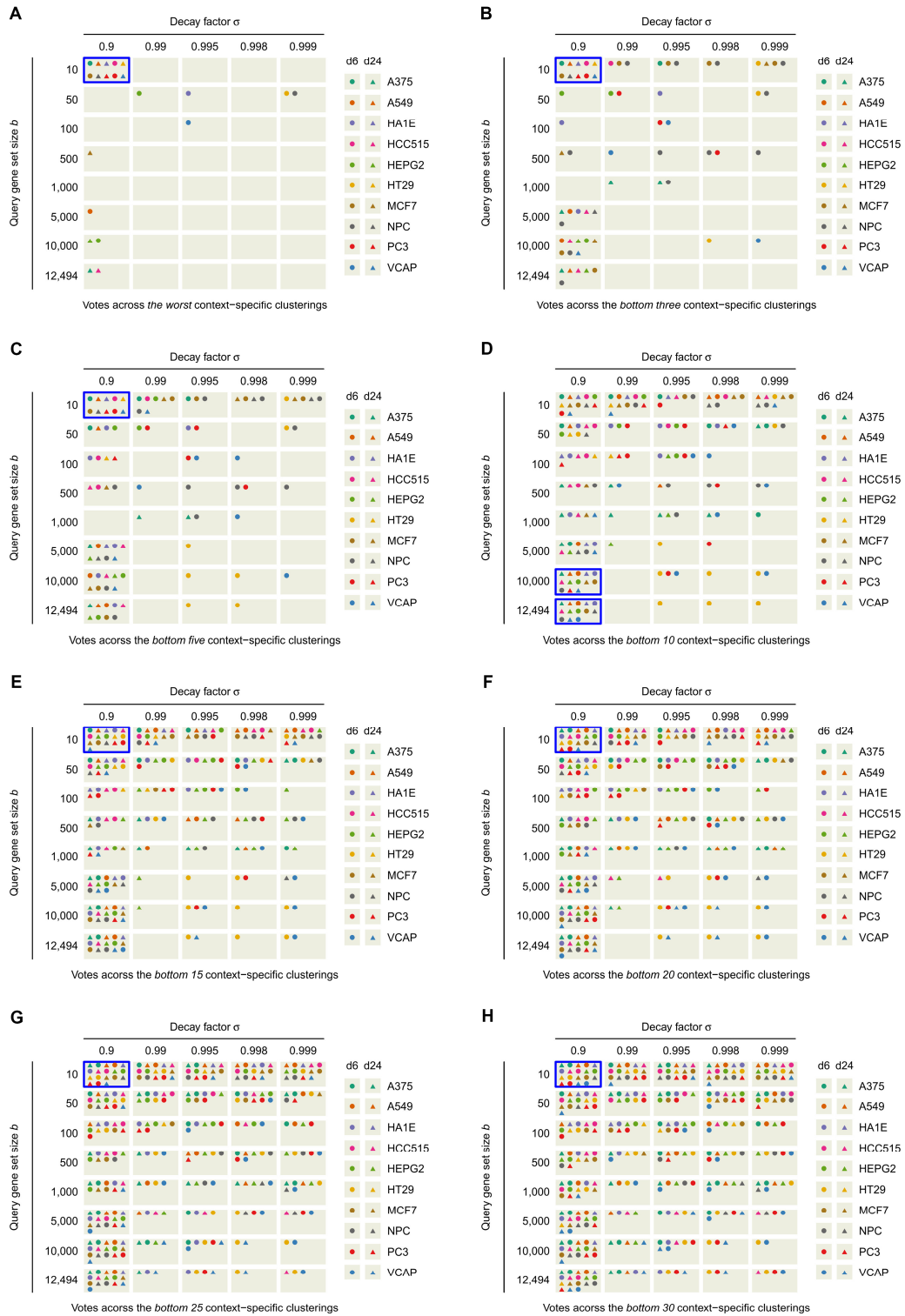
Figure S3. Votes for the worst parameter set for the intensity-based metric across context-specific clusterings, related to Figure 2.

Shown are the votes among the worst (A), bottom three (B), bottom five (C), bottom 10 (D), bottom 15 (E), bottom 20 (F), bottom 25 (G), or bottom 30 (H) context-specific intensity-based clusterings across all contexts, in which the parameter set (or sets) that received the most votes was boxed by a blue rectangle.
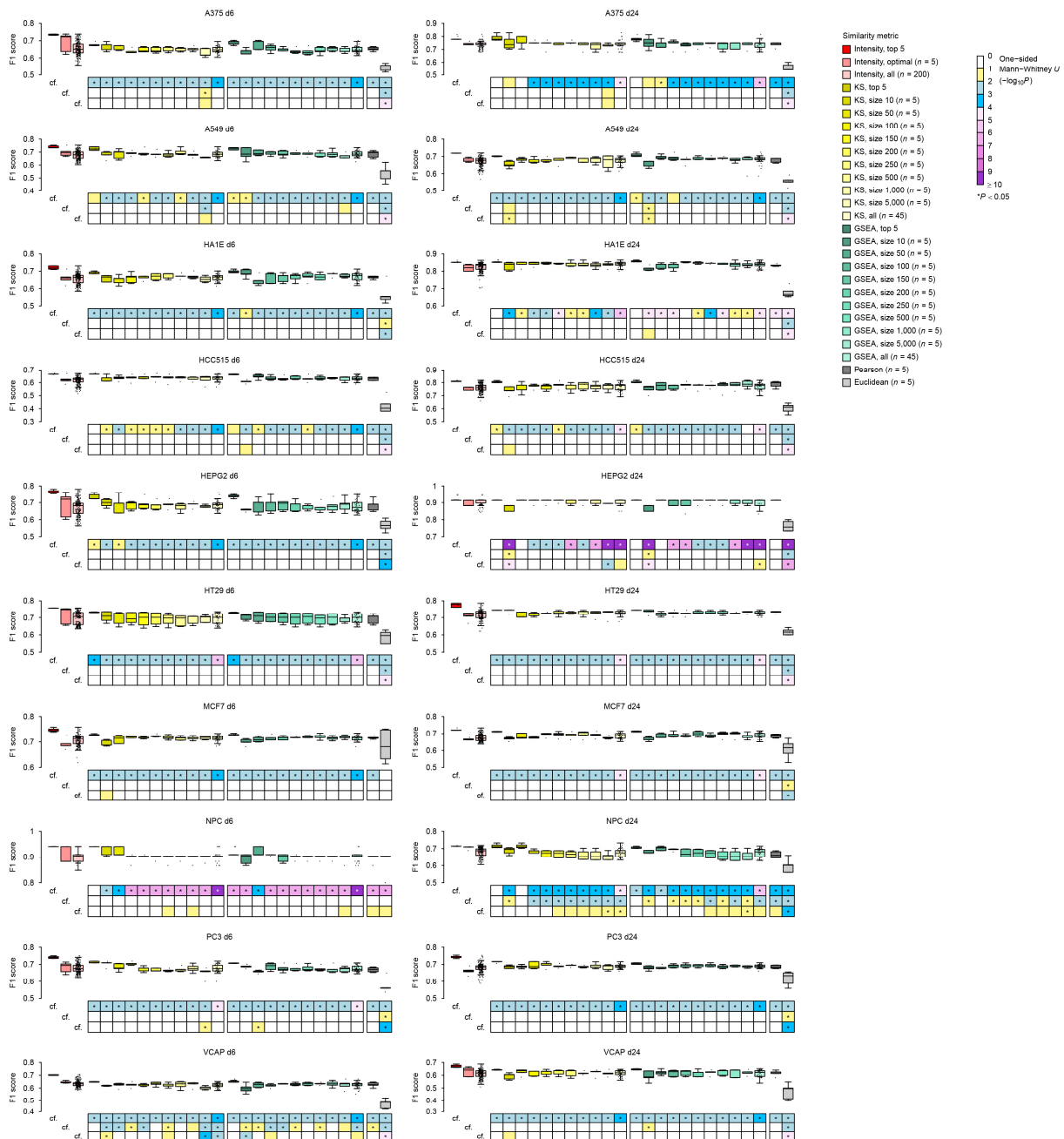
Figure S4. Comparison of F1 scores of intensity-based and common clusterings for each context, related to Figure 2.

With five biomedical clustering methods, F1 scores of intensity-based clusterings within the top five, using the optimal parameter set ($n = 5$), or across all parameter sets ($n = 200$) were compared with KS or GSEA-based clusterings within the top five, using a fixed query gene set size ($n = 5$), or across all gene set sizes considered ($n = 45$) and compared with common clusterings using the Pearson ($n = 5$) or Euclidean ($n = 5$) metric by one-sided Mann–Whitney U tests (with $-\log_{10}(P$-value) represented by the corresponding colored box at the bottom of each plot). Box plots depict the interquartile range (IQR) and whiskers represent 1.5× IQR. *$P < 0.05$.
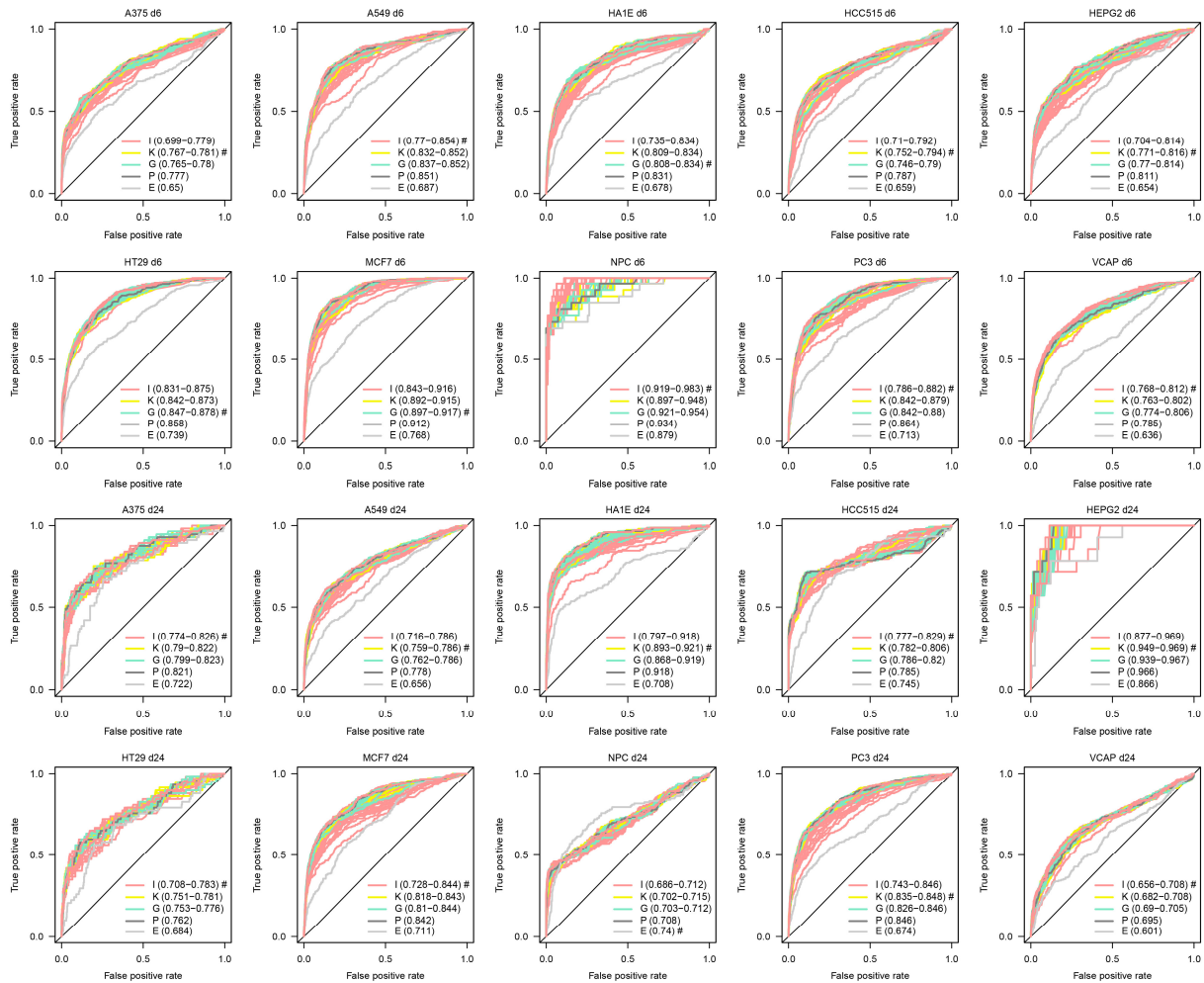
Figure S5. Comparison of AUROC performance among intensity-based and common metrics with respect to the gold-standard clustering, related to Figure 2.

For each context, an ROC curve was drawn using the intensity-based metric (I; light red lines) with all other parameter sets, KS (K; yellow lines) or GSEA (G; aquamarine lines) metric with all query gene set sizes considered, Pearson (P; dark gray line) metric, or Euclidean (E: light gray line) metric accompanied by a range of AUROC values in each parentheses, and the metric that achieved the best AUROC value was indicated by a hash (#).

A375 d6, similarity heat map
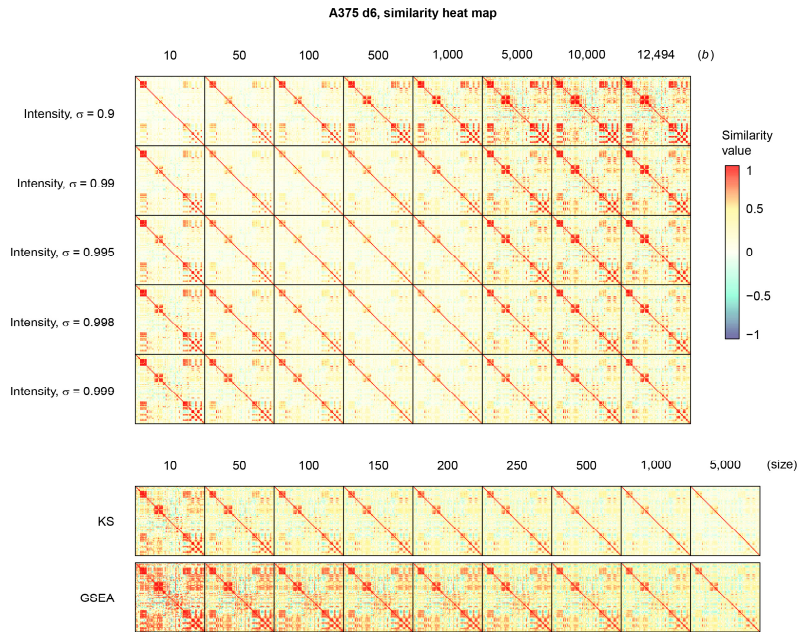
Figure S6. Similarity heat map of the gold-standard drugs for each intensity-based metric with a fixed parameter set (*b* and *σ*) and for each KS or GSEA metric with a fixed gene set size in the context of d6 perturbation in A375 cells, related to Figure 2.

For each similarity heat map, the drugs were ordered with MoAs being the same as in Figure 2A.

A375 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.652	0.667	0.722	0.707	0.681	0.649	0.643	0.639
$\sigma = 0.99$	0.703	0.699	0.723	0.648	0.648	0.667	0.667	0.662
$\sigma = 0.995$	0.707	0.699	0.712	0.645	0.632	0.642	0.635	0.642
$\sigma = 0.998$	0.683	0.691	0.687	0.642	0.642	0.637	0.648	0.636
$\sigma = 0.999$	0.678	0.672	0.687	0.642	0.642	0.656	0.647	0.648

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.677	0.672	0.638	0.642	0.642	0.642	0.656	0.648	0.618
GSEA	0.643	0.696	0.662	0.652	0.638	0.632	0.642	0.658	0.638

A375 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.714	0.75	0.75	0.708	0.714	0.7	0.737	0.7
$\sigma = 0.99$	0.737	0.737	0.737	0.737	0.737	0.737	0.75	0.75
$\sigma = 0.995$	0.737	0.737	0.737	0.737	0.737	0.737	0.75	0.75
$\sigma = 0.998$	0.737	0.737	0.737	0.737	0.737	0.737	0.75	0.75
$\sigma = 0.999$	0.737	0.737	0.737	0.737	0.737	0.737	0.75	0.75

|	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.737	0.75	0.75	0.75	0.75	0.75	0.75	0.737	0.727
GSEA	0.75	0.75	0.75	0.737	0.75	0.75	0.746	0.75	0.75

A549 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.652	0.681	0.686	0.681	0.676	0.638	0.617	0.631
$\sigma = 0.99$	0.641	0.694	0.696	0.694	0.704	0.694	0.694	0.694
$\sigma = 0.995$	0.649	0.705	0.698	0.684	0.688	0.694	0.689	0.694
$\sigma = 0.998$	0.652	0.706	0.711	0.672	0.672	0.713	0.697	0.688
$\sigma = 0.999$	0.642	0.694	0.681	0.672	0.681	0.694	0.681	0.677

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.698	0.701	0.694	0.684	0.682	0.678	0.689	0.678	0.656
GSEA	0.684	0.688	0.683	0.683	0.692	0.692	0.682	0.672	

A549 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.602	0.651	0.689	0.678	0.667	0.655	0.667	0.626
$\sigma = 0.99$	0.65	0.672	0.672	0.684	0.672	0.683	0.683	0.676
$\sigma = 0.995$	0.661	0.678	0.678	0.678	0.678	0.689	0.683	0.678
$\sigma = 0.998$	0.672	0.678	0.69	0.672	0.681	0.689	0.672	0.672
$\sigma = 0.999$	0.661	0.683	0.69	0.681	0.689	0.689	0.672	0.681

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.65	0.686	0.667	0.672	0.689	0.689	0.689	0.689	0.682
GSEA	0.672	0.691	0.689	0.683	0.683	0.683	0.689	0.681	0.693

HA1E d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.696	0.687	0.672	0.65	0.658	0.657	0.647	0.642
$\sigma = 0.99$	0.696	0.689	0.662	0.658	0.667	0.675	0.662	0.662
$\sigma = 0.995$	0.687	0.661	0.672	0.657	0.656	0.672	0.662	0.662
$\sigma = 0.998$	0.683	0.667	0.671	0.672	0.667	0.672	0.667	0.667
$\sigma = 0.999$	0.642	0.663	0.676	0.675	0.672	0.671	0.667	0.667

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.667	0.658	0.667	0.662	0.672	0.685	0.672	0.667	0.653
GSEA	0.685	0.646	0.686	0.675	0.675	0.672	0.667	0.686	0.672

HA1E d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.708	0.811	0.807	0.828	0.822	0.8	0.793	0.793
$\sigma = 0.99$	0.793	0.825	0.82	0.84	0.84	0.84	0.84	0.84
$\sigma = 0.995$	0.792	0.825	0.832	0.84	0.848	0.84	0.84	0.84
$\sigma = 0.998$	0.807	0.837	0.84	0.84	0.84	0.84	0.84	0.84
$\sigma = 0.999$	0.807	0.837	0.84	0.846	0.835	0.835	0.843	0.843

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.84	0.851	0.848	0.846	0.843	0.843	0.846	0.835	0.835
GSEA	0.816	0.84	0.84	0.84	0.854	0.846	0.846	0.84	0.84

HCC515 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.556	0.571	0.6	0.617	0.611	0.6	0.592	0.603
$\sigma = 0.99$	0.562	0.6	0.622	0.626	0.63	0.631	0.627	0.634
$\sigma = 0.995$	0.622	0.609	0.632	0.645	0.632	0.632	0.629	0.626
$\sigma = 0.998$	0.6	0.613	0.638	0.635	0.631	0.635	0.638	0.629
$\sigma = 0.999$	0.577	0.613	0.622	0.634	0.631	0.629	0.638	0.638

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.613	0.637	0.639	0.643	0.644	0.642	0.642	0.641	0.642
GSEA	0.609	0.649	0.63	0.625	0.63	0.628	0.634	0.634	0.641

HCC515 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.759	0.804	0.755	0.745	0.733	0.714	0.737	0.727
$\sigma = 0.99$	0.757	0.782	0.763	0.78	0.8	0.788	0.758	0.759
$\sigma = 0.995$	0.747	0.771	0.765	0.78	0.78	0.772	0.768	0.76
$\sigma = 0.998$	0.747	0.763	0.765	0.78	0.78	0.792	0.771	0.772
$\sigma = 0.999$	0.742	0.756	0.8	0.78	0.78	0.792	0.776	0.772

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.739	0.78	0.78	0.774	0.78	0.78	0.788	0.772	0.779
GSEA	0.766	0.779	0.779	0.78	0.78	0.78	0.787	0.787	0.788

HEPG2 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.65	0.689	0.69	0.655	0.621	0.655	0.661	0.656
$\sigma = 0.99$	0.689	0.705	0.721	0.687	0.696	0.691	0.682	0.682
$\sigma = 0.995$	0.694	0.671	0.687	0.697	0.691	0.682	0.686	0.686
$\sigma = 0.998$	0.69	0.733	0.687	0.682	0.691	0.686	0.682	0.681
$\sigma = 0.999$	0.678	0.682	0.696	0.696	0.691	0.686	0.691	0.676

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.7	0.696	0.69	0.687	0.686	0.691	0.692	0.692	0.679
GSEA	0.657	0.701	0.701	0.696	0.692	0.682	0.672	0.677	0.691

HEPG2 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.914	0.889	0.914	0.914	0.914	0.914	0.882	0.848
$\sigma = 0.99$	0.914	0.914	0.882	0.914	0.914	0.914	0.914	0.914
$\sigma = 0.995$	0.914	0.914	0.914	0.914	0.914	0.914	0.914	0.914
$\sigma = 0.998$	0.914	0.914	0.914	0.914	0.914	0.914	0.914	0.914
$\sigma = 0.999$	0.889	0.914	0.914	0.914	0.914	0.914	0.914	0.914

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.662	0.914	0.914	0.914	0.914	0.914	0.914	0.914	0.095
GSEA	0.882	0.914	0.914	0.914	0.914	0.914	0.914	0.914	0.895

HT29 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.719	0.723	0.712	0.712	0.7	0.71	0.676	0.696
$\sigma = 0.99$	0.715	0.726	0.748	0.715	0.706	0.706	0.71	0.693
$\sigma = 0.995$	0.726	0.721	0.738	0.713	0.73	0.701	0.697	0.697
$\sigma = 0.998$	0.732	0.71	0.731	0.705	0.712	0.697	0.701	0.701
$\sigma = 0.999$	0.703	0.71	0.722	0.705	0.701	0.699	0.701	0.701

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.71	0.723	0.693	0.701	0.701	0.701	0.697	0.708	0.704
GSEA	0.715	0.711	0.702	0.701	0.701	0.706	0.697	0.702	0.697

HT29 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.638	0.667	0.667	0.696	0.727	0.708	0.708	0.692
$\sigma = 0.99$	0.667	0.717	0.714	0.733	0.727	0.727	0.727	0.727
$\sigma = 0.995$	0.681	0.737	0.721	0.737	0.733	0.727	0.727	0.727
$\sigma = 0.998$	0.696	0.737	0.727	0.727	0.727	0.737	0.737	0.733
$\sigma = 0.999$	0.692	0.733	0.727	0.733	0.733	0.733	0.737	0.737

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.746	0.704	0.727	0.727	0.727	0.727	0.727	0.737	0.737
GSEA	0.737	0.727	0.727	0.727	0.727	0.727	0.727	0.727	0.737

MCF7 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.672	0.69	0.696	0.685	0.694	0.696	0.667	0.671
$\sigma = 0.99$	0.701	0.693	0.693	0.717	0.72	0.718	0.723	0.723
$\sigma = 0.995$	0.696	0.693	0.693	0.716	0.716	0.716	0.718	0.718
$\sigma = 0.998$	0.69	0.693	0.716	0.725	0.722	0.717	0.718	0.718
$\sigma = 0.999$	0.686	0.697	0.716	0.722	0.721	0.718	0.718	0.718

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.687	0.715	0.721	0.715	0.722	0.716	0.716	0.713	0.718
GSEA	0.699	0.707	0.716	0.718	0.717	0.718	0.718	0.716	0.717

MCF7 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.651	0.671	0.667	0.647	0.662	0.648	0.652	0.661
$\sigma = 0.99$	0.662	0.662	0.662	0.676	0.676	0.68	0.688	0.691
$\sigma = 0.995$	0.662	0.656	0.671	0.676	0.681	0.696	0.697	0.692
$\sigma = 0.998$	0.653	0.656	0.671	0.683	0.698	0.701	0.701	0.701
$\sigma = 0.999$	0.652	0.652	0.667	0.683	0.706	0.697	0.708	0.695

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.672	0.676	0.675	0.7	0.697	0.691	0.706	0.691	0.676
GSEA	0.662	0.685	0.681	0.685	0.699	0.681	0.702	0.701	0.688

NPC d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.882	0.909	0.909	0.875	0.909	0.875	0.875	0.909
$\sigma = 0.99$	0.875	0.909	0.882	0.909	0.909	0.909	0.909	0.909
$\sigma = 0.995$	0.875	0.941	0.941	0.909	0.909	0.909	0.903	0.903
$\sigma = 0.998$	0.875	0.941	0.909	0.909	0.903	0.903	0.903	0.903
$\sigma = 0.999$	0.875	0.909	0.909	0.903	0.903	0.903	0.903	0.903

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.941	0.909	0.903	0.903	0.903	0.903	0.903	0.903	0.903
GSEA	0.903	0.909	0.909	0.903	0.903	0.903	0.903	0.903	0.903

NPC d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.636	0.659	0.705	0.69	0.667	0.635	0.644	0.652
$\sigma = 0.99$	0.69	0.699	0.71	0.702	0.697	0.688	0.688	0.702
$\sigma = 0.995$	0.69	0.699	0.71	0.698	0.696	0.688	0.674	0.674
$\sigma = 0.998$	0.651	0.71	0.71	0.681	0.667	0.674	0.674	0.674
$\sigma = 0.999$	0.651	0.71	0.71	0.674	0.667	0.66	0.674	0.674

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.696	0.711	0.688	0.674	0.673	0.667	0.653	0.653	0.653
GSEA	0.675	0.705	0.696	0.667	0.674	0.674	0.653	0.653	0.653

PC3 d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.642	0.662	0.657	0.713	0.722	0.712	0.69	0.676
$\sigma = 0.99$	0.671	0.671	0.694	0.701	0.694	0.685	0.693	0.693
$\sigma = 0.995$	0.657	0.676	0.702	0.695	0.687	0.671	0.662	0.686
$\sigma = 0.998$	0.676	0.7	0.693	0.688	0.671	0.662	0.662	0.662
$\sigma = 0.999$	0.676	0.694	0.695	0.663	0.671	0.662	0.662	0.676

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.709	0.679	0.705	0.678	0.663	0.662	0.671	0.676	0.662
GSEA	0.69	0.658	0.693	0.672	0.676	0.671	0.671	0.671	0.676

PC3 d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.607	0.673	0.656	0.662	0.667	0.685	0.685	0.684
$\sigma = 0.99$	0.655	0.681	0.662	0.687	0.684	0.696	0.694	0.696
$\sigma = 0.995$	0.656	0.677	0.676	0.687	0.683	0.688	0.699	0.698
$\sigma = 0.998$	0.66	0.68	0.682	0.688	0.691	0.688	0.689	0.689
$\sigma = 0.999$	0.658	0.671	0.683	0.692	0.691	0.686	0.7	0.696

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.687	0.688	0.688	0.697	0.689	0.695	0.686	0.684	0.692
GSEA	0.681	0.683	0.694	0.697	0.697	0.697	0.689	0.689	0.69

VCAP d6, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.619	0.65	0.65	0.662	0.653	0.622	0.615	0.605
$\sigma = 0.99$	0.637	0.655	0.642	0.624	0.634	0.639	0.639	0.627
$\sigma = 0.995$	0.671	0.637	0.642	0.626	0.636	0.634	0.625	0.619
$\sigma = 0.998$	0.662	0.63	0.642	0.63	0.613	0.633	0.633	0.622
$\sigma = 0.999$	0.683	0.638	0.635	0.624	0.636	0.643	0.633	0.622

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.619	0.63	0.624	0.629	0.637	0.617	0.633	0.641	0.606
GSEA	0.596	0.629	0.625	0.633	0.63	0.636	0.633	0.633	0.633

VCAP d24, F1-score performance

Intensity	10	50	100	500	1,000	5,000	10,000	12,494 (b)
$\sigma = 0.9$	0.554	0.574	0.596	0.607	0.604	0.629	0.557	0.562
$\sigma = 0.99$	0.612	0.628	0.641	0.632	0.621	0.626	0.62	0.641
$\sigma = 0.995$	0.602	0.632	0.636	0.623	0.636	0.627	0.615	0.612
$\sigma = 0.998$	0.596	0.636	0.642	0.627	0.636	0.631	0.621	0.621
$\sigma = 0.999$	0.582	0.632	0.632	0.632	0.636	0.625	0.625	0.625

	10	50	100	150	200	250	500	1,000	5,000 (size)
KS	0.605	0.636	0.632	0.631	0.636	0.636	0.614	0.63	0.617
GSEA	0.587	0.62	0.628	0.632	0.623	0.623	0.636	0.641	0.626

Figure S7. Robustness analysis of F1-score performance of the intensity-based, KS, and GSEA metrics to the gene set size, related to Figure 2.

For each context, the value in each cell represents the median F1 score among the clusterings produced by the five clustering methods for a given metric (an intensity-based metric with fixed *b* and *σ*, or a KS or GSEA metric with a fixed set size).

A375 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.699	0.725	0.736	0.754	0.753	0.742	0.74	0.737	
Intensity, σ = 0.99	0.733	0.75	0.748	0.768	0.768	0.769	0.77	0.769	
Intensity, σ = 0.995	0.743	0.755	0.753	0.772	0.772	0.774	0.775	0.775	
Intensity, σ = 0.998	0.748	0.758	0.758	0.776	0.775	0.777	0.779	0.779	
Intensity, σ = 0.999	0.747	0.758	0.76	0.777	0.777	0.778	0.779	0.778	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.767	0.77	0.772	0.78	0.779	0.781	0.78	0.78	0.77	
GSEA	0.768	0.765	0.769	0.776	0.777	0.78	0.78	0.779	0.775	

A375 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.826	0.791	0.804	0.789	0.808	0.794	0.803	0.803	
Intensity, σ = 0.99	0.814	0.774	0.787	0.78	0.797	0.812	0.816	0.816	
Intensity, σ = 0.995	0.813	0.778	0.787	0.784	0.799	0.815	0.822	0.821	
Intensity, σ = 0.998	0.81	0.784	0.79	0.792	0.804	0.821	0.824	0.825	
Intensity, σ = 0.999	0.807	0.791	0.793	0.799	0.81	0.822	0.825	0.825	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.79	0.801	0.813	0.811	0.817	0.815	0.814	0.82	0.822	
GSEA	0.799	0.809	0.818	0.815	0.815	0.813	0.817	0.823	0.821	

A549 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.77	0.805	0.82	0.835	0.834	0.826	0.827	0.825	
Intensity, σ = 0.99	0.811	0.83	0.839	0.849	0.849	0.848	0.847	0.848	
Intensity, σ = 0.995	0.82	0.835	0.843	0.852	0.852	0.85	0.849	0.849	
Intensity, σ = 0.998	0.826	0.837	0.846	0.854	0.854	0.851	0.85	0.85	
Intensity, σ = 0.999	0.827	0.838	0.847	0.854	0.854	0.851	0.85	0.85	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.832	0.85	0.848	0.851	0.851	0.851	0.852	0.851	0.84	
GSEA	0.837	0.851	0.85	0.851	0.852	0.852	0.852	0.852	0.846	

A549 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.716	0.742	0.755	0.764	0.769	0.761	0.75	0.747	
Intensity, σ = 0.99	0.748	0.76	0.767	0.78	0.783	0.779	0.779	0.779	
Intensity, σ = 0.995	0.755	0.764	0.772	0.784	0.785	0.781	0.78	0.781	
Intensity, σ = 0.998	0.76	0.768	0.777	0.786	0.784	0.78	0.779	0.78	
Intensity, σ = 0.999	0.763	0.771	0.779	0.786	0.783	0.778	0.778	0.778	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.759	0.778	0.784	0.785	0.786	0.785	0.783	0.78	0.764	
GSEA	0.762	0.782	0.786	0.785	0.785	0.785	0.783	0.782	0.774	

HA1E d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.735	0.774	0.78	0.799	0.797	0.793	0.793	0.79	
Intensity, σ = 0.99	0.763	0.808	0.815	0.827	0.828	0.83	0.829	0.828	
Intensity, σ = 0.995	0.771	0.816	0.82	0.828	0.83	0.832	0.831	0.83	
Intensity, σ = 0.998	0.779	0.822	0.825	0.83	0.832	0.834	0.831	0.83	
Intensity, σ = 0.999	0.783	0.825	0.827	0.832	0.833	0.833	0.831	0.83	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.809	0.831	0.829	0.828	0.828	0.832	0.833	0.834	0.82	
GSEA	0.808	0.828	0.83	0.828	0.828	0.831	0.831	0.834	0.827	

HA1E d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.797	0.845	0.86	0.875	0.882	0.883	0.879	0.879	
Intensity, σ = 0.99	0.859	0.874	0.884	0.897	0.903	0.904	0.904	0.905	
Intensity, σ = 0.995	0.87	0.88	0.89	0.902	0.907	0.908	0.909	0.91	
Intensity, σ = 0.998	0.88	0.887	0.895	0.907	0.912	0.914	0.915	0.916	
Intensity, σ = 0.999	0.883	0.891	0.898	0.909	0.915	0.917	0.918	0.918	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.893	0.907	0.912	0.913	0.914	0.914	0.914	0.919	0.921	0.916
GSEA	0.868	0.897	0.904	0.908	0.909	0.911	0.916	0.919	0.919	

HCC515 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.71	0.735	0.728	0.733	0.729	0.724	0.72	0.718	
Intensity, σ = 0.99	0.726	0.756	0.763	0.773	0.774	0.78	0.782	0.783	
Intensity, σ = 0.995	0.728	0.758	0.767	0.778	0.781	0.787	0.789	0.79	
Intensity, σ = 0.998	0.729	0.759	0.77	0.784	0.786	0.79	0.791	0.792	
Intensity, σ = 0.999	0.728	0.759	0.773	0.786	0.788	0.789	0.79	0.79	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.752	0.784	0.784	0.788	0.794	0.794	0.792	0.786	0.78	
GSEA	0.746	0.782	0.781	0.784	0.788	0.788	0.79	0.788	0.787	

HCC515 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.829	0.829	0.822	0.804	0.794	0.785	0.785	0.785	
Intensity, σ = 0.99	0.821	0.797	0.789	0.778	0.777	0.784	0.784	0.784	
Intensity, σ = 0.995	0.815	0.789	0.783	0.777	0.777	0.785	0.785	0.785	
Intensity, σ = 0.998	0.809	0.783	0.779	0.777	0.78	0.786	0.786	0.785	
Intensity, σ = 0.999	0.806	0.779	0.778	0.779	0.782	0.787	0.785	0.785	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.806	0.793	0.79	0.79	0.791	0.791	0.791	0.789	0.782	
GSEA	0.82	0.794	0.789	0.788	0.788	0.789	0.79	0.789	0.786	

HEPG2 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.704	0.728	0.733	0.756	0.759	0.764	0.762	0.762	
Intensity, σ = 0.99	0.731	0.759	0.773	0.793	0.795	0.8	0.8	0.8	
Intensity, σ = 0.995	0.739	0.766	0.781	0.801	0.803	0.807	0.808	0.808	
Intensity, σ = 0.998	0.747	0.773	0.788	0.808	0.81	0.813	0.813	0.813	
Intensity, σ = 0.999	0.751	0.776	0.791	0.81	0.813	0.814	0.814	0.814	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.771	0.797	0.805	0.811	0.814	0.814	0.816	0.813	0.81	
GSEA	0.77	0.787	0.797	0.807	0.812	0.814	0.814	0.812	0.812	

HEPG2 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.877	0.929	0.941	0.95	0.95	0.943	0.939	0.95	
Intensity, σ = 0.99	0.928	0.966	0.953	0.955	0.957	0.953	0.954	0.955	
Intensity, σ = 0.995	0.937	0.968	0.951	0.957	0.96	0.956	0.955	0.956	
Intensity, σ = 0.998	0.94	0.969	0.95	0.959	0.96	0.963	0.963	0.962	
Intensity, σ = 0.999	0.939	0.967	0.948	0.959	0.96	0.966	0.967	0.967	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.954	0.949	0.952	0.96	0.963	0.959	0.963	0.969	0.965	
GSEA	0.947	0.941	0.939	0.949	0.953	0.955	0.959	0.965	0.967	

HT29 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.831	0.852	0.863	0.872	0.871	0.863	0.863	0.867	
Intensity, σ = 0.99	0.855	0.868	0.871	0.874	0.875	0.87	0.87	0.873	
Intensity, σ = 0.995	0.861	0.872	0.873	0.873	0.873	0.867	0.866	0.868	
Intensity, σ = 0.998	0.865	0.875	0.874	0.871	0.869	0.861	0.86	0.861	
Intensity, σ = 0.999	0.866	0.875	0.874	0.869	0.865	0.856	0.855	0.856	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.859	0.873	0.873	0.87	0.867	0.863	0.859	0.854	0.842	
GSEA	0.862	0.878	0.875	0.871	0.87	0.868	0.863	0.858	0.847	

HT29 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.708	0.751	0.747	0.751	0.749	0.754	0.761	0.767	
Intensity, σ = 0.99	0.745	0.761	0.762	0.769	0.767	0.779	0.778	0.783	
Intensity, σ = 0.995	0.748	0.763	0.763	0.768	0.768	0.777	0.775	0.777	
Intensity, σ = 0.998	0.747	0.768	0.766	0.767	0.766	0.77	0.769	0.77	
Intensity, σ = 0.999	0.745	0.77	0.769	0.767	0.764	0.764	0.762	0.763	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.751	0.779	0.781	0.772	0.771	0.764	0.765	0.756	0.754	
GSEA	0.753	0.775	0.776	0.773	0.771	0.771	0.767	0.762	0.754	

MCF7 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.843	0.873	0.884	0.9	0.905	0.897	0.887	0.888	
Intensity, σ = 0.99	0.876	0.895	0.9	0.914	0.916	0.915	0.914	0.914	
Intensity, σ = 0.995	0.883	0.902	0.906	0.916	0.916	0.915	0.914	0.914	
Intensity, σ = 0.998	0.889	0.908	0.912	0.916	0.915	0.913	0.912	0.912	
Intensity, σ = 0.999	0.89	0.911	0.914	0.916	0.914	0.911	0.911	0.911	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.892	0.914	0.915	0.915	0.915	0.914	0.911	0.91	0.906	
GSEA	0.897	0.915	0.916	0.916	0.917	0.916	0.913	0.912	0.908	

MCF7 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.728	0.757	0.772	0.811	0.819	0.835	0.836	0.833	
Intensity, σ = 0.99	0.769	0.779	0.789	0.818	0.826	0.841	0.842	0.842	
Intensity, σ = 0.995	0.786	0.791	0.799	0.823	0.83	0.842	0.844	0.844	
Intensity, σ = 0.998	0.805	0.805	0.812	0.83	0.835	0.843	0.844	0.844	
Intensity, σ = 0.999	0.817	0.814	0.821	0.834	0.838	0.843	0.844	0.844	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.818	0.831	0.837	0.84	0.841	0.843	0.842	0.842	0.838	
GSEA	0.81	0.83	0.837	0.839	0.84	0.843	0.843	0.844	0.844	

NPC d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.919	0.965	0.961	0.978	0.983	0.98	0.973	0.972	
Intensity, σ = 0.99	0.955	0.973	0.965	0.971	0.966	0.951	0.948	0.948	
Intensity, σ = 0.995	0.96	0.976	0.967	0.965	0.96	0.942	0.941	0.942	
Intensity, σ = 0.998	0.965	0.976	0.967	0.958	0.949	0.933	0.932	0.935	
Intensity, σ = 0.999	0.969	0.973	0.964	0.953	0.944	0.927	0.928	0.928	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.946	0.948	0.938	0.936	0.932	0.931	0.929	0.925	0.897	
GSEA	0.928	0.954	0.944	0.945	0.945	0.942	0.939	0.932	0.921	

NPC d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.7	0.694	0.693	0.7	0.701	0.695	0.689	0.686	
Intensity, σ = 0.99	0.693	0.696	0.702	0.711	0.71	0.711	0.71	0.71	
Intensity, σ = 0.995	0.695	0.699	0.705	0.711	0.71	0.71	0.709	0.71	
Intensity, σ = 0.998	0.696	0.703	0.71	0.71	0.708	0.709	0.709	0.71	
Intensity, σ = 0.999	0.697	0.706	0.712	0.709	0.708	0.708	0.709	0.71	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.709	0.715	0.713	0.711	0.71	0.709	0.708	0.706	0.702	
GSEA	0.703	0.712	0.709	0.708	0.71	0.71	0.71	0.709	0.706	

PC3 d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.786	0.809	0.811	0.835	0.835	0.841	0.837	0.833	
Intensity, σ = 0.99	0.805	0.848	0.856	0.879	0.878	0.876	0.873	0.871	
Intensity, σ = 0.995	0.808	0.858	0.865	0.882	0.88	0.874	0.872	0.87	
Intensity, σ = 0.998	0.811	0.866	0.871	0.884	0.877	0.868	0.866	0.865	
Intensity, σ = 0.999	0.814	0.869	0.874	0.878	0.873	0.863	0.861	0.861	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.842	0.877	0.879	0.875	0.872	0.87	0.865	0.859	0.842	
GSEA	0.842	0.873	0.88	0.875	0.875	0.873	0.871	0.866	0.852	

PC3 d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.743	0.774	0.79	0.81	0.819	0.82	0.821	0.818	
Intensity, σ = 0.99	0.786	0.816	0.82	0.838	0.842	0.844	0.842	0.841	
Intensity, σ = 0.995	0.796	0.824	0.825	0.839	0.844	0.845	0.842	0.841	
Intensity, σ = 0.998	0.804	0.832	0.832	0.843	0.845	0.845	0.843	0.842	
Intensity, σ = 0.999	0.808	0.836	0.836	0.844	0.846	0.846	0.844	0.843	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.835	0.843	0.846	0.846	0.848	0.848	0.846	0.846	0.838	
GSEA	0.826	0.842	0.843	0.844	0.846	0.846	0.846	0.846	0.843	

VCAP d6, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.768	0.785	0.793	0.807	0.803	0.799	0.79	0.79	
Intensity, σ = 0.99	0.791	0.803	0.807	0.811	0.808	0.8	0.794	0.794	
Intensity, σ = 0.995	0.796	0.808	0.81	0.808	0.804	0.795	0.79	0.79	
Intensity, σ = 0.998	0.798	0.812	0.811	0.803	0.798	0.788	0.785	0.785	
Intensity, σ = 0.999	0.797	0.812	0.81	0.799	0.794	0.784	0.781	0.781	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.789	0.802	0.799	0.794	0.794	0.793	0.785	0.781	0.763	
GSEA	0.792	0.803	0.806	0.802	0.801	0.8	0.792	0.788	0.774	

VCAP d24, AUROC performance

	10	50	100	500	1,000	5,000	10,000	12,494	(b)
Intensity, σ = 0.9	0.656	0.684	0.698	0.698	0.703	0.698	0.691	0.695	
Intensity, σ = 0.99	0.689	0.701	0.703	0.704	0.708	0.703	0.701	0.702	
Intensity, σ = 0.995	0.693	0.703	0.703	0.703	0.706	0.7	0.698	0.699	
Intensity, σ = 0.998	0.696	0.707	0.706	0.703	0.705	0.696	0.694	0.695	
Intensity, σ = 0.999	0.697	0.708	0.708	0.703	0.702	0.693	0.691	0.692	

	10	50	100	150	200	250	500	1,000	5,000	(size)
KS	0.7	0.708	0.7	0.697	0.697	0.697	0.695	0.692	0.682	
GSEA	0.69	0.705	0.704	0.704	0.703	0.703	0.701	0.697	0.69	

Figure S8. Robustness analysis of AUROC performance of the intensity-based, KS, and GSEA metrics to the gene set size, related to Figure 2.
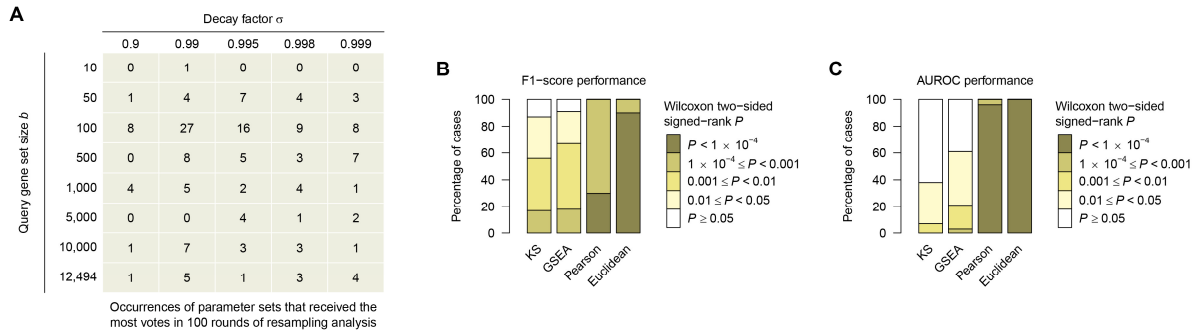
A

Query gene set size b	Decay factor σ				
	0.9	0.99	0.995	0.998	0.999
10	0	1	0	0	0
50	1	4	7	4	3
100	8	27	16	9	8
500	0	8	5	3	7
1,000	4	5	2	4	1
5,000	0	0	4	1	2
10,000	1	7	3	3	1
12,494	1	5	1	3	4

Occurrences of parameter sets that received the most votes in 100 rounds of resampling analysis

Figure S9. Resampling analysis for the parameter tuning and performance evaluation, related to Figure 2.

(A) Occurrences of parameter sets that received the most votes for the best set in a total of 100 rounds of the resampling analysis.

(B) Comparison of F1-score performance between the intensity-based metric with the KS, GSEA, Pearson-based, and Euclidean-based metrics in a total of 100 rounds of the resampling analysis (Wilcoxon two-sided paired signed-rank test).

(C) Comparison of AUROC performance between the intensity-based metric with the KS, GSEA, Pearson-based, and Euclidean-based metrics in a total of 100 rounds of the resampling analysis (Wilcoxon two-sided paired signed-rank test).
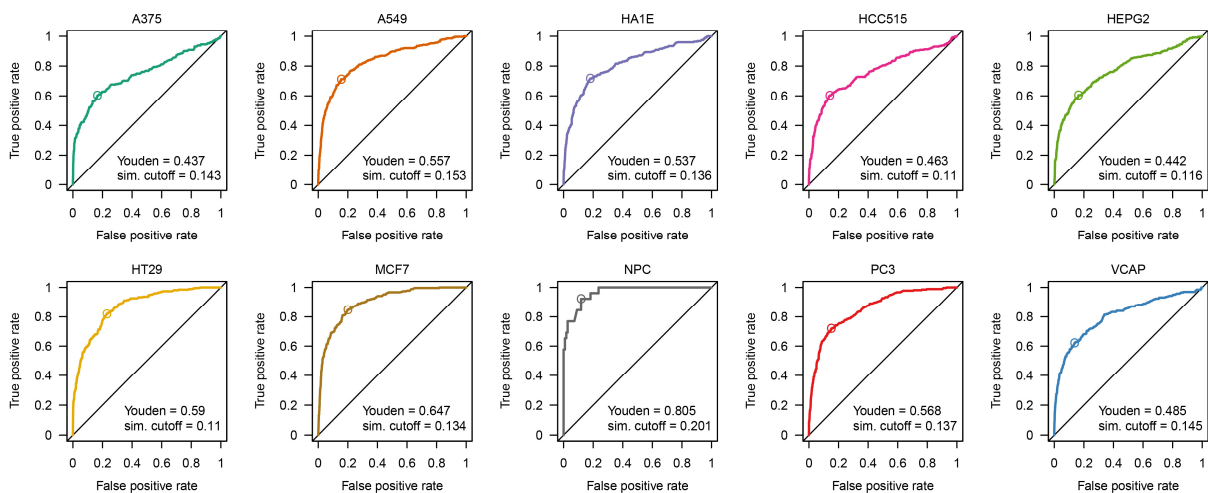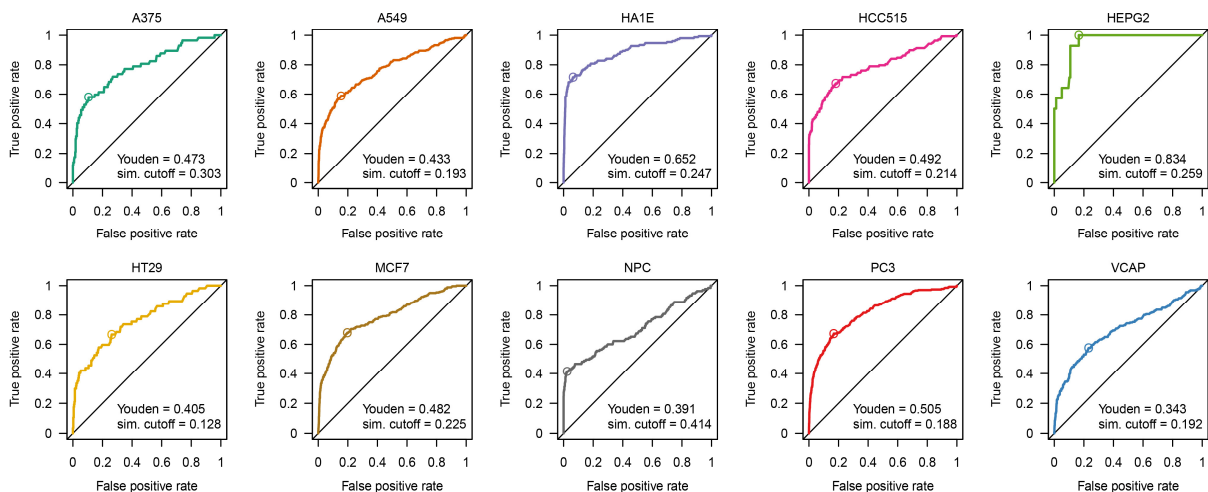
Figure S10. Similarity cutoffs defined by Youden's index using the optimal intensity-based similarity metric with respect to the gold-standard clustering, related to Figure 3.

A

Query: MCF7, vorinostat (GSE69845)
Reference: MCF7, d6 (LINCS)

Ordered compounds (*n* = 3,063)

Top 10 similar compounds:
vorinostat (#)
apicidin
belinostat
THM−I−94
trichostatin−a
HC−toxin
dacinostat
scriptaid
panobinostat
NCH−51

B

Query: Ishikawa, vorinostat (GSE69849)
Reference: MCF7, d6 (LINCS)

Ordered compounds (*n* = 3,063)

Top 10 similar compounds:
vorinostat (#)
apicidin
belinostat
trichostatin−a
HC−toxin
dacinostat
givinostat
THM−I−94
ISOX
panobinostat

C

Query: MCF7, triptolide (GSE28662)
Reference: MCF7, d6 (LINCS)

Ordered compounds (*n* = 3,063)

Top 10 similar compounds:
triptolide (#)
dactinomycin
CGP−60474
daunorubicin
alvocidib
BMS−387032
PIK−75
chromomycin−a3
mitoxantrone
doxorubicin

D

Query: MCF7, triptolide (GSE28662)
Reference: A549, d6 (LINCS)

Ordered compounds (*n* = 2,610)

Top 10 similar compounds:
daunorubicin
dactinomycin
triptolide (#)
alvocidib
A−443644
PIK−75
ER−27319
chromomycin−a3
JNK−9L
AT−7519

E

Query: MCF7, triptolide (GSE28662)
Reference: HA1E, d6 (LINCS)

Ordered compounds (*n* = 2,665)

Top 10 similar compounds:
doxorubicin
ZG−10
ER−27319
PIK−75
bisindolylmaleimide−ix
JNK−9L
pidorubicine
daunorubicin
triptolide (#)
mitoxantrone

Figure S11. Validation of the intensity-based similarities using independently derived perturbation profiles to query the LINCS database, related to Figure 3.

Shown are the intensity-based similarities (using optimal intensity-based similarity metric) between a query chemical perturbation in a cell type (with a GEO accession number) and the reference chemical perturbations in a cell type (LINCS) as indicated. The top 10 similar reference compounds (red points) are listed below, with a hash (#) indicating the same reference compound as the query. For each plot, a horizontal dash gray line indicates the similarity cutoff defined by the Youden's index in the corresponding context as in Figure S10.
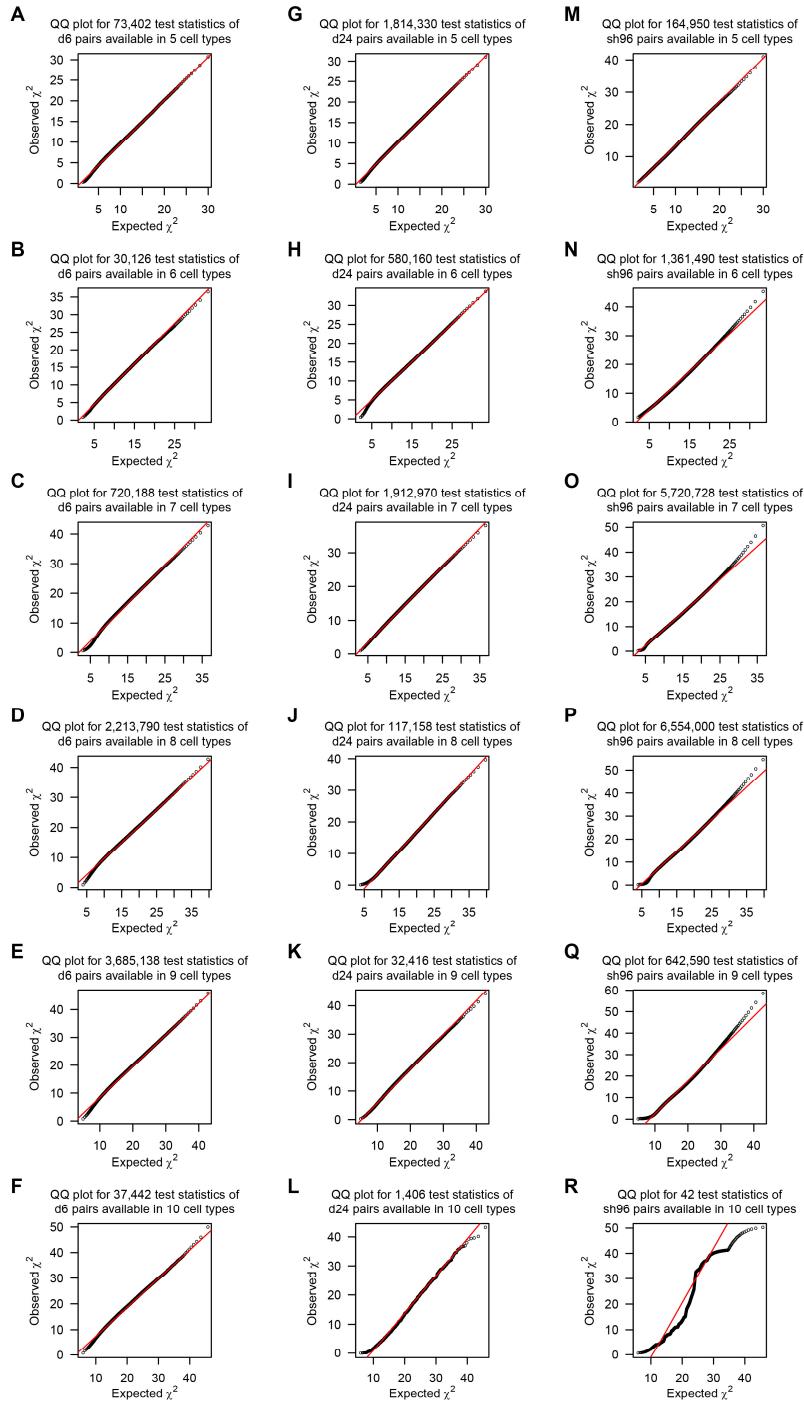
Figure S12. Quantile–quantile (Q–Q) plot for the chi-squared test statistics of the perturbation pairs used for RSS analysis, related to Figure 4.

For each of the d6 (A–F), d24 (G–L) and sh96 (M–R) perturbation types, we used the relative ranks of intensity-based similarities to generate a Q–Q plot for the chi-squared test statistics of the perturbation pairs available in n = 5 (A, G, M), 6 (B, H, N), 7 (C, I, O), 8 (D, J, P), 9 (E, K, Q), or 10 (F, L, R) cell types, respectively (corresponding to chi-squared distribution with 10, 12, 14, 16, 18, or 20 degrees of freedom, respectively). For each plot, we drew a red line that passes through the 100th and 900th permilles (i.e., 1000-quantiles).
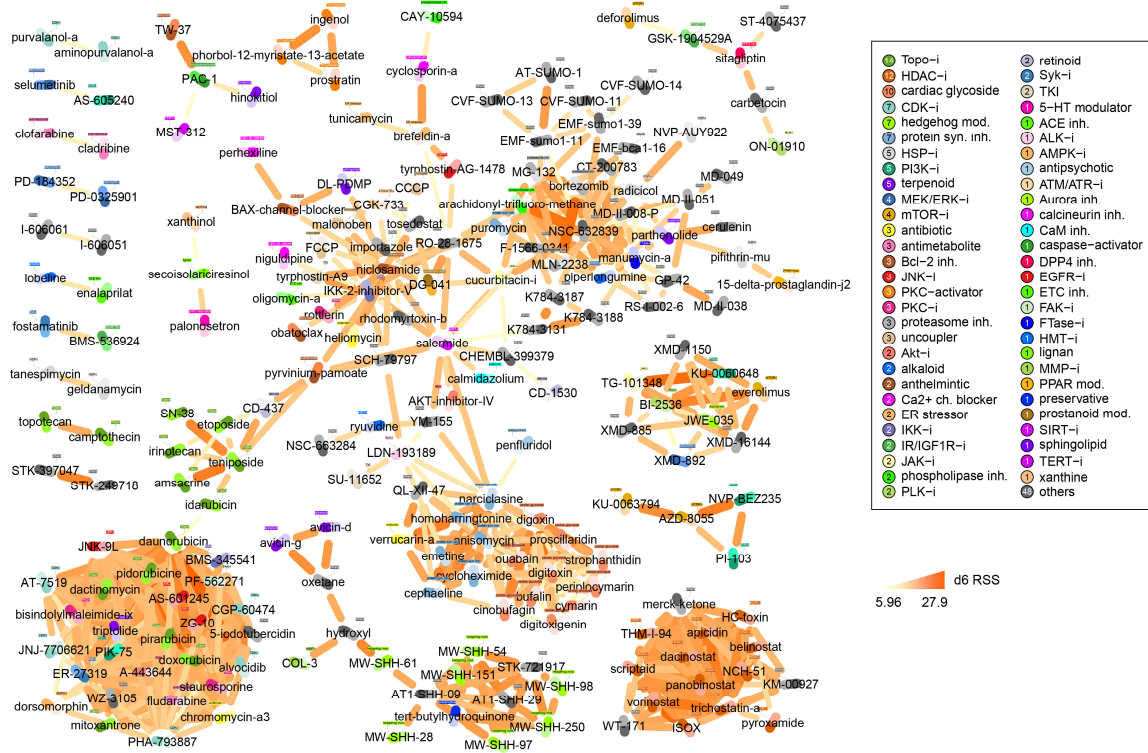
Figure S13. d6 RSS network, related to Figure 4.

Displayed are d6 chemical perturbation pairs that were recurrently similar across cell types (FDR < 0.001; Table S4; see Table S7 for network communities and their enrichment analyses). A primary MoA is assigned to each chemical perturbagen and color-coded on the darker side of each icon as well as in the text inset. The right box summarizes the occurrences of primary MoAs in the network.
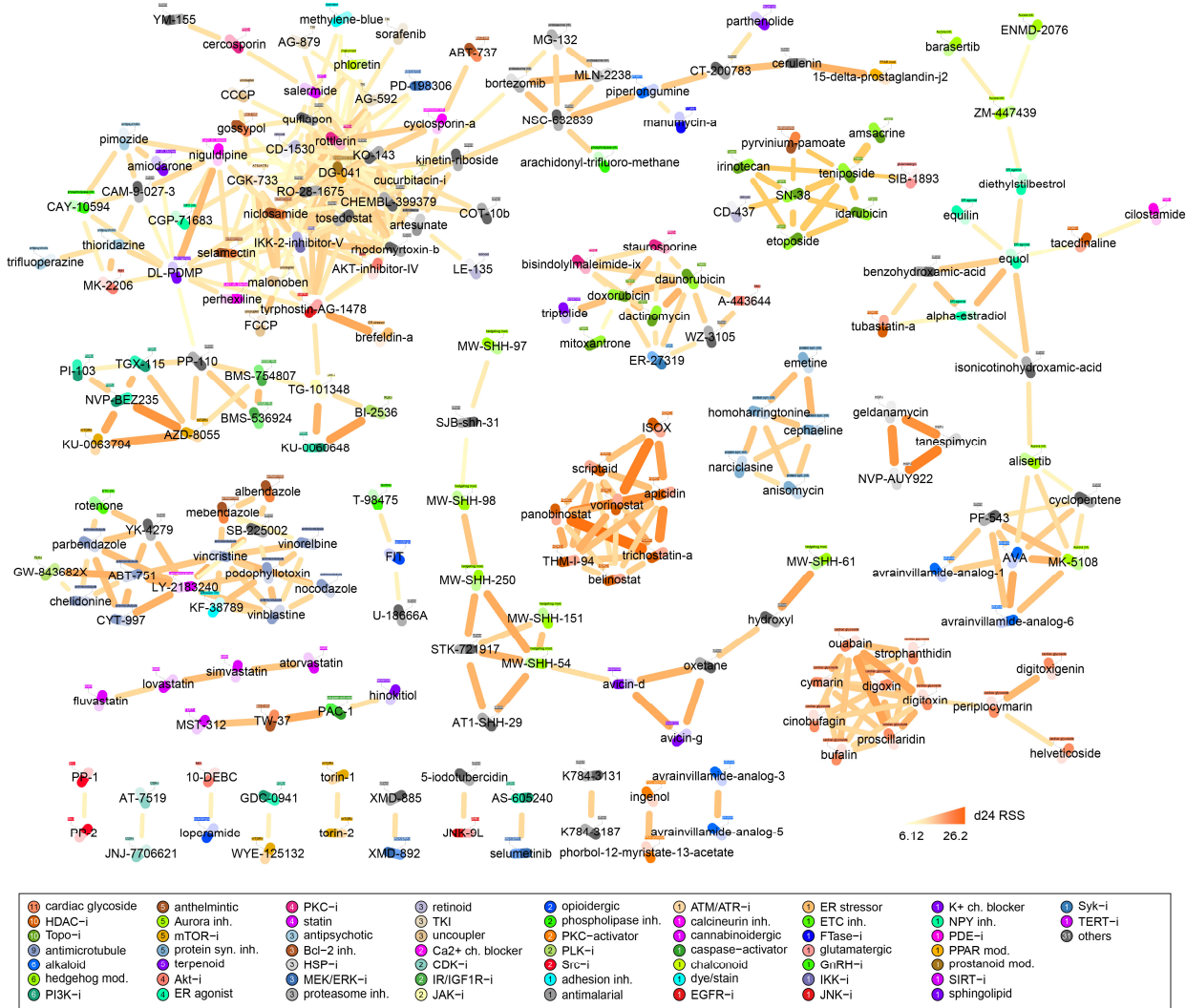
Figure S14. d24 RSS network, related to Figure 4.

Displayed are d24 chemical perturbation pairs that were recurrently similar across cell types (FDR < 0.001; Table S5; see Table S8 for network communities and their enrichment analyses). A primary MoA is assigned to each chemical perturbagen and color-coded on the darker side of each icon as well as in the text inset. The box below summarizes the occurrences of primary MoAs in the network.
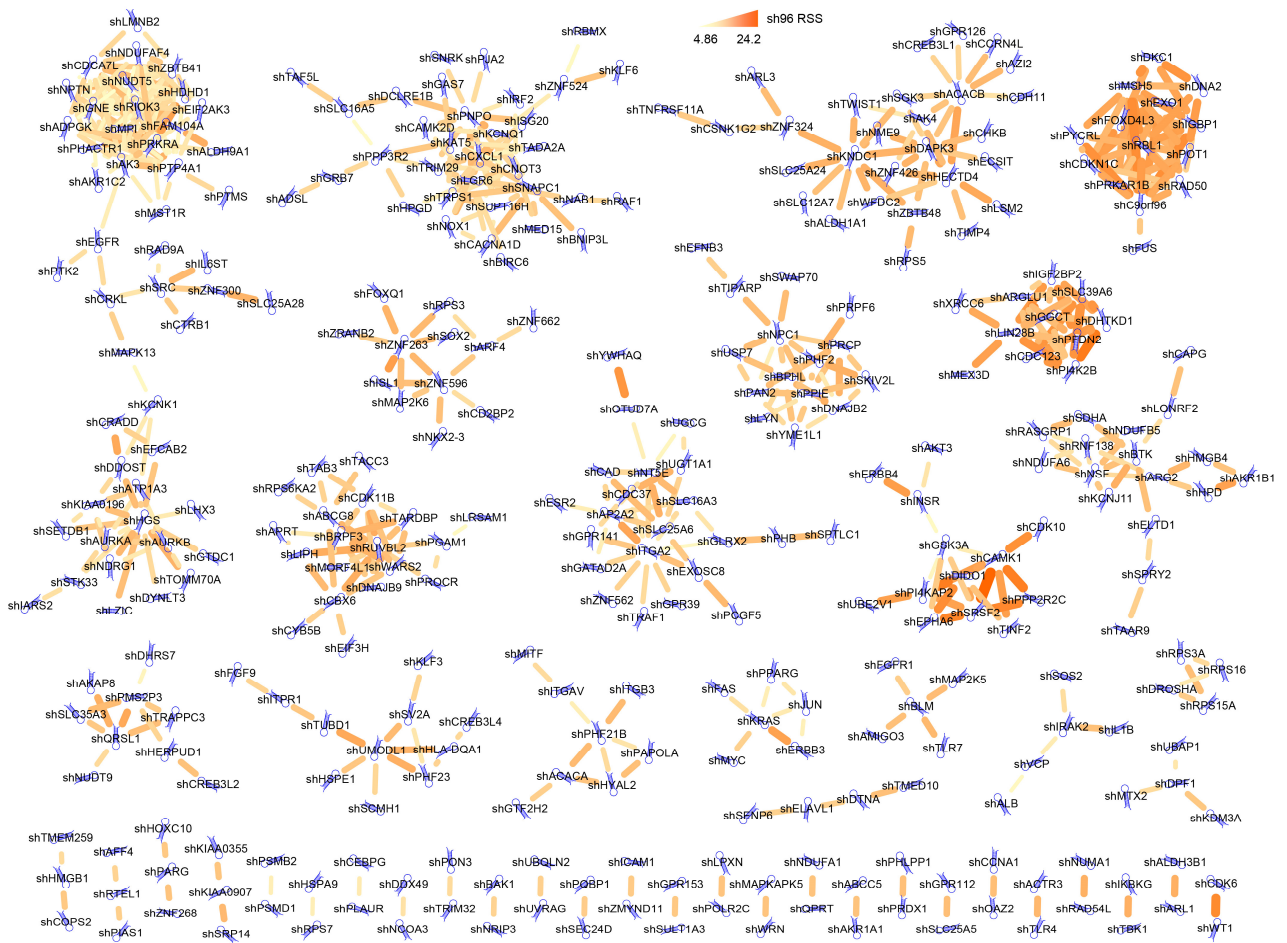
Figure S15. sh96 RSS network, related to Figure 4.

Displayed are sh96 genetic perturbation pairs that were recurrently similar across cell types (FDR < 0.001; Table S6; see Table S9 for network communities and their enrichment analyses).
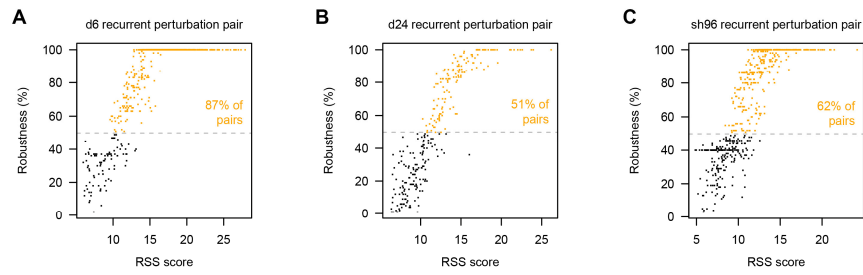
Figure S16. Resampling analysis for the recurrent perturbation pairs, related to Figure 4.
For recurrent perturbation pairs in each of the d6 (A), d24 (B), and sh96 (C) perturbation types, the percentages of cases in which they were recovered in the resampling analysis (as robustness; *y*-axis) were plotted against their real RSS scores (i.e., those from the RSS analysis when 100% of cell types were used; *x*-axis). For each plot, the robustness at 50% recovery is indicated by a horizontal dash gray line, above which the proportion of pairs (orange) is shown.
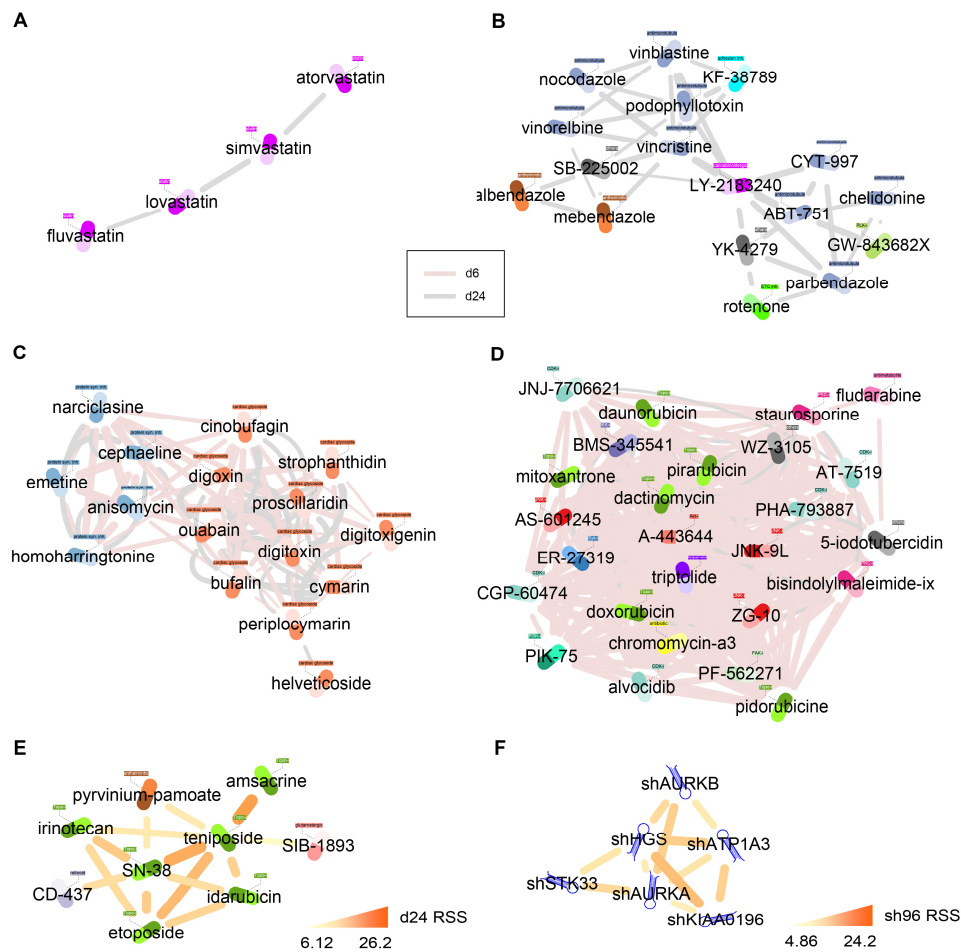
Figure S17. Subnetworks related to RSS analyses, related to Figure 4.

(A) A connected component involving lipid-lowering statins in the drug RSS network (merging d6 and d24 chemical perturbation networks).

(B) A connected component involving most antimicrotubules in the drug RSS network.

(C) A subnetwork spanning protein synthesis inhibitors and cardiac glycosides in the drug RSS network.

(D) A subnetwork of the PKC inhibitor staurosporine in the drug RSS network.

(E) A connected component involving the anthelmintic pyrvinium pamoate in the d24 RSS network.

(F) A subnetwork of the shAURKA perturbation in the sh96 RSS network.

Figure S18. Enrichments of hallmark gene sets for RSS network communities, related to Figure 4.
Significant enrichments (corrected hypergeometric *P* < 0.001) of hallmark gene set collections from MSigDB for drug (A) and sh96 (B) RSS network communities (Tables S9 and S10). For each drug community, the number of drugs and the composition of MoAs are shown.

Figure S19. Enrichments of canonical pathways for RSS network communities, related to Figure 4. Significant enrichments (corrected hypergeometric *P* < 0.001) of canonical pathways from MSigDB for drug (A) and sh96 (B) RSS network communities (Tables S9 and S10). For each drug community, the number of drugs and the composition of MoAs are shown.

Figure S20. Enrichments of immunologic signatures for drug RSS network communities, related to Figure 4.

Shown are significant enrichments (corrected hypergeometric *P* < 0.0001) of immunologic signatures from MSigDB (Table S10) with the number of drugs and the composition of MoAs for each drug community indicated below.
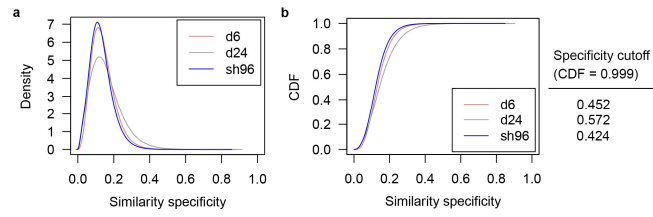
Figure S21. Similarity specificities of perturbation pairs, related to Figure 5.
Empirical density functions (A) and CDFs (B) against similarity specificities are shown for each perturbation type.

Transparent Methods

Data acquisition and preprocessing. We accessed LINCS (Subramanian et al., 2017) (Library of Integrated Network-based Cellular Signatures) L1000 gene expression datasets (level 3, normalized using invariant set scaling followed by quantile normalization; http://www.lincscloud.org/, now available at Gene Expression Omnibus; downloaded on 18 September 2014), which comprises 1,328,098 perturbation profiles for 77 cell types. All associated metadata were retrieved via Lincscloud API (http://api.lincscloud.org/, now replaced by https://clue.io/api) using R package 'rjson' (version 0.2.14). To facilitate similarity computations and to avoid redundancies, we reduced each probe-wise expression vector (dimension = 22,268; Affymetrix Human Genome U133A Array) to a gene-wise expression vector (dimension = 12,494) by taking the median of each gene with multiple probes in reference to R package 'hgu133a.db' (version 2.14.0).

We considered the following three types of perturbation: exposure to chemical drugs for 6 hours (abbreviated as d6); exposure to chemical drugs for 24 hours (d24); and exposure to short hairpin RNAs (shRNAs) for 96 hours (sh96). The following 10 selected cell types were analyzed: NPC (a human induced pluripotent stem cell-derived neural progenitor cell line); HA1E (a human kidney epithelial immortalized cell line); MCF7 (a human breast adenocarcinoma cell line); A549 and HCC515 (two human non-small cell lung adenocarcinoma cell lines); HT29 (a human colorectal adenocarcinoma cell line); HepG2 (a human hepatocellular carcinoma cell line); PC3 (a human prostate adenocarcinoma cell line); VCaP (a human metastatic prostate cancer cell line); and A375 (a human malignant melanoma cell line). All expression profiles were $\log_2$-transformed. The names of the perturbagens were set to those listed under the field 'pert_iname' in the metadata. For chemical perturbagens, we excluded names that were unclear in their meanings and might be unfamiliar to general users outside the Broad Institute; these names started with 'BRD-', 'SA-', 'BG-', 'ARG-', 'Broad-Sai-', 'FU-', 'JAS07-', 'KU-C', 'KUC', 'RAN-', 'SD-', 'ST-0', 'ST-2', 'ST-4', 'TL-', 'TUL', 'VU-', 'WZ-4-', 'DAC-', or 'WY-01'. The total number of non-redundant perturbagens analyzed in each cell type are summarized in Table S3, corresponding to a union of 3,332 chemical and 3,934 genetic perturbagens. We manually curated primary mechanisms of action (MoAs) for all chemical perturbagens with associated LINCS metadata and MoA tag descriptions (Figure S1 and Table S2).

Intensity-based similarity metric. We used perturbation intensity (i.e., the gene expression difference between perturbed and unperturbed conditions) and rank-based scoring functions with two tunable parameters to define an intensity-based similarity metric, inspired by the general observation from a biological perspective that a given perturbation should only affect a restricted panel of genes (Felix and Barkoulas, 2015). We therefore proposed that the most extremely perturbed genes compared to the unperturbed counterpart might be appropriate and sufficient to represent a given perturbation. For each perturbation u treated in a given cell type with N detection instances (one detection instance indicates one well of measurement in a 384-well plate for fluorescence detection

and the corresponding experiments may be performed in technical and biological replicates), we computed the 'perturbed profile' d_u:

$$d_u = \frac{1}{N} \sum_{j=1}^{N} \left(r_{u,j} - e_{u,j} \right)$$

where r and e are expression vectors of the perturbation instance j and the corresponding control on the same plate of j, respectively. This perturbed profile represents the average perturbation intensity from available experiments in a way that reduces cross-plate batch effects, the extent of which became increasingly perceptible for some perturbations with multiple experimental replicates. Note that for most perturbations in a given cell type, experiments were performed in only one biological replicate with no more than three technical replicates, together with cross-plate batch effects, making it impractical to run standard differential expression analyses on LINCS data. Next, we derived an expression intensity rank vector ρ_u, where the 1st most perturbed entry (regardless of the direction of gene regulation) is assigned 1, the 2nd most perturbed entry is assigned 2, and so on. This intensity rank allowed us to define a gene set size b that could be tuned to yield a query gene set B in which $|B| = b$. That is, the number of the top most perturbed genes for B could be fixed at $b = 10, 50, 100, 500, 1,000, 5,000, 10,000,$ or $12,494$ (the maximum size) with a corresponding query vector θ_u for each sample u for which the selected entries (within B) are assigned 1 (up) or −1 (down) and the unselected (outside B) are assigned 0. Then, for any two perturbations u and v when v was used as a query, we computed the forward and reverse match scores $fwd_{u \leftarrow v}$ and $rev_{u \leftarrow v}$ (of v to u), respectively, as follows:

$$fwd_{u \leftarrow v} = \sum_{g \in G} \left[\!\left[sign\left(d_u[g] \right) = \theta_v[g] \right]\!\right] \times \sigma^{\left(\rho_u[g]-1 \right)}$$

$$rev_{u \leftarrow v} = \sum_{g \in G} \left[\!\left[sign\left(d_u[g] \right) = -\theta_v[g] \right]\!\right] \times \sigma^{\left(\rho_u[g]-1 \right)}$$

where G is the gene space ($|G| = 12,494$), $[\![\cdot]\!]$ returns 1 if the expression inside gives a TRUE value and returns 0 otherwise, $sgn(\cdot)$ is the sign function, $d_u[g]$ is the entry of the perturbed profile d_u corresponding to the gene g, $\theta_v[g]$ is the entry of the query vector θ_v corresponding to the gene g, and the other tunable parameter decay factor σ to adjust the weights of ranks for σ fixed at 0.9, 0.99, 0.995, 0.998, or 0.999. Subsequently, we derived the directional similarity score of v to u as $sim_{u \leftarrow v} = (fwd_{u \leftarrow v} - rev_{u \leftarrow v})/S$, where S is the theoretical maximum match score (i.e., the sum of the geometric progression $\sigma^0, \sigma^1, ..., \sigma^{b-1}$ as the geometric series). Under this scenario, a +1 similarity score represents the perfect forward match whereas −1 represents the perfect reverse match. Similarly, we defined the

similarity score of u to v (when u was used as a query), where $sim_{u \leftarrow v} \neq sim_{v \leftarrow u}$. In this way, an undirected similarity score for any two perturbations v and u was obtained by averaging their directional similarity scores as $sim_{u \leftrightarrow v} = (sim_{u \leftarrow v} + sim_{v \leftarrow u})/2$.

Parameter optimization, comparison, and performance evaluation. We leveraged a ground truth clustering of a selected panel of chemical perturbagens with well-established primary MoAs ($n = 74$, but not all of these perturbagens were available in a given cell type at 6 or 24 h; thus, the exact number was variable; Table S1) to determine the optimal intensity-based similarity metric and facilitate the comparison among the intensity-based and all other similarity metrics.

For comparison, we considered four types of gene expression similarity metrics defined by the Kolmogorov–Smirnov (KS) statistic (Massey, 1951) (as the state of the art (Lamb et al., 2006; Iorio et al., 2010; Sirota et al., 2011; Kidd et al., 2016)), Gene Set Enrichment Analysis (GSEA) method (in which exponent = 1 was used for the standard enrichment score (Subramanian et al., 2005)), Pearson correlation, and Euclidean distance (two in common use (D'Haeseleer, 2005)). All these metrics were defined at the level of the 'perturbed' profiles $\boldsymbol{d_u}$ and $\boldsymbol{d_v}$ for any two perturbations u and v. For KS and GSEA metrics, we considered a range of set sizes at 10, 50, 100, 150, 200, 250, 500, 1,000 and 5,000 for each of the top- and bottom-ranked genes used for their matching algorithms; a metric was then defined by taking the difference of the matching scores between these two extremes ('top' minus 'bottom') over a value of 2 (the maximum of the difference when a perturbagen uses its ranked list to query itself, as did in those state-of-the-art studies (Iorio et al., 2010; Sirota et al., 2011; Kidd et al., 2016) except CMap (Lamb et al., 2006), which takes the largest value (≤ 2) among the absolute values of all possible differences produced when a gene signature is used to query the entire database).

For each similarity metric (intensity-based, KS, GSEA, Pearson, or Euclidean), the clustering performance was evaluated using the external cluster validity index F1 score based on a mapping approach to define true positives, true negatives, false positives, and false negatives (Wiwie et al., 2015). We explicitly considered the following five biomedical clustering methods: hierarchical clustering (HC) using average (HC-avg; using hclust in R base) or Ward (HC-ward; using agnes in R package cluster version 1.15.2) linkage; partitioning around medoids (PAM; using pam in R package cluster version 1.15.2); affinity propagation (AP; using apclusterK in R package apcluster version 1.4.3); and exemplar-based agglomerative clustering (EBC; using aggExCluster in R package apcluster version 1.4.3), with default function arguments. For each context (a perturbation type in a cell type) and each 'metric–method' setting (one metric plus one clustering method), we computed F1 scores for all possible clusters across a full range of number of clusters $k = 2,..., n - 1$ (where n is the number of drugs in the ground truth clustering) and determined an optimal clustering by picking the k with the best F1 score. We then prioritized these intensity-based clusters and all other clusters separately according to F1 scores and visualized these results with clusters of each clustering color-coded in a way that starting from the last row, the color consistency was maximized to the first row (i.e., the best of the intensity-based clusterings; Figure 2A provides an illustrative example of d6 perturbation

type in A375).

An optimal parameter set (b^*, σ^*) for the intensity-based similarity metric was determined by majority voting of occurrences of parameter sets among the top five intensity-based clusterings (including equal rank whenever possible) across 20 contexts (two perturbation types plus 10 cell types; Figure 2B). We also examined the voting results when varying the number of the top-scoring (Figure S2) or the bottom-scoring (Figure S3) intensity-based clusterings used for voting.

We first compared intensity-based clusterings with all other clusterings using these F1 scores. We extracted the highest F1 score for each context from clusterings using the intensity-based (across all parameter sets or using the optimal parameter set), KS or GSEA (across all gene set sizes or using a fixed gene set size), Pearson-based, or Euclidean-based metric and tested the statistical significance of their distributions across all contexts using Wilcoxon two-sided paired signed-rank tests (intensity-based over the others; Figure 2C). For each context, we also assessed whether the F1 scores of the intensity-based clusterings within the top five, using the optimal parameter set, or across all parameter sets were significantly higher than other clusterings using one-sided Mann–Whitney U tests (Figure S4). Next, we compared the intensity-based similarity metric with all other similarity metrics using a strategy based on the area under the receiver operating characteristic (ROC) curve (AUROC) with respect to the ground truth. For each context, we varied the threshold τ on a selected metric such that a perturbation pair with the similarity value beyond τ indicates that they are in the same cluster, whereas a value below τ not is in the same cluster; thus, we generated a ROC with an associated AUROC score. We aggregated the results using the optimal intensity-based metric across all contexts (Figure 2D) or each metric for each context (Figure S5, in which a hash (#) indicates the metric that achieved the best context-specific AUROC). We also extracted the highest AUROC score for each context achieved by the intensity-based, KS, GSEA, Pearson-based, or Euclidean-based metric and tested the statistical significance of their distributions across all contexts using Wilcoxon two-sided paired signed-rank tests (intensity-based over the others; Figure 2E).

For each context, we assessed the robustness of intensity-based, KS, and GSEA similarity metrics to the gene set size (i.e., the query gene set size b for the intensity-based metric with a fixed decay factor σ, or the set size for each of the top- and bottom-ranked genes for the KS or GSEA metric) based on the F1-score (Figure S7) and AUROC (Figure S8) performance. For the analysis using F1-score performance, we used the median F1 score among the clusterings produced by the five clustering methods for a given metric (an intensity-based metric with fixed b and σ, or a KS or GSEA metric with a fixed set size) in each context.

To examine whether the results of the analysis for parameter tuning and performance evaluation were robust against chemical perturbagens, we repeated the analysis with resampling of 75% of chemical perturbagens for 100 times. For parameter tuning for the intensity-based metric, in each round of the resampling analysis, votes for parameter sets were collected among the top five intensity-based clusterings across all contexts (as did in Figure 2B), and the parameter set that received the most votes was counted (more than one parameter set is possible if votes are equal in each round). The occurrence

of each parameter set that received the most votes in a total of 100 rounds of the resampling analysis is summarized in Figure S9A. For performance comparison, in each round of the resampling analysis, the best F1 scores for clusterings across all contexts (as did in Figure 2C) or the best AUROC scores across all contexts (as did in Figure 2E) achieved by the intensity-based metric were compared with those by the KS, GSEA, Pearson-based, or Euclidean-based metric using Wilcoxon two-sided paired signed-rank tests. The resulting *P*-values of the tests on the F1-score and AUROC performance from a total of 100 rounds of the resampling analysis are summarized in Figure S9B and S9C, respectively.

Similarity cutoffs. We used the optimal parameter set (b^*, σ^*) to compute pairwise intensity-based similarities for all available perturbations in LINCS. For each perturbation type (d6, d24, or sh96), we binned similarity values into bins at 0.001 resolution (i.e., rounded to the nearest three decimals) and plotted the empirical cumulative distribution function (CDF) of these bins within each cell type or across all cell types (Figure 3A). This overall CDF (across all cell types) was then used to derive a positive (sim^+_{d6}, sim^+_{d24}, or sim^+_{sh96}) and a negative (sim^-_{d6}, sim^-_{d24}, or sim^-_{sh96}) heuristic similarity cutoffs such that the positive value is the binned similarity value where CDF is greater than or approximately equal to 0.999 and the negative value is the binned similarity value where CDF is less than or approximately equal to 0.001. For each d6 or d24 perturbation type in a given cell type, we also determined a similarity cutoff defined by Youden's index (Youden, 1950) (i.e., sensitivity + specificity − 1) of the ROC curve by associating the intensity-based similarities with the ground truth clustering as a comparison to its corresponding positive heuristic similarity cutoff (Figure S10).

To explore the ability of the intensity-based similarity metric to recover small-molecule connections between LINCS and other resources, we computed the perturbation intensities of some chemical perturbagens independently derived from Gene Expression Omnibus (GEO) to query the LINCS database. We used the GEO datasets with the following accession numbers: GSE69845 (vorinostat at 6 h in MCF7 cells), GSE69849 (vorinostat at 6 h in Ishikawa cells), and GSE28662 (triptolide at 2, 4, and 6 h in MCF7 cells). For each dataset, we converted the probe set IDs into the gene symbols while taking the median expression values for each gene with multiple probes. The query results are shown in Figure S11.

Recurrent similarity score (RSS) analysis. We used a strategy combining the intensity-based similarities and a modified Fisher's method (Jacobsen et al., 2013) *ad hoc* to infer recurrent perturbation–perturbation relationships. Given a perturbation pair (j, μ) available in cell type c, we first fixed j and computed the relative rank of its similarities to all other perturbations containing μ as follows:

$$rr_{\mu,j,c} = \frac{r_{\mu,j,c}}{\left|L_{j,c}\right|} - \frac{1}{2\left|L_{j,c}\right|}$$

where $L_{j,c}$ is the ordered (from negative to positive) vector of similarities involving perturbation j (excluding itself) in cell type c and $r_{\mu,j,c}$ is the rank for perturbation μ in the vector $L_{j,c}$ such that $r_{\mu,j,c}$ is uniformly distributed on the interval $(0, 1)$. Under the null hypothesis that no negative score exists between perturbations μ and j across all n cell types ($n = 10$), we would expect twice the negative natural logarithm of the product of these uniformly distributed relative ranks to approximately follow a chi-squared distribution with $2n$ degrees of freedom as follows (Figure S12):

$$H_0 : -2\sum_c \ln\left(rr_{\mu,j,c}\right) \sim X^2_{2n}$$

In this way, the chi-squared test gives a one-sided P-value called p_-. Alternatively, we could test the hypothesis at another side using the inverted rank ($|L_{j,c}| - r_{\mu,j,c} + 1$), giving another statistic called p_+. These two, one-sided P-values, p_- and p_+, were multiplied by 2 to adjust for multiple testing. Finally, for a given perturbation pair (j, μ) available in at least half of the cell types (≥ 5 in this case), we could define a recurrence score $REC_{j\to\mu}$ (when fix j and see μ) as the signed logarithm of these P-values as follows:

$$REC_{j\to\mu} = \begin{cases} \log_{10}\left(2 \times p_-\right), & \text{if } p_- < p_+ \\ -\log_{10}\left(2 \times p_+\right), & \text{if } p_+ < p_- \\ 0 & , \text{if } p_- = p_+ \end{cases}$$

Note that the rank of the similarity value of the perturbation pair (j, μ) in the context of j (i.e., fix j and see μ) should be different from that in the context of μ (i.e., fix μ and see j). Therefore, a recurrence score associating a perturbation pair (j, μ) in the context of μ could also be defined such that $REC_{\mu\to j} \neq REC_{j\to\mu}$. To resolve multiple testing problems, we used the Benjamini–Hochberg method to correct P-values as false discovery rates (FDRs) for all perturbation pairs. A recurrent similarity score (RSS) for any two perturbations j and μ was then defined as $RSS_{j,\mu} = (REC_{j\to\mu} + REC_{\mu\to j})/2$. A perturbation pair (j, μ) is considered recurrently similar if (1) both $FDR_{j\to\mu}$ and $FDR_{\mu\to j} < 0.001$, and (2) $sim_{j\leftrightarrow\mu} >$ the positive heuristic similarity cutoff (sim^+_{d6}, sim^+_{d24}, or sim^+_{sh96}) or $<$ the negative heuristic similarity cutoff (sim^-_{d6}, sim^-_{d24}, or sim^-_{sh96}) in at least three cell types. We performed analyses of RSS scores separately for each perturbation type (d6, d24, or sh96), and all recurrently similar perturbation pairs are provided in Tables S4–S6 and visualized using Cytoscape (version 3.0.2) in Figures S13–S15 (d6 plus d24 were merged as in Figure 4).

To assess the robustness of the recurrent perturbation pairs against cell types, we resampled 60%

of cell types and repeated the RSS analysis for 100 times for each perturbation type (d6, d24, or sh96). A recurrent perturbation pair (j, μ) is recovered in one round of the resampling analysis if both $FDR_{j \to \mu}$ and $FDR_{\mu \to j} < 0.001$. The robustness of a given recurrent perturbation pair is then defined as the ratio of the number of cases in which they are recovered over the number of tested cases (Figure S16).

To detect perturbation communities in each recurrent network (d6, d24, sh96, or 'd6+d24'), we used a Markov cluster algorithm (MCL (Enright et al., 2002); using mcl in R package MCL version 1.0 with default function arguments while addLoops = TRUE), which simulates random walks on a given adjacency matrix. The resulting clusters were labeled according to the number of perturbagens in descending order. We then exploited a strategy based on hypergeometric tests to assess enrichments in each of Molecular Signature Database (MSigDB (Subramanian et al., 2005); version 5.0) gene set collections across all perturbation clusters. For each perturbation u (of d6, d24, or sh96), we defined an average perturbed profile $\overline{d_u}$ across l cell type-specific perturbed profiles available ($l \leq n$ where n = 10 is the total number of cell types) as $\overline{d_u} = (\Sigma_{i=1}^{l} d_u^{(i)})/l$. For a chemical perturbagen with both d6 and d24 data in the recurrent drug network, we took the mean of its average perturbed profiles in d6 and d24. Next, for a given perturbation cluster, we computed the cluster average perturbed profile as the mean of all average perturbed profiles of involved perturbations, from which the most extremely perturbed genes (here, top 100) were used for the enrichment analysis. We then performed hypergeometric tests on enrichments for any MSigDB collection of interest across all perturbation communities in each recurrent network separately and adjusted the P-values using the Benjamini–Hochberg correction. All perturbation communities and their enrichment results for each recurrent network are provided in Tables S7–S10. To visualize enrichment results, we selected perturbation clusters and gene sets with at least one significant enrichment (corrected $P < 0.001$ in Figures S18 and S19, and $P < 0.0001$ in Figure S20 for presentation purposes) and then defined ubiquitous gene sets as having significant enrichments in at least 50% of considered perturbation clusters. The ubiquitous gene sets and perturbation clusters were ordered by the counts of valid enrichment P-values and, if counts were equal, they were ordered by the sum of magnitudes of $-\log_{10}(P\text{-value})$. The remaining non-ubiquitous gene sets were grouped according to which perturbation cluster had the most significant enrichment P-value. For each enriched chemical perturbation cluster, we showed the proportion of MoAs and indicated a predominant MoA (accounting for > 50% of MoAs in the cluster) when possible.

Highly specific similarity (HSS) analysis. For each perturbation pair available in at least half of n cell types ($n = 10$; ≥ 5 in this case), we defined a similarity specificity score as $specificity(X) = 1 - (entropy(X)/\log_2(l))$ where X is the similarity vector $[s_1, s_2,..., s_n]$ (a NA value was assigned to an entry without available experiments to define the similarity in that cell type), $entropy(\cdot)$ is the Shannon entropy such that $entropy(X) = 1 - \Sigma p_i \times \log_2(p_i)$ in which $p_i = |s_i|/\Sigma|s_i|$ for $i = 1,..., l$, and l is the number of available similarity values in X (i.e., non-NA entries; $n/2 \leq l \leq n$). We then used these similarity specificity values (binned at 0.001 resolution, i.e., rounded to the nearest three decimals) to generate an empirical CDF from which a similarity specific cutoff was derived as the binned value where CDF

is greater than or approximately equal to 0.999 (Figure S21). A perturbation pair is considered highly specific if the following conditions were met: (1) measurements occurred in at least half of cell types; (2) a similarity value was greater than the positive or less than the negative heuristic similarity cutoff in the dominant cell type as the HSS; and (3) its specificity was greater than the similarity specificity cutoff. We performed HSS analyses separately for each perturbation type (d6, d24, or sh96), and all highly specific perturbation pairs are summarized in Tables S11–S13 with a d24 HSS network visualized in Figure 5.

Experimental validation. *Cell culture and chemicals.* MCF7, A549, H1299, and H1975 cells were purchased from American Type Culture Collection (ATCC). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and incubated at 37 °C in humidified atmosphere with 5% CO_2 and routinely passaged when 90–95% confluent. Cells tested negative for mycoplasma. Pyrvinium pamoate (P0027), etacrynic acid (SML1083), and crystal violet (C3886) were purchased from Sigma-Aldrich. DMEM (12800), FBS (A3160601), and trypsin-EDTA (15400054) were obtained from Thermo Fisher. MTS (G1111) was ordered from Promega. Antibodies against γH2AX (9718, Cell Signaling), p53 (sc-126, Santa Cruz), TOP2A (ab74715, Abcam), p-TOP2A (Ser1106; ab75765, Abcam), and β-actin (MAB1501, Merck) were purchased from commercial vendors. Horseradish peroxidase-conjugated secondary antibodies were obtained from Abcam. Enhanced chemiluminescence reagent was ordered from Bio-Rad.

MTS assay. Cells (2×10^3 per well) were seeded in a 96-well plate for 24 h followed by indicated chemical treatments. At the endpoint, 20 μL MTS solution was added per well. The plate was incubated for 2 h at 37 °C with 5% CO_2 before measurement of the absorbance at 490 nm (A_{490}) using a ELISA reader (Bio-Rad). Data are represented by A_{490} values with background correction and normalized with the corresponding control group (DMSO) as 100%.

Clonogenic assay. Cells (3×10^4 per well) were seeded in a 6-well plate for 24 h followed by the indicated chemical treatments for 72 h. The remaining cells were then trypsinized and seeded in a new 6-well plate (500 per well) for 14 days. The colonies were fixed with 100% methanol overnight and stained with 1% crystal violet solution.

Western blot analysis. Cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.4; Bioman), 150 mM NaCl, 1% NP40, 0.1% SDS (Bioshop), and 0.5% sodium deoxycholate (Sigma) with a protease inhibitor (16012540, Bioshop) and protein phosphatase inhibitor cocktail (FC0050-0001, Bionovas) immediately added before use. After sonication, protein concentrations were determined using the BCA method (23225, Thermo Fisher). Each sample with equal protein concentrations was subjected to SDS-PAGE and transferred to a PVDF membrane. Nonspecific binding was blocked with 5% nonfat milk. After immunoblotting with the desired primary antibodies, membranes were incubated with appropriate secondary antibodies. Protein bands were detected using ECL reagent.