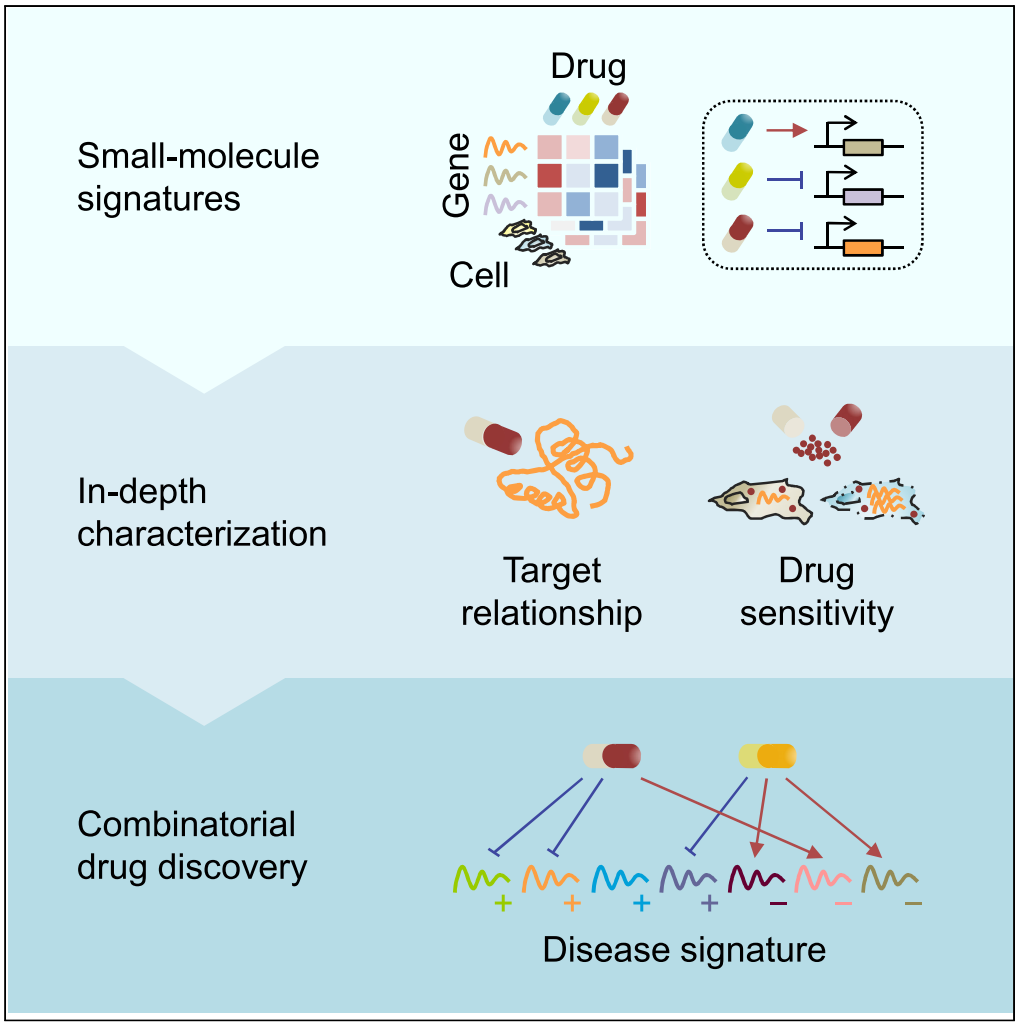
iScience

CellPress

Article

Perturbational Gene-Expression Signatures for Combinatorial Drug Discovery



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

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

HIGHLIGHTS

Compound signatures targeting non-oncogene addiction for combinatorial drug discovery

These signatures are reproducible and linked to cancer hallmarks and drug sensitivity

Two synergistic drug combinations are experimentally confirmed *in vitro*

Huang et al., iScience *15*, 291–306
May 31, 2019
https://doi.org/10.1016/j.isci.2019.04.039

iScience

CellPress

Article

Perturbational Gene-Expression Signatures for Combinatorial Drug Discovery

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

SUMMARY

Cancer is a complex disease that relies on both oncogenic mutations and non-mutated genes for survival, and therefore coined as *oncogene and non-oncogene addictions*. The need for more effective combination therapies to overcome drug resistance in oncology has been increasingly recognized, but the identification of potentially synergistic drugs at scale remains challenging. Here we propose a gene-expression-based approach, which uses the recurrent perturbation-transcript regulatory relationships inferred from a large compendium of chemical and genetic perturbation experiments across multiple cell lines, to engender a testable hypothesis for combination therapies. These transcript-level recurrences were distinct from known compound-protein target counterparts, were reproducible in external datasets, and correlated with small-molecule sensitivity. We applied these recurrent relationships to predict synergistic drug pairs for cancer and experimentally confirmed two unexpected drug combinations *in vitro*. Our results corroborate a gene-expression-based strategy for combinatorial drug screening as a way to target non-mutated genes in complex diseases.

INTRODUCTION

Cancer is a genetic disease marked by extensive heterogeneities within the cells of a primary tumor (intratumoral heterogeneity) and also between the primary tumors of different patients (intertumoral heterogeneity) (Dagogo-Jack and Shaw, 2018). Our increasing comprehension of cancer genome has informed diverse mutational spectra that are responsible for the malignant transformation across various cancer types (Vogelstein et al., 2013). Together with the dependency on these oncogenic driver mutations for their growth and survival, cancer cells can develop an addiction to certain genes that are themselves not oncogenic but whose functions are required for maintenance of the tumorigenic state (Luo et al., 2009). These needs of both oncogenes and non-mutated genes for cell survival, coined as *oncogene and non-oncogene addictions*, respectively, add further complexity to the classical hallmarks and stress phenotypes of cancer (Hanahan and Weinberg, 2011; Luo et al., 2009). Strategies that interfere with this oncogenic circuitry have been proved to serve as attractive points for cancer therapies (Hanahan and Weinberg, 2011; Luo et al., 2009; Nagel et al., 2016).

Despite substantial efforts, anticancer drug resistance has become one of the biggest challenges in global human health (Holohan et al., 2013). Initial responses to conventional chemotherapeutics and molecularly targeted agents are greatly affected by intertumoral heterogeneity, whereas a gradual loss of drug sensitivity can often arise from intratumoral heterogeneity (Dagogo-Jack and Shaw, 2018; Saunders et al., 2012). Functional interplays between distinct hallmark capabilities are also enabling drug-tolerant cancer cells to evolve a resistance mechanism during the prolonged treatment (Hata et al., 2016; Ramirez et al., 2016). To thwart acquired resistance, combination drug therapies are now garnering much attention for their ability to cooperatively eliminate heterogeneous cancer cells within a single tumor (Al-Lazikani et al., 2012; Flemming, 2015; Jia et al., 2009; Webster, 2016). In view of high attrition rates in the *de novo* drug development (Waring et al., 2015), one tantalizing option for exploring combination opportunities in oncology is to consider already launched drugs that have well-documented safety profiles (Bertolini et al., 2015; Corsello et al., 2017). However, a critical step toward fully realizing the potential of combination therapy is to identify synergistic drug interactions before clinical testing (Lopez and Banerji, 2017). Given the tremendous number of all possible drug combinations, a systems-level method that can predict drug synergy to an unprecedented scale for every disease is urgently required (Kwong et al., 2013; Ryall and Tan, 2015; Sun et al., 2013).

Cellular molecular responses to small-molecule treatments have provided many possibilities for cancer drug discovery (Moffat et al., 2014). The complex polypharmacological interactions following small-molecule

[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.2019.04.039

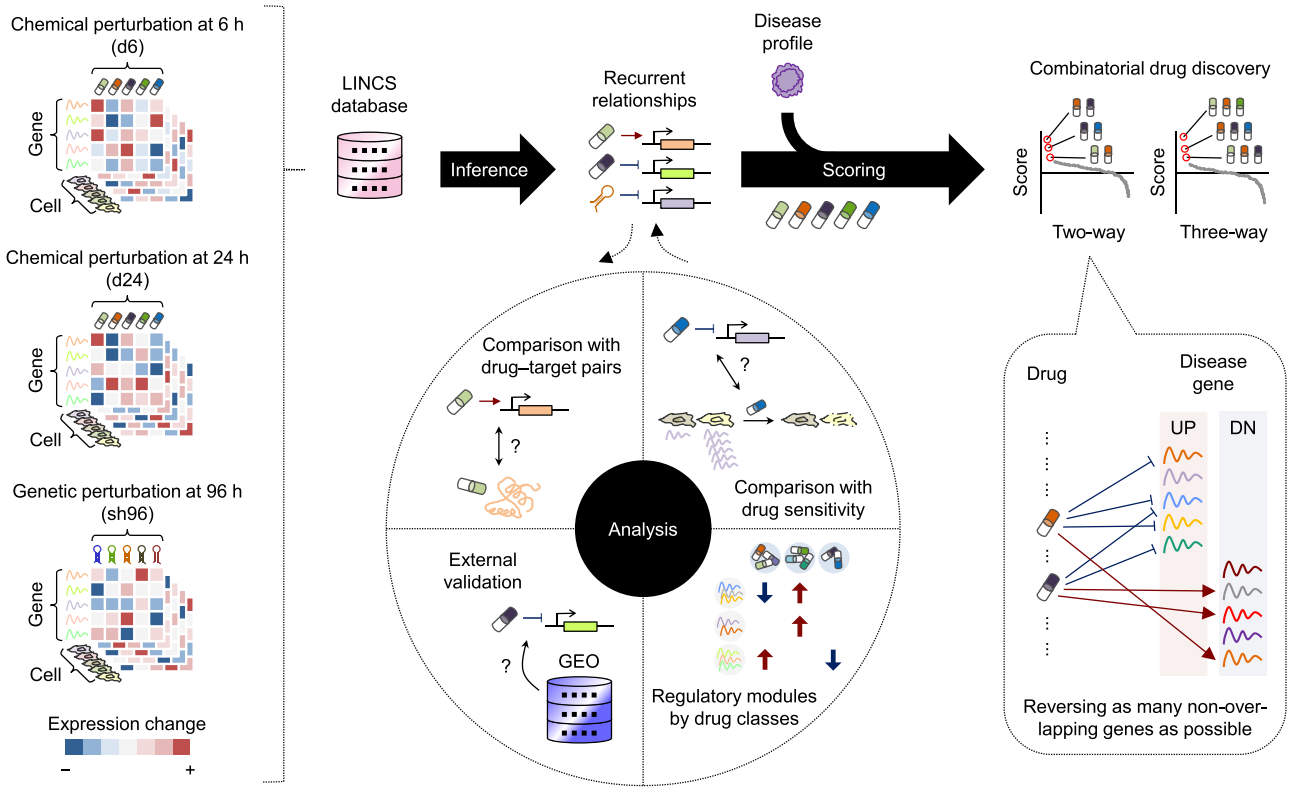


Figure 1. Study Workflow

To identify recurrent relationships between perturbations and regulated transcripts, we used the gene-expression profiles of more than 7,000 different chemical and genetic perturbagens across 10 selected cell lines from the LINCS L1000 resource. The inferred transcript-level recurrences were first compared with known small molecule-target annotations, benchmarked against independent perturbation datasets, characterized at the small-molecule-class regulatory level, and then connected with small-molecule sensitivity. We hypothesized that drug synergy can be informed by correlating a given disease signature with the combined patterns of gene reversal achieved by small-molecule-regulated recurring transcripts. To investigate this possibility, we applied this method to predict synergistic drug combinations across cancer types and validated the predictions *in vitro*.

treatments would be reflected on variations in several molecular features (Feng et al., 2009), among which gene expression is popularly used for such efforts (Lamb et al., 2006). Although most molecular efficacy targets of approved drugs are rarely transcription factors (Santos et al., 2017), there is increasing evidence that changes in gene expression following small-molecule treatments are usually pervasive and may link to drug activity. For example, histone deacetylase (HDAC) inhibitors may provide neuroprotection by downregulating KEAP1 (kelch-like ECH-associated protein 1), a binding suppressor of the transcription factor NRF2 (nuclear factor, erythroid 2 like 2; also known as NFE2L2), thereby inducing nuclear translocation of NRF2 and its downstream transcriptional program (Wang et al., 2012). Inhibition of the mammalian target of rapamycin (mTOR) has been known to profoundly affect gene expression by regulating the activity of a wide range of transcription factors, such as peroxisome proliferator-activated receptor alpha in the blockade of hepatic ketogenesis (Laplante and Sabatini, 2013). Moreover, the transcription factor NRF1 (also known as NFE2L1) was found to be required for the transcriptional activation of proteasome subunit genes upon proteasome inhibition in mammalian cells (Radhakrishnan et al., 2010). These recent findings further strengthen the hypothesis that gene-expression changes can, to some extent, mirror drug activities and therefore be used to inform the selection of drug combinations.

Here we address this demand by introducing a strategy that combines the gene-expression signatures between chemical perturbations and disease statuses to provide a testable hypothesis for combination therapies (Figure 1). Using more than 1.3 million publicly accessible perturbational gene-expression profiles obtained from the Library of Integrated Network-Based Cellular Signatures (LINCS) (Subramanian et al., 2017), we first generated hundreds of thousands of recurrent perturbation-transcript regulatory associations inferred between 3,332 chemical and 3,934 genetic perturbagens and 12,494 transcripts across

10 cell types. These perturbation-transcript regulatory relationships are noticeably distinct from known annotated compound-target pairs, but can still be recapitulated in independent perturbation datasets. Our analysis also reveals shared and divergent transcriptional modules across small-molecule mechanisms of action (MoAs), of which some modules are common to many MoAs that target cancer-specific hallmarks. Together with the relationships that correlate chemical sensitivity across many cancer cell lines with basal (unperturbed) gene expression (Rees et al., 2016), the connections between chemical perturbations and recurrently regulated transcripts can further expand our understanding of the molecular mechanisms associated with chemical sensitivity. On the basis of the identified perturbational gene-expression signatures, we predicted synergistic drug combinations for cancer, using an algorithm to search for therapeutics that can reverse as many non-overlapping disease transcripts as possible. The synergistic inhibitory effects of two drug combinations—CD-437 (a retinoid) combined with sirolimus (an mTOR inhibitor) and narciclasine (a protein synthesis inhibitor) combined with purvalanol A (a cyclin-dependent kinase [CDK] inhibitor)—were confirmed *in vitro*, indicating their potential as the general combination therapies for cancer. This first large-scale, unbiased, *in silico* approach will offer significant improvements in our ability to rationalize combination therapies for clinical success.

RESULTS

Recurrent Perturbation-Transcript Regulatory Relationships

We used publicly available large-scale chemical and genetic perturbation datasets for 10 cell types in LINCS, for which three perturbation types—exposure to small molecules or drugs for 6 h (denoted as d6), exposure to small molecules or drugs for 24 h (d24), and exposure to short hairpin RNAs (shRNAs) for 96 h (sh96)—comprised the majority of perturbation experiments (Table S1) (Huang et al., 2018). To investigate recurrent relationships between perturbations and regulated transcripts, we adapted a method and rank-based statistical score, the REC score (Jacobsen et al., 2013), to evaluate the recurrence of a given perturbation-transcript association across selected cell types (Transparent Methods). Among more than 100 million perturbation-transcript pairs (>10,000 chemical and genetic perturbations times 12,949 measured transcripts), we identified 77,691, 30,151, and 162,628 significantly recurring interactions for d6, d24, and sh96 perturbation types, respectively (Benjamini-Hochberg [BH]-corrected false discovery rate [FDR] < 0.001) (Figure 2 and Table S2). Several of these significantly recurring pairs were recovered in the analysis with resampling of 60% cell types, indicating their general robustness to variations in cell identity in our screen (Figure S1). We observed that whereas most perturbations tended to regulate a balanced number of genes in each direction (up- or downregulation), most genes exhibited a propensity to be either up- or downregulated by different effective perturbations in an exclusive manner (Figure 2A). For chemical perturbations, drugs or small molecules appeared to consistently regulate significant transcripts at both 6 and 24 h (Figure 2B). Of 3,190 genes targeted by the LINCS shRNA library and also measured for their expression, only 440 were recurrently downregulated by their corresponding "on-target" shRNAs, whose associated recurrent pairs accounted for 20.1% of all identified sh96 relationships (Figure 2C), probably in part due to the "off-target" activity (Jackson and Linsley, 2010) or other indirect responses 96 h after specific perturbations. Despite this observation, we found that the strongest "on-target" negative associations were captured in the most significantly recurring pairs (Figure 2D). Moreover, we note that some known small-molecule-regulated transcripts were captured by our analysis—for example, *KEAP1* downregulated by HDAC inhibitors (such as vorinostat and dacinostat) (Wang et al., 2012) and *ATF3* (activating transcription factor 3) upregulated by proteasome inhibitors (such as bortezomib and MLN-2238) (Zimmermann et al., 2000) (Table S2). These data substantiate our approach for identifying biologically meaningful relationships.

Comparison to Annotated Compound-Target Pairs

To assess the degree of correlation between small-molecule-regulated transcripts and their assigned protein targets, we compared the transcript-level recurrences for chemical perturbations with those compound-target annotations made by DrugBank (Figure S2 and Table S3) (Law et al., 2014). Of 11,022 annotated compound-target pairs, nearly 71% had no available REC scores, owing to the lack of either the target transcript or the matched compound for a given pair in LINCS (Figure 3A, as "not available"). The REC scores for an additional <10% pairs (<1,000) were not defined due to the insufficient number of cross-cell-type perturbation experiments per compound that is required for robust statistical inference (for a given compound-transcript pair, the REC score is defined only if the compound has been treated in at least half of the 10 cell lines analyzed in LINCS; Transparent Methods; Figure 3A, as "not defined"). Of the remaining compound-target pairs with correspondent REC scores (~2,500; Figure 3A, the first

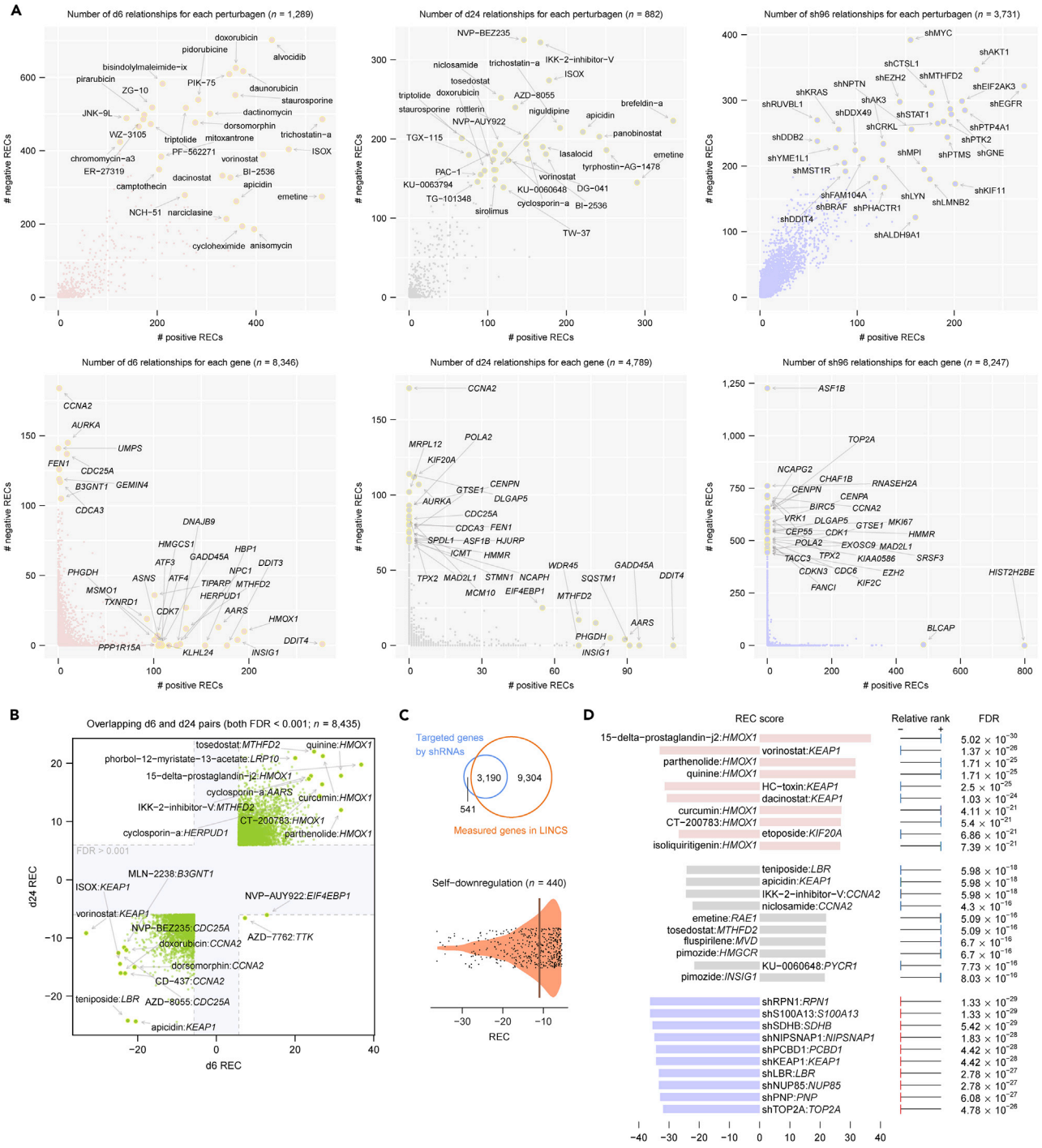


Figure 2. Recurrent Perturbation-Transcript Regulatory Relationships

(A) Summary of the recurrent relationships between the chemical and genetic perturbations and the regulated transcripts. The upper panels show the number of significant relationships per perturbation, whereas the lower panels show the same per gene. d6, small-molecule or drug treatment for 6 h; d24, small-molecule or drug treatment for 24 h; sh96, short hairpin RNA treatment for 96 h. REC, recurrence score.

(B) Comparison between d6 and d24 recurrent relationships. FDR, false discovery rate.

(C) On-target sh96 recurrent relationships. The REC scores for significant on-target pairs are shown in a violin plot; line, median.

(D) Top 10 significant d6, d24, and sh96 recurrent pairs.

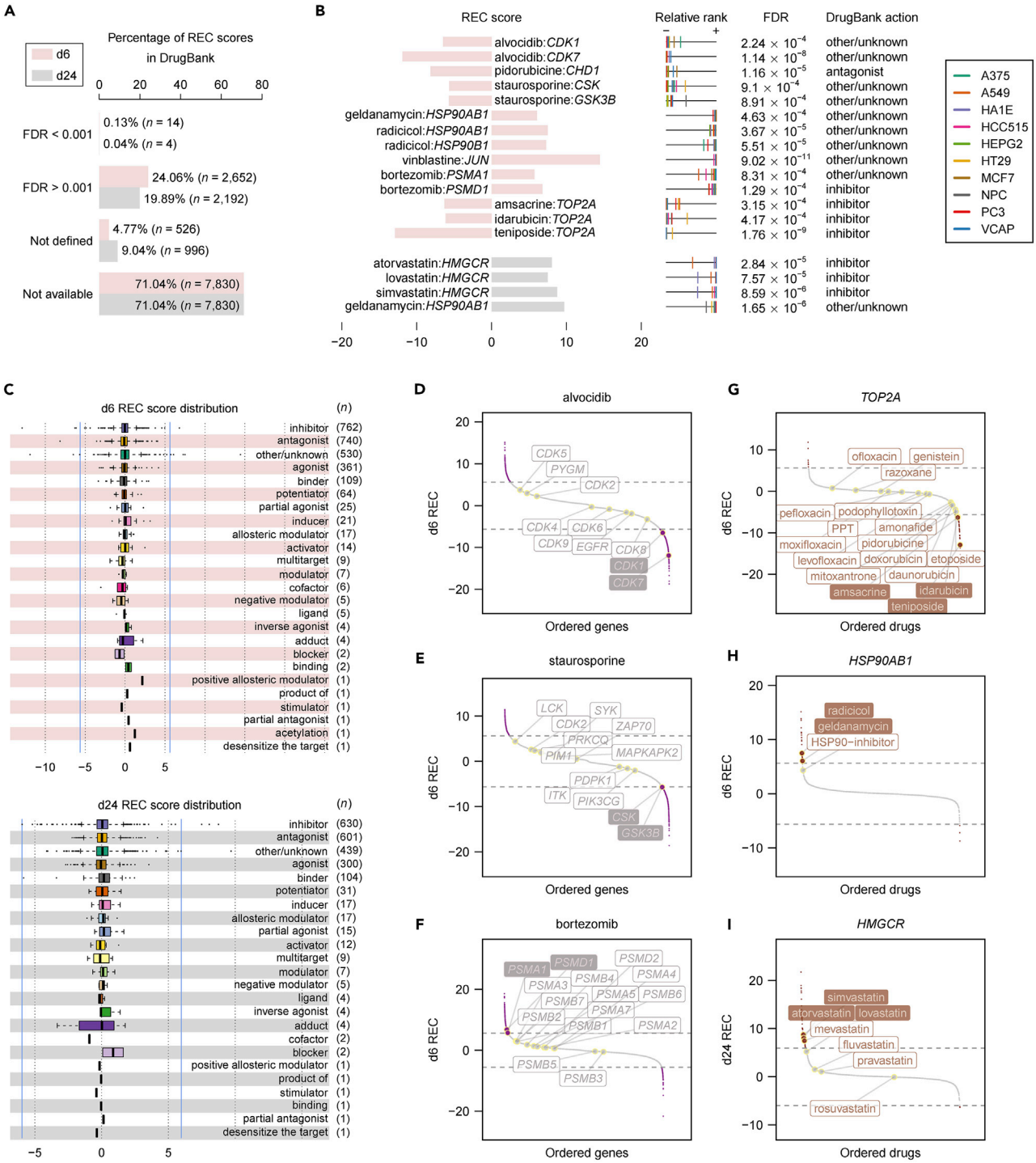


Figure 3. Comparison with Annotated Compound-Target Relationships

(A) Distribution of REC scores in DrugBank.

(B) The significant d6 or d24 pairs (FDR < 0.001) overlapping with DrugBank.

(C) Distribution of available REC scores for each compound action type. Box plots depict the interquartile range (IQR), and whiskers represent 1.5 × IQR. Blue vertical lines indicate the REC scores for which FDR = 0.001.

(D–I) Distribution of REC scores in LINCS for each of selected compounds (D–F) or genes (G–I) with at least one significant REC score overlapping with compound-target annotations. For each plot, all available DrugBank annotations are shown, and horizontal dashed lines indicate the REC scores for which FDR = 0.001.

two categories), only no more than 20 displayed statistically significant transcript-level recurrences (BH-corrected FDR < 0.001) (Figure 3B) and there was no apparent bias in the distribution of REC scores for each compound action type (for example, many outlier relationships for the "inhibitor" or "antagonist" and the "agonist" categories were readily found in both directions of regulation) (Figure 3C). For those compounds with at least one annotated target that exhibited a significant transcript-level recurrence (BH-corrected FDR < 0.001), not all other target transcripts for each compound were necessarily significantly regulated. For example, the pan-CDK inhibitor alvocidib seemed not to recurrently regulate all *CDK* transcripts at 6 h (only for *CDK1* and *CDK7*; Figure 3D), ditto for the promiscuous kinase inhibitor staurosporine (only for *CSK* and *GSK3B*; Figure 3E) and the proteasome inhibitor bortezomib (only for *PSMA1* and *PSMD1*; Figure 3F). However, we highlighted some intriguing observations from annotated targets shared by compounds that harbored at least one significantly recurring relationship toward the target transcript—topoisomerase inhibitors (teniposide, amsacrine, and idarubicin and, less significantly, mitoxantrone and etoposide) were most likely to downregulate *TOP2A* (DNA topoisomerase II alpha) (Figure 3G), heat shock protein 90 (HSP90) inhibitors (radicicol and geldanamycin) inclined to upregulate *HSP90AB1* (heat shock protein 90 alpha family class B member 1) (Figure 3H), and 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase (HMGCR) inhibitors (simvastatin, atorvastatin, and lovastatin and, less significantly, mevastatin, fluvastatin, and pravastatin) tended to upregulate *HMGCR* (Figure 3I). Overall, the transcript-level recurrences identified by our analysis represent completely distinct relationships from those known compound-target annotations.

Validating Transcript-Level Recurrences in Independent Datasets

We next sought to determine whether the recurrent perturbation-transcript interactions inferred with the LINCS resource can be reflected in independent datasets while generalizable to other cell lines. To this end, we performed differential gene-expression analysis in compound perturbed and unperturbed samples and looked at the overlap between these externally derived, differentially expressed (DE) genes and the corresponding transcript-level recurrences. A significant overlap of concordantly regulated DE (two-sided unpaired Student's t test, BH-corrected $p < 0.05$) and REC (BH-corrected FDR < 0.001 or 10^{-6}) genes for each direction was observed for many compound treatments in MCF7 cells, a human breast adenocarcinoma cell line included for evaluation of transcript-level recurrence (hypergeometric $p < 0.05$) (Figures 4A–4D and S3). In other two cell lines, the Ishikawa (human endometrial adenocarcinoma) and HepaRG (human hepatoma) cells, both of which were not used in this study, the patterns of perturbation-induced changes in gene expression also overlapped substantially with the inferred transcript-level recurrences (Figures 4E–4H and S3). This suggests that the recurring perturbation-transcript associations that we inferred from a few core cell lines are sufficiently characteristic of the unique compound actions that depend less on cell identity.

Regulatory Modules across Small-Molecule Classes

We next aimed at exploring the transcriptional regulatory landscape for drugs and small molecules. For each compound, we first manually annotated its primary MoA and searched for, if available, the associated World Health Organization Anatomical Therapeutic Chemical Classification codes (ATC codes) as the standard therapeutic indicator (Figure S4 and Table S1). For each class label (a primary MoA or an ATC code at any level), we calculated the proportion of hits of recurring regulation for each gene targeted by the tested compounds belonging to the class. This process has highlighted many transcriptional aspects of small-molecule-class actions that might be of therapeutic importance. For example, we found that many glucocorticoids (>10) can recurrently upregulate *NFKBIA* (nuclear factor-κB inhibitor alpha), consistent with the previously reported promoter activation by the glucocorticoid dexamethasone (Deroo and Archer, 2001), and *DDIT4* (DNA damage-inducible transcript 4), whose upregulation by dexamethasone has been shown to mediate the drug-induced autophagy in lymphocytes (Molitoris et al., 2011) (Figures 5A and S5). Several genes involved in cholesterol homeostasis, including *INSIG1* (insulin induced gene 1), *HMGCS1* (HMG-CoA synthase 1), *FDFT1* (farnesyl-diphosphate farnesyltransferase 1), *SQLE* (squalene epoxidase), *HMGCR*, and *LDLR* (low-density lipoprotein receptor), were found to be recurrently regulated by many antipsychotics (Figure S6), supporting the lipid disturbance associated with antipsychotic medications (Perez-Iglesias et al., 2009; Wiklund et al., 2010). In line with our observation (Figure 3H), HSP inhibitors are prone to transcriptionally upregulating several *HSP* family members, including *HSPH1*, *HSPA6*, *HSP90AB1*, *HSPD1*, *HSPE1*, *HSPB1*, and *HSP90AA1* (Figure S7).

To gain a global picture of small-molecule transcriptional responses, we focused on genes frequently targeted at the level of the primary MoA or ATC code separately (for a given class, proportion of hits ≥20%

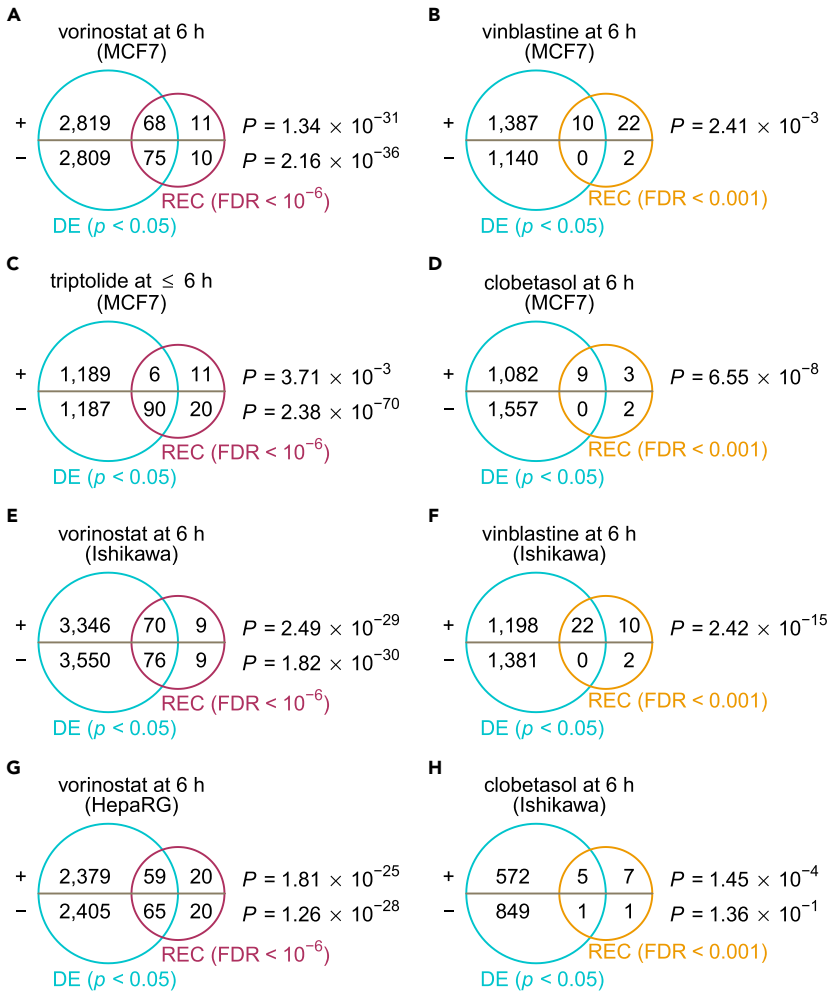


Figure 4. External Validation of Transcript-Level Recurrences

For each dataset corresponding to a given small-molecule perturbation in a given cell line for indicated time duration, the differentially expressed (DE) genes (two-sided unpaired Student's t test, Benjamini-Hochberg (BH)-corrected p < 0.05) were compared with corresponding transcript-level recurrences (FDR < 0.001 or 10^{-6}) for upregulated and downregulated genes separately using the hypergeometric test. (A–D) Validation of transcript-level recurrences for vorinostat (A), vinblastine (B), triptolide (C), or clobetasol (D) treated in MCF7 cells. (E) Validation of transcript-level recurrences for vorinostat treated in Ishikawa cells. (F) Validation of transcript-level recurrences for vinblastine treated in Ishikawa cells. (G) Validation of transcript-level recurrences for vorinostat treated in HepaRG cells. (H) Validation of transcript-level recurrences for clobetasol treated in Ishikawa cells. For detailed information about the comparison with different criteria, see Figure S3.

and consistent hits by ≥ 3 compounds). Hierarchical clustering of these frequently targeted transcripts revealed 226 gene modules across 50 primary MoAs (Figure 5B), covering 140 gene singletons and three gene clusters harboring ubiquitously regulated genes among primary MoAs (Figure 5C and Table S4). Of these gene modules, some displayed specific enrichments for cancer hallmarks. For example, one ubiquitous cluster of 63 frequently downregulated genes by several cytotoxic drugs and kinase inhibitors (ucl1) was enriched for mitotic cell cycle regulation (E2F targets, BH-corrected hypergeometric $p = 4.15 \times 10^{-15}$; G2/M checkpoint, $p = 1.95 \times 10^{-13}$; and mitotic spindle, p = 0.0167), whereas another cluster of 15 ubiquitously upregulated genes (ucl2) was enriched for stress phenotypes (mTOR complex 1 [mTORC1] signaling, $p = 1.76 \times 10^{-6}$; p53 pathway, $p = 4.21 \times 10^{-8}$; hypoxia, p = 0.00174; apoptosis, $p = 6.72 \times 10^{-4}$; and unfolded protein response, p = 0.00505) (Figure 5D). Using the same criteria for analysis on ATC categories, we, nevertheless, found the limited sharing of gene modules across therapeutic areas (Figure S8 and Table S5), owing to the relatively less emphasis on small-molecule action in the ATC classification system. This allowed us to distinguish extensively targeted transcripts across different ATC classes, for example, *NFKBIA* and *DDIT4*, which were prone to being downregulated by drugs approved

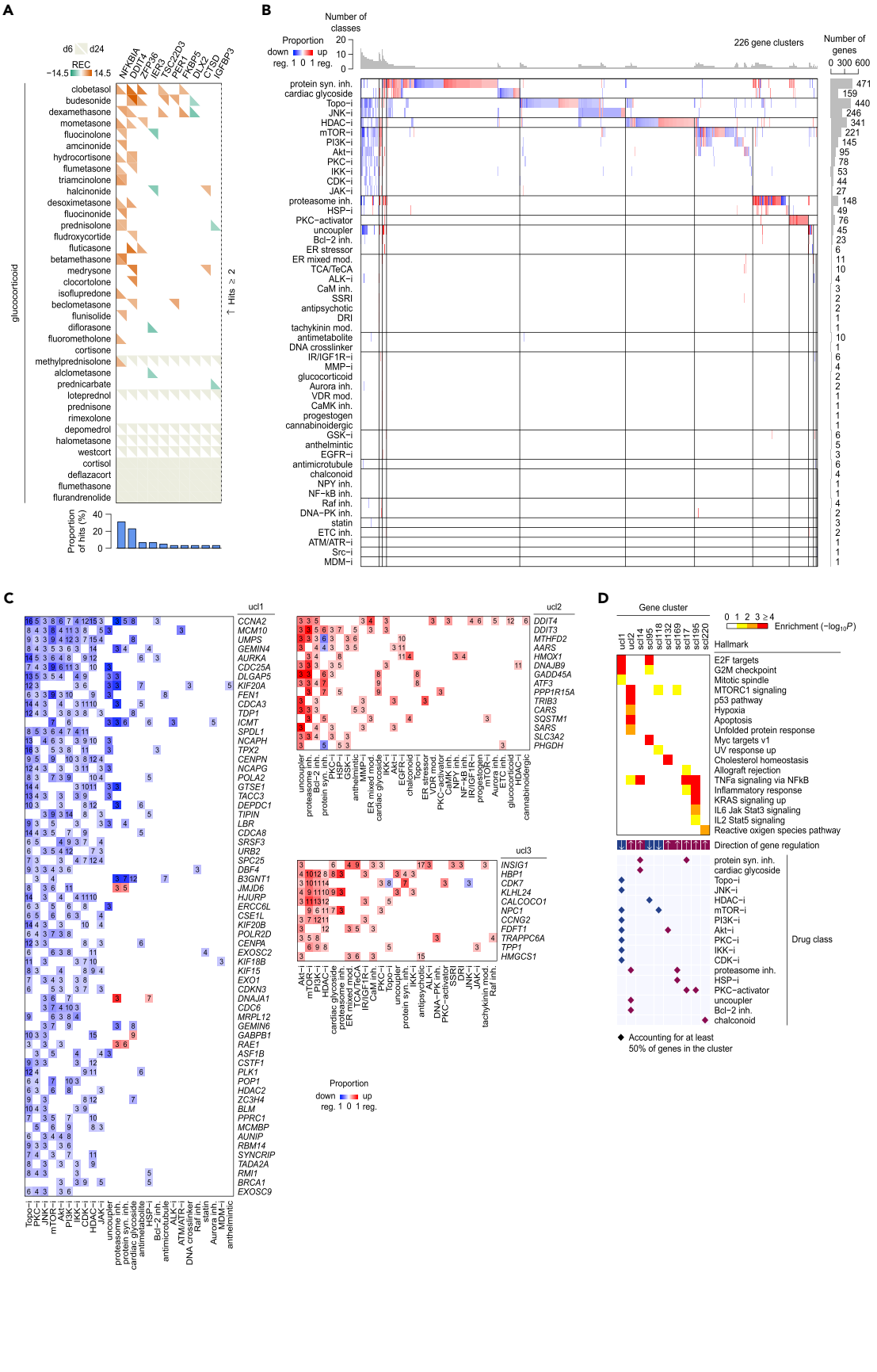

A
d6
d24
REC
−14.5
14.5
NFKBIA
DDIT4
ZFP36
IER3
TSC22D3
PER1
FKBP5
DUSP1
KLF2
CTGF
IGFBP3
glucocorticoid
clobetasol
budesonide
dexamethasone
mometasone
fluocinolone
amcinonide
hydrocortisone
flumetasone
triamcinolone
halcinonide
desoximetasone
fluocinonide
prednisolone
fludroxycortide
fluticasone
betamethasone
medrysone
clocortolone
isoflupredone
beclometasone
flunisolide
diflorasone
fluorometholone
cortisone
methylprednisolone
alclometasone
prednicarbate
loteprednol
prednisone
rimexolone
depomedrol
halometasone
westcort
cortisol
deflazacort
flumethasone
flurandrenolide
↑ Hits ≥ 2
Proportion of hits (%)
40
20
0
B
Number of classes
Proportion
down
up
reg. 1 0 1 reg.
20
10
0
226 gene clusters
Number of genes
0 300 600
protein syn. inh.
cardiac glycoside
Topo−i
JNK−i
HDAC−i
mTOR−i
PI3K−i
Akt−i
PKC−i
IKK−i
CDK−i
JAK−i
proteasome inh.
HSP−i
PKC−activator
uncoupler
Bcl−2 inh.
ER stressor
ER mixed mod.
TCA/TeCA
ALK−i
CaM inh.
SSRI
antipsychotic
DRI
tachykinin mod.
antimetabolite
DNA crosslinker
IR/IGF1R−i
MMP−i
glucocorticoid
Aurora inh.
VDR mod.
CaMK inh.
progestogen
cannabinoidergic
GSK−i
anthelmintic
EGFR−i
antimicrotubule
chalconoid
NPY inh.
NF−kB inh.
Raf inh.
DNA−PK inh.
statin
ETC inh.
ATM/ATR−i
Src−i
MDM−i
471
159
440
246
341
221
145
95
78
53
44
27
148
49
76
45
23
6
11
10
1
6
4
3
2
2
1
1
10
1
6
4
2
2
1
1
1
1
6
5
3
6
4
1
1
4
2
3
2
1
1
C
ucl1
ucl2
ucl3
CCNA2
MCM10
UMPS
GEMIN4
AURKA
CDC25A
DLGAP5
KIF20A
FEN1
CDCA3
TDP1
ICMT
SPDL1
NCAPH
TPX2
CENPN
NCAPG
POLA2
GTSE1
TACC3
DEPDC1
TIPIN
LBR
CDCA8
SRSF3
URB2
SPC25
DBF4
B3GNT1
JMJD6
HJURP
ERCC6L
CSE1L
KIF20B
POLR2D
CENPA
EXOSC2
KIF18B
KIF15
EXO1
CDKN3
DNAJA1
CDC6
MRPL12
GEMIN6
GABPB1
RAE1
ASF1B
CSTF1
PLK1
POP1
HDAC2
ZC3H4
BLM
PPRC1
MCMBP
AUNIP
RBM14
SYNCRIP
TADA2A
RMI1
BRCA1
EXOSC9
DDIT4
DDIT3
MTHFD2
AARS
HMOX1
DNAJB9
GADD45A
ATF3
PPP1R15A
TRIB3
CARS
SQSTM1
SARS
SLC3A2
PHGDH
INSIG1
HBP1
CDK7
KLHL24
CALCOCO1
NPC1
CCNG2
FDFT1
TRAPPC6A
TPP1
HMGCS1
Topo−i
PKC−i
IKK−i
mTOR−i
Akt−i
PI3K−i
HDAC−i
CDK−i
uncoupler
proteasome inh.
protein syn. inh.
cardiac glycoside
antimetabolite
HSP−i
Bcl−2 inh.
antimicrotubule
ALK−i
ATM/ATR−i
DNA crosslinker
Raf inh.
statin
Aurora inh.
MDM−i
anthelmintic
Proportion
down
up
reg. 1 0 1 reg.
D
Gene cluster
ucl1
ucl2
scl14
scl95
scl118
scl132
scl169
scl17
scl195
scl220
0 1 2 3 ≥4
Enrichment (−log10P)
Hallmark
E2F targets
G2M checkpoint
Mitotic spindle
MTORC1 signaling
p53 pathway
Hypoxia
Apoptosis
Unfolded protein response
Myc targets v1
UV response up
Cholesterol homeostasis
Allograft rejection
TNFa signaling via NFkB
Inflammatory response
KRAS signaling up
IL6 Jak Stat3 signaling
IL2 Stat5 signaling
Reactive oxigen species pathway
Direction of gene regulation
Drug class
protein syn. inh.
cardiac glycoside
Topo−i
JNK−i
HDAC−i
mTOR−i
PI3K−i
Akt−i
PKC−i
IKK−i
CDK−i
proteasome inh.
HSP−i
PKC−activator
uncoupler
Bcl−2 inh.
chalconoid
◆ Accounting for at least
50% of genes in the cluster

Figure 5. Regulatory Gene Modules across Diverse Drug Classes

(A) Recurrently regulated genes by glucocorticoids.

(B) Transcriptional modules across small-molecule primary MoAs. For detailed descriptions of primary MoAs, see Table S1.

(C) Three ubiquitous gene clusters identified in (B). For each small-molecule-class hit, the number of within-class small molecules accounting for the hit is shown.

(D) Enrichments of gene regulatory modules for cancer hallmarks.

in the dermatological (category D), alimentary (category A), respiratory (category R), and sensory (category S) systems (Figure S8C). Overall, through our systematic and unbiased analysis, these small-molecule transcriptional modules can advance further studies to identify novel compound action.

Correlation with Small-Molecule Sensitivity

We next studied the relationships between compound-induced transcriptional responses and chemical sensitivity patterns generated by correlating an area under the dose-response curve metric for each of 481 tested compounds with 18,543 basal transcripts across 823 cancer cell lines (Rees et al., 2016). By overlapping significant perturbation-transcript regulatory interactions (BH-corrected FDR <0.001) and sensitivity-expression associations (two-sided Bonferroni-corrected $p < 0.05$), we identified 9,413 compound-transcript connections (6,431 for d6 plus 2,982 for d24) that could help enhance the molecular interpretation of small-molecule action (Figure 6A and Table S6). For example, we found that high expression of *TIMELESS*, a circadian regulator whose overexpression has been implicated in various cancer types (Mao et al., 2013), was strongly correlated with sensitivity to the HDAC inhibitor vorinostat, which, in turn, can recurrently downregulate *TIMELESS* after the treatment (Figures S9A and S9B). In contrast, low expression of *BAG3* (BCL2-associated athanogene 3) (Rosati et al., 2011) was correlated with sensitivity to the proteasome inhibitor bortezomib that can upregulate *BAG3* after the treatment (Figures S9C and S9D), consistent with those observed in leukemia cells where *BAG3* knockdown restored sensitivity to bortezomib (Liu et al., 2009). To gain a broader insight into the relationships between expression-sensitivity correlation and reciprocal gene regulation, we extended this analysis to the level of compound MoA (Figure 6B and Table S7). Interestingly, SYNCRIP (synaptotagmin-binding cytoplasmic RNA-interacting protein), an RNA-binding protein recently shown to be required for the development of leukemia, was strongly correlated with sensitivity to topoisomerase and HDAC inhibitors, both of which can recurrently downregulate *SYNCRIP* (Figure 6B) (Vu et al., 2017). Overall, these data have enabled us to explore molecular mechanisms associated with small-molecule sensitivity for the treatment of disease.

Combinatorial Drug Discovery for Cancer and Experimental Validation

We hypothesized that the perturbation-transcript regulatory recurrences could be exploited to predict synergistic drug interactions by correlating the combined patterns of perturbation-regulated recurring transcripts with a given disease gene signature. Specifically, we developed an algorithm that examines the extent of gene reversal achieved by a combination of small molecules with individual transcript-level recurrences in reference to the disease signature, in which the combination is called synergistic if it reverses a significant proportion of disease transcripts in a nearly non-overlapping manner (Figure S10, Transparent Methods). The therapeutic score by this approach provides a rough interpretation of gene reversal such that suppose there are 1,000 disease genes and a given perturbation can reverse exactly 100 genes with REC scores for which all FDRs are 0.001, the therapeutic score is 0.1. To address this possibility, we used gene-expression signatures between primary tumors and tumor-adjacent normal tissues among eight cancer types from The Cancer Genome Atlas (TCGA) to inform pan-cancer synergistic drug combinations (RNA sequencing reads with Bonferroni-corrected $p < 0.05$, $|\log_2(\text{fold change})| > 1$, and $\log_2(\text{count per million}) > 3$) (Table S8) (Aran et al., 2017). Using our expression-based approach, we first identified ~300 single agents that could significantly reverse disease gene signatures across many cancer types, including several chemotherapeutics and investigational drugs that target cancer-related pathways (BH-corrected $p < 0.05$, by permutation analysis; 377 of 2,628 tested single agents had at least one significant therapeutic score) (Figure 7A and Table S9). Based on the patterns of gene reversal for these individual therapeutics ($n = 218$ and 162 for d6 and d24 perturbations, respectively, corresponding to a union of 288 compounds with at least one therapeutic score ≥ 0.01 with BH-corrected $p < 0.05$), we then predicted synergistic drug pairs for cancer in general, most of which were combinations of a topoisomerase inhibitor and an mTOR or phosphoinositide 3-kinase (PI3K) inhibitor (BH-corrected $p < 0.05$, by permutation analysis; 20,050 of 32,508 tested combinations had at least one significant therapeutic score) (Figures 7B and 7C and Table S10), consistent with the reported synergistic toxicity in some cancer cells (Babichev et al., 2016; Itamochi et al., 2011).

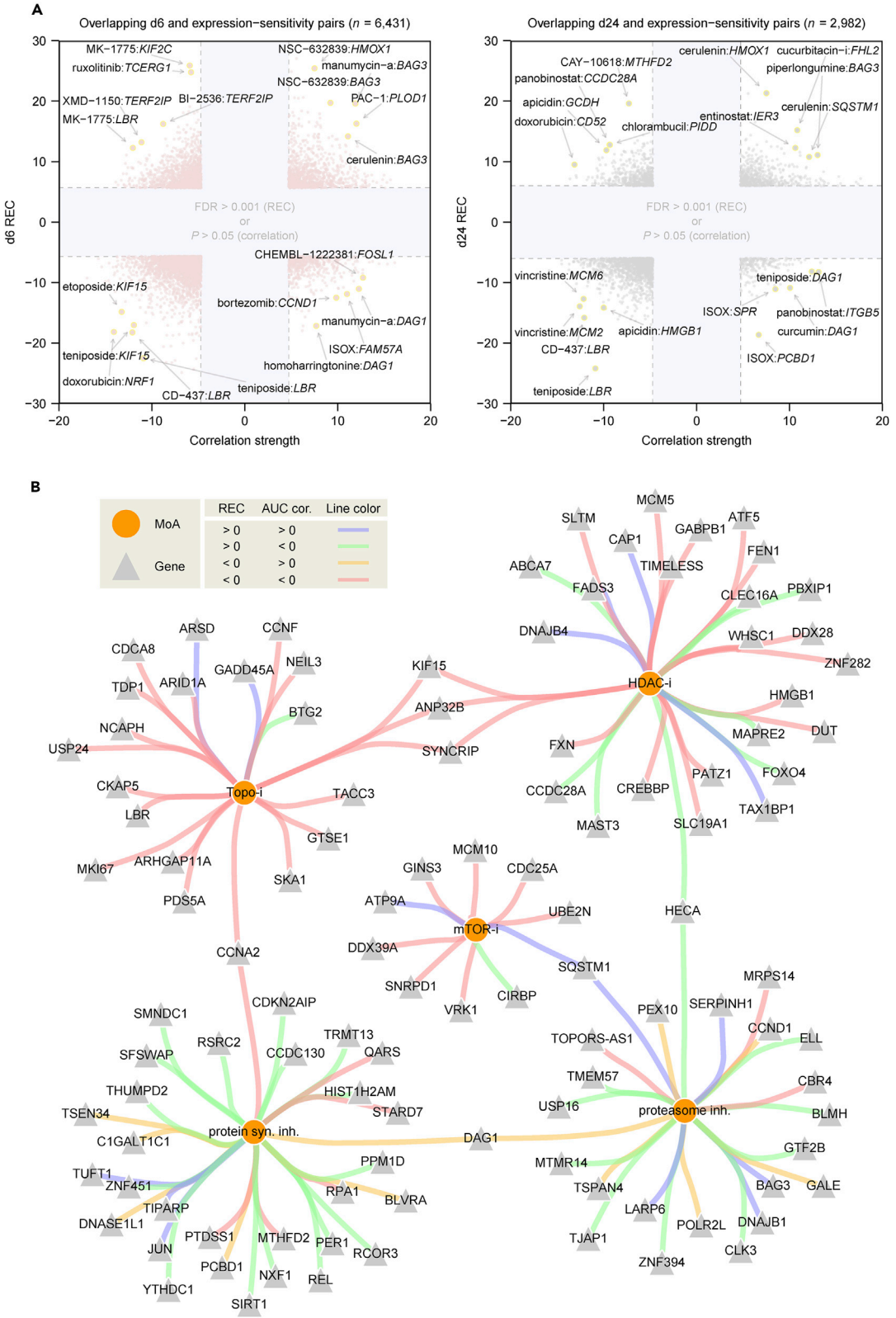

A
Overlapping d6 and expression−sensitivity pairs (n = 6,431)
MK−1775:KIF2C
ruxolitinib:TCERG1
XMD−1150:TERF2IP
BI−2536:TERF2IP
MK−1775:LBR
NSC−632839:HMOX1
manumycin−a:BAG3
NSC−632839:BAG3
PAC−1:PLOD1
cerulenin:BAG3
FDR > 0.001 (REC)
or
P > 0.05 (correlation)
CHEMBL−1222381:FOSL1
etoposide:KIF15
bortezomib:CCND1
manumycin−a:DAG1
ISOX:FAM57A
homoharringtonine:DAG1
teniposide:KIF15
doxorubicin:NRF1
CD−437:LBR
teniposide:LBR
d6 REC
Correlation strength
Overlapping d24 and expression−sensitivity pairs (n = 2,982)
CAY−10618:MTHFD2
panobinostat:CCDC28A
apicidin:GCDH
doxorubicin:CD52
chlorambucil:PIDD
cerulenin:HMOX1
cucurbitacin−i:FHL2
piperlongumine:BAG3
entinostat:IER3
cerulenin:SQSTM1
FDR > 0.001 (REC)
or
P > 0.05 (correlation)
vincristine:MCM6
vincristine:MCM2
CD−437:LBR
teniposide:LBR
apicidin:HMGB1
teniposide:DAG1
ISOX:SPR
panobinostat:ITGB5
curcumin:DAG1
ISOX:PCBD1
d24 REC
Correlation strength
B
MoA
Gene
REC
AUC cor.
Line color
> 0 > 0
> 0 < 0
< 0 > 0
< 0 < 0
HDAC-i
Topo-i
mTOR-i
protein syn. inh.
proteasome inh.
MCM5
SLTM
CAP1
GABPB1
ATF5
FEN1
ABCA7
TIMELESS
FADS3
CLEC16A
PBXIP1
DNAJB4
WHSC1
DDX28
ZNF282
HMGB1
MAPRE2
DUT
FXN
FOXO4
PATZ1
CCDC28A
CREBBP
TAX1BP1
MAST3
SLC19A1
ARSD
CCNF
CDCA8
GADD45A
NEIL3
TDP1
ARID1A
BTG2
NCAPH
USP24
CKAP5
LBR
TACC3
GTSE1
MKI67
ARHGAP11A
SKA1
PDS5A
KIF15
ANP32B
SYNCRIP
MCM10
GINS3
CDC25A
ATP9A
UBE2N
HECA
DDX39A
CCNA2
SQSTM1
SNRPD1
CIRBP
VRK1
CDKN2AIP
SMNDC1
RSRC2
TRMT13
SFSWAP
CCDC130
QARS
THUMPD2
HIST1H2AM
TSEN34
STARD7
C1GALT1C1
DAG1
TUFT1
PPM1D
ZNF451
RPA1
BLVRA
TIPARP
DNASE1L1
PTDSS1
MTHFD2
PER1
RCOR3
JUN
PCBD1
NXF1
REL
YTHDC1
SIRT1
MRPS14
PEX10
SERPINH1
CCND1
TOPORS-AS1
ELL
TMEM57
CBR4
USP16
BLMH
GTF2B
MTMR14
TSPAN4
BAG3
GALE
LARP6
POLR2L
DNAJB1
TJAP1
CLK3
ZNF394

Figure 6. Comparison with Small-Molecule Sensitivity Data

(A) Overlapping transcript-perturbagen (REC scores with FDR <0.001, identified by our analysis; y axis) and expression-sensitivity (Z-scored area under the dose-response-curve correlation strength with Bonferroni-corrected two-sided $p < 0.05$, identified by Rees et al (Rees et al., 2016); x axis) relationships. See Table S6.

(B) Overlapping relationships at the level of primary MoA. Shown are significant MoA-transcript overlapping pairs with occurrences ≥ 3 and the proportion $\geq 50\%$, or simply with occurrences ≥ 8, as defined in Table S7.

Among the top 15 predicted synergistic drug pairs, we selected two unexpected interactions—a retinoid CD-437 in combination with an mTOR inhibitor sirolimus and a protein synthesis inhibitor narciclasine in combination with a CDK inhibitor purvalanol A—for testing in tumor-type-matched breast (BRCA; $n = 2$) and lung (LUAD; $n = 2$) cancer cell lines (Figure 7B). To evaluate drug synergy, we used the delta score, which has been shown to exhibit superior performance over other commonly used drug interaction scores, for example, based on the Loewe additivity or Bliss independence model (Yadav et al., 2015). The delta score is symmetric, ranging from −1 (antagonism) through 0 (additive effect) to +1 (synergy), and also biologically meaningful, which can inform the additional effect produced over an expected drug interaction. For example, a delta score of 0.1 indicates 10% of additional effects of growth inhibition in a cytotoxic assay for a given drug combination, whereas a delta score of −0.1 indicates antagonism with 10% effects below the expected interaction. For known synergistic drug combinations, high delta scores of 0.03 or more over the dose-response matrices were confirmed by Yadav et al (Yadav et al., 2015). Of eight tested conditions in this study (two drug combinations for four cell lines), five revealed synergistic antiproliferative effects (average delta score ≥ 0.04; Figures 7D and S11). The observed interaction patterns of synergy were also reproducible across biological replicates (Figure S12). For the most synergistic case (CD-437 combined with sirolimus in H1299 cells), the average and maximum delta scores (0.14 and 0.54, respectively) were comparable with those of the top synergistic interaction revealed by Yadav et al. (0.176 and 0.40, respectively). We also note that the degree of synergy varied across sensitive cell lines for each drug combination, consistent with cell-type-specific responses to most anticancer drugs (Niepel et al., 2017).

The concept of targeting non-oncogene dependencies based on gene expression is being actively explored for combinatorial drug discovery. A recent analysis of breast cancer cell lines and molecularly targeted agents reveals that drug synergy may arise from compounds not only between different gene-expression clusters but also within the same cluster (Niepel et al., 2017). With compound screen data, it seems feasible that predictive drug-response gene biomarkers may translate to synergistic drug combinations. Jiang et al. have developed a computational method to predict gene interactions of drug response and show that several genes are frequently associated with resistance or sensitivity to multiple targeted drugs (Jiang et al., 2018). For example, their approach identified *AXL* to be a resistance gene for the ALK inhibitor NVP-TAE684, consistent with our observation that low expression of *AXL* correlated with sensitivity to NVP-TAE684, which can in turn downregulate *AXL* after treatment (Table S6). However, these predictive methodologies are limited in their ability to provide a general solution to identifying efficacious drug combinations at a sufficiently large scale.

Recent studies are revealing the potential to inform new drugs through their ability to invert the non-oncogene dependency program. It has been demonstrated that reversal of cancer gene expression positively correlates, although not outstandingly, with drug efficacy (Chen et al., 2017). This has led to the identification of the top four compounds that have not been previously reported for liver cancer—a cardiac glycoside strophanthidin, a mitochondrial uncoupler FCCP, an ATM/ATR kinase inhibitor CGK-733, and an anthelmintic drug pyrvinium pamoate—all of which were effectively recapitulated by our approach in liver cancer (Table S9). Indeed, the compounds predicted for liver cancer in this study were highly enriched in the top compound list generated by Chen et al. ($p \approx 0$, by enrichment analysis, Figure S13) (Chen et al., 2017), further verifying our approach for identifying effective drug therapies. More recently, a precision oncology method has been proposed to identify compounds that could achieve the global inversion of master regulator protein activities in gastroenteropancreatic neuroendocrine tumors (GEP-NETs), a rare malignancy originating in the pancreas, small bowel, and rectum (Alvarez et al., 2018). Of 107 compounds that were validated *in vitro* to be differentially active in GEP-NET-related cells, 68 were included in our inference analysis of transcript-level recurrences. Of these, 63% (43 of 68) achieved significant reversal of at least one cancer signature in this study, including entinostat (an HDAC inhibitor, which was selected for validation *in vivo* by Alvarez et al.), CD-437, alvocidib, bortezomib, and GDC-0941 (also known as pictilisib, a PI3K inhibitor) (Figures 7A and 7B and Table S11). This validates that our gene-expression-based approach is able to identify small-molecule compounds that could effectively target non-oncogene dependencies. However, these current methodologies still cannot be extended to combinatorial drug discovery,

Figure 7. Predicted Synergistic Drug Combinations across Cancer Types

(A) Expression-based single therapeutics for cancers with corresponding median cell line sensitivity available from the Cancer Therapeutics Response Portal (CTRP). Shown are small molecules with at least one significant therapeutic score ≥0.05 (BH-corrected $p < 0.05$) across cancers types. For full results, see Table S9.

(B) Top 15 synergistic drug combinations across cancer types. For full results, see Table S10.

(C) Occurrences of the combinations of small-molecule classes in top 100.

(D) Validation of two predicted synergistic drug pairs in breast and lung cancer cells using clonogenic assays. Drug synergy was evaluated using the delta score (Yadav et al., 2015). For full results, see Figures S11 and S12.

owing to the shared limitation of all gene-expression-based similarity approaches to connect more than one perturbation (query) to a given disease of interest (reference) at a time. Inspired by the fact that a given perturbation should only affect a restricted panel of genes (Felix and Barkoulas, 2015), our approach leverages small-molecule-regulated recurring transcripts that more closely represent the unique compound mechanisms independent of cell identity, which further enables us to predict effective drug pairs by comparing the compound mechanisms mirrored in their gene-expression signatures. Together, these data demonstrate the feasibility of using transcript-level recurrences for unbiased data-driven prediction of synergistic drug interactions.

DISCUSSION

Our study presents a systematic computational approach for combinatorial drug discovery by correlating the patterns of gene reversal in a given disease signature achieved by small-molecule-regulated recurring transcripts. These transcript-level recurrences inferred within the LINCS resource are reproducible enough in external datasets to represent the unique small-molecule action and are clearly distinct from annotated compound-target relationships. In-depth analysis revealed that some small-molecule-class transcriptional modules can regulate common hallmarks and stress phenotypes of cancer and that some recurrently regulated transcripts are associated with chemical and small-molecule-class sensitivity that warrants further exploration.

Although gene expression provides a basis for predicting drug synergy, the extent of gene reversal by single agents does not necessarily correlate with chemical sensitivity, suggesting that cellular processes to which cancer cells are addicted may shape the response to cancer therapies. In addition, although we predicted synergistic drug combinations using "bulk" patient gene-expression profiles, cell-line-intrinsic characteristics (for example, coding mutations that alter actionable sites for drug binding or other mechanisms related to drug inactivation) may also influence the evaluation of drug synergy in our experimental setting. However, the presented approach can only capture some synergistic drug pairs that exhibited nearly non-overlapping patterns of gene reversal and it is possible for two drugs without significant reversal effects to display synergistic killing. Interestingly, this approach may also be combined with a drug similarity network approach to expand the repertoire of potential synergistic drug pairs using co-clustering relationships (Huang et al., 2018), as suggested by a recent study showing that synergy can be observed with pairs of drugs that elicit both similar and distinct transcriptional responses (Niepel et al., 2017).

To predict effective drug combinations for cancer in this study, we assumed that all disease transcripts are equally important except their differential expression values. However, the predicting algorithm can be improved by adjusting the weights of disease transcripts (for example, by integrative analysis of the cancer transcriptome to narrow down the genes essential for maintaining the tumorigenic state) or by imposing additional constraints that require some specific transcripts to be targeted. The need for more refined cancer signatures was reflected on the generally low therapeutic scores (of around 0.1) in this study, indicating that only a small proportion of genes in a given signature (roughly 10% for a therapeutic score of 0.1) can be reversed by small-molecule treatments. Alternatively, our gene-expression-based approach provides a rational framework on which to combine standard cancer therapies (chemotherapeutic agents, molecularly targeted drugs, or immunotherapies) with a drug that would be able to reverse the signature of an emerging resistance mechanism, which may relate to the acquired addiction of cancer cells to non-mutated genes that are hard to decipher by genetic analysis (Flemming, 2015).

In addition to being placed in the context of a global reversal of a disease signature for predicting synergistic drug interactions, these transcript-level recurrences can also be integrated with small-molecule sensitivity to develop a combinatorial strategy for treating diseases. For example, in cases where low expression of a given transcript correlates with sensitivity to a small molecule or a small-molecule class that can in turn upregulate the transcript after treatment (i.e., the relationships in the first quadrant in each panel of Figure 6A and those connected with blue edges in the network of Figure 6B), a strategy that combines the small-molecule treatment with the inhibition of the correlated transcript could augment the small-molecule sensitivity in cancer cells harboring high level of the transcript (for example, the bortezomib:*BAG3* pair for which *BAG3* knockdown sensitized leukemia cells to bortezomib-induced apoptosis; Liu et al., 2009). Conversely, for pairs for which a given small-molecule treatment is able to downregulate a target transcript whose expression positively correlates with the small-molecule sensitivity (i.e., the relationships in the third quadrant in each panel of Figure 6A and those connected with red edges in the network of Figure 6B), enforced overexpression of the target transcript may enhance the cytotoxic effects in small-molecule-sensitive cell lines.

To benchmark our inferred transcript-level regulatory recurrences, we used independent perturbation profiles to gauge the reproducibility while, at the same time, using annotated compound-target relationships to unveil a disparate regulatory nature of the direct target proteins and the corresponding transcripts. We observed examples in which a protein inhibitor may either upregulate (HSP inhibitors versus *HSP* family members) or downregulate (topoisomerase inhibitors versus *TOP2A*) the expression of the corresponding transcript and also found evidence of no regulatory bias in target transcripts for each small-molecule-binding type. In contrast to the ability of expression-sensitivity correlations made between unperturbed gene-expression data and small-molecule sensitivity profiles across cell lines to identify some direct target connections (Rees et al., 2016), our transcript-level regulatory recurrences, in a way, represent small-molecule mechanisms at the level of treatment-induced transcriptional changes other than direct target associations. Moreover, these transcript-level recurrences could be integrated with hallmark network models to map the small-molecule mechanisms onto cancer cell vulnerabilities and predict effective drug treatments (Wang et al., 2015; Zaman et al., 2013). The transcriptional modules identified in our analysis further bring to light several unexplored relationships that might be of potential therapeutic relevance (Molitoris et al., 2011; Perez-Iglesias et al., 2009; Wiklund et al., 2010).

We also note that the cell line identities have a great influence on the composition of the inferred transcript-level regulatory recurrences. For example, recurrent findings might be favored in situations in which the cell lines used for inference can respond similarly to a given perturbation, but less likely in situations when most of the cell lines lack the expression of a transcript relevant to the mechanism of a given perturbation. Consistent with this notion, the compounds with the strongest transcript-level regulatory relationships inferred in this study were less enriched for some molecularly targeted drugs that are thought to be effective in only a subset of cancer cells harboring the relevant gene mutations or pathway activities, such as the receptor tyrosine kinase epidermal growth factor receptor or HER2 inhibitors lapatinib, neratinib, and erlotinib (Figure 1A and Table S2). This is in contrast to those studies where targeted therapies are the main focus and tested in responsive cells (Jiang et al., 2018; Niepel et al., 2017). Despite this concern, the recurring regulatory relationships identified in this study can adequately reproduce some small-molecule transcriptional mechanisms in a separate panel of cell lines. Nonetheless, there are many circumstances in which this inference method can be applied—for example, in a desired subset of immune cell types to identify therapeutic opportunities for manipulating the immune function or in selected cell lines to develop effective combination therapies against a specific cancer type or subtype for which some molecularly targeted drugs are deemed more likely to induce similar transcriptional responses.

Importantly, our approach has potential clinical applicability to guide future combination therapy in individual patients with cancer. To this end, a tumor sample and its matched normal tissue for each patient are required to generate a cancer gene-expression signature that represents the transition from normal to malignant state, thus enabling the prediction of drug combinations that can effectively reverse this transition in a patient-specific manner. Given the observed variation of gene expression within human tissues (GTEx Consortium, 2015), it is often necessary for our approach to use patient-specific tumor-adjacent normal tissues rather than a pooled reference from patients with the same tumor type. In this study, we chose to predict pan-cancer drugs by comparison between pooled tumors and normal tissues for each cancer type, owing to the general paucity of the matched normal samples in TCGA. However, *in vitro* testing of the top predictions in randomly chosen tumor-type-matched lines verified our proof-of-concept analysis, indicating that these top prioritized drug combinations may be effective in a large fraction of patients with cancer.

In summary, we establish a systems-level computational strategy through gene signature analysis to inform testable predictions for effective combination therapies. The small-molecule-regulated recurring transcripts identified here might not only be particularly useful for targeting non-oncogene addiction from which a resistance mechanism arises but also enable further investigation of the transcriptional mechanisms that underlie small-molecule sensitivity and side effects. Our findings provide a resource for future research on small-molecule mechanisms and will pave the way for more rational combination therapies in clinical trials.

Limitations of the Study

For precision oncology, this approach requires a tumor sample (or a cancer cell line) and its matched normal tissue of each patient (or a lineage-matched untransformed cell line with similar genetic background) to generate a cancer gene-expression signature that represents the transition from the normal

to malignant state, which enables the prediction of drug combinations that can effectively reverse this transition. Therefore it might not be appropriate to use a pooled reference from those tumor-adjacent normal tissues of patients with the same tumor type (or a pooled reference from all other cancer cell lines in pan-cancer analysis). In addition, the quality of a disease gene-expression signature will affect the ability of this approach to identify effective therapies.

METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.04.039.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Technology (MOST 105-2320-B-002-057-MY3 and MOST 106-2320-B-002-053-MY3) and the National Health Research Institutes (NHRI-EX107-10530PI and NHRI-EX107-10709BI) in Taiwan.

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. and Y.-H.C. performed the drug combination 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.

Received: September 13, 2018
Revised: February 2, 2019
Accepted: April 29, 2019
Published: May 31, 2019

REFERENCES

Al-Lazikani, B., Banerji, U., and Workman, P. (2012). Combinatorial drug therapy for cancer in the post-genomic era. Nat. Biotechnol. *30*, 679–692.

Alvarez, M.J., Subramaniam, P.S., Tang, L.H., Grunn, A., Aburi, M., Rieckhof, G., Komissarova, E.V., Hagan, E.A., Bodei, L., Clemons, P.A., et al. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nat. Genet. *50*, 979–989.

Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M., and Butte, A.J. (2017). Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat. Commun. *8*, 1077.

Babichev, Y., Kabaroff, L., Datti, A., Uehling, D., Isaac, M., Al-Awar, R., Prakesch, M., Sun, R.X., Boutros, P.C., Venier, R., et al. (2016). PI3K/AKT/mTOR inhibition in combination with doxorubicin is an effective therapy for leiomyosarcoma. J. Transl. Med. *14*, 67.

Bertolini, F., Sukhatme, V.P., and Bouche, G. (2015). Drug repurposing in oncology–patient and health systems opportunities. Nat. Rev. Clin. Oncol. *12*, 732–742.

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.

Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., Johnston, S.E., Vrcic, A., Wong, B., Khan, M., et al. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. Nat. Med. *23*, 405–408.

Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. *15*, 81–94.

Deroo, B.J., and Archer, T.K. (2001). Glucocorticoid receptor activation of the I kappa B alpha promoter within chromatin. Mol. Biol. Cell *12*, 3365–3374.

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.

Flemming, A. (2015). Anticancer drugs: finding the perfect combination. Nat. Rev. Drug Discov. *14*, 13.

GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hata, A.N., Niederst, M.J., Archibald, H.L., Gomez-Caraballo, M., Siddiqui, F.M., Mulvey, H.E., Maruvka, Y.E., Ji, F., Bhang, H.E., Krishnamurthy Radhakrishna, V., et al. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat. Med. *22*, 262–269.

Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. Nat. Rev. Cancer *13*, 714–726.

Huang, C.T., Hsieh, C.H., Oyang, Y.J., Huang, H.C., and Juan, H.F. (2018). A large-scale gene expression intensity-based similarity metric for drug repositioning. iScience *7*, 40–52.

Itamochi, H., Oishi, T., Shimada, M., Sato, S., Uegaki, K., Naniwa, J., Sato, S., Nonaka, M., Terakawa, N., Kigawa, J., et al. (2011). Inhibiting the mTOR pathway synergistically enhances cytotoxicity in ovarian cancer cells induced by etoposide through upregulation of c-Jun. Clin. Cancer Res. *17*, 4742–4750.

Jackson, A.L., and Linsley, P.S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat. Rev. Drug Discov. *9*, 57–67.

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.

Jia, J., Zhu, F., Ma, X., Cao, Z., Cao, Z.W., Li, Y., Li, Y.X., and Chen, Y.Z. (2009). Mechanisms of drug combinations: interaction and network perspectives. Nat. Rev. Drug Discov. *8*, 111–128.

Jiang, P., Lee, W., Li, X., Johnson, C., Liu, J.S., Brown, M., Aster, J.C., and Liu, X.S. (2018). Genome-scale signatures of gene interaction from compound screens predict clinical efficacy of targeted cancer therapies. Cell Syst. *6*, 343–354.e5.

Kwong, L.N., Heffernan, T.P., and Chin, L. (2013). A systems biology approach to personalizing therapeutic combinations. Cancer Discov. *3*, 1339–1344.

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.

Laplante, M., and Sabatini, D.M. (2013). Regulation of mTORC1 and its impact on gene expression at a glance. J. Cell Sci. *126*, 1713–1719.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. *42*, D1091–D1097.

Liu, P., Xu, B., Li, J., and Lu, H. (2009). BAG3 gene silencing sensitizes leukemic cells to Bortezomib-induced apoptosis. FEBS Lett. *583*, 401–406.

Lopez, J.S., and Banerji, U. (2017). Combine and conquer: challenges for targeted therapy combinations in early phase trials. Nat. Rev. Clin. Oncol. *14*, 57–66.

Luo, J., Solimini, N.L., and Elledge, S.J. (2009). Principles of cancer therapy: oncogene and non-oncogene addiction. Cell *136*, 823–837.

Mao, Y., Fu, A., Leaderer, D., Zheng, T., Chen, K., and Zhu, Y. (2013). Potential cancer-related role of circadian gene TIMELESS suggested by expression profiling and in vitro analyses. BMC Cancer *13*, 498.

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.

Molitoris, J.K., McColl, K.S., Swerdlow, S., Matsuyama, M., Lam, M., Finkel, T.H., Matsuyama, S., and Distelhorst, C.W. (2011). Glucocorticoid elevation of dexamethasone-induced gene 2 (Dig2/RTP801/REDD1) protein mediates autophagy in lymphocytes. J. Biol. Chem. *286*, 30181–30189.

Nagel, R., Semenova, E.A., and Berns, A. (2016). Drugging the addict: non-oncogene addiction as a target for cancer therapy. EMBO Rep. *17*, 1516–1531.

Niepel, M., Hafner, M., Duan, Q., Wang, Z., Paull, E.O., Chung, M., Lu, X., Stuart, J.M., Golub, T.R., Subramanian, A., et al. (2017). Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling. Nat. Commun. *8*, 1186.

Perez-Iglesias, R., Mata, I., Pelayo-Teran, J.M., Amado, J.A., Garcia-Unzueta, M.T., Berja, A., Martinez-Garcia, O., Vazquez-Barquero, J.L., and Crespo-Facorro, B. (2009). Glucose and lipid disturbances after 1 year of antipsychotic treatment in a drug-naive population. Schizophr. Res. *107*, 115–121.

Radhakrishnan, S.K., Lee, C.S., Young, P., Beskow, A., Chan, J.Y., and Deshaies, R.J. (2010). Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. Mol. Cell *38*, 17–28.

Ramirez, M., Rajaram, S., Steininger, R.J., Osipchuk, D., Roth, M.A., Morinishi, L.S., Evans, L., Ji, W., Hsu, C.H., Thurley, K., et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat. Commun. *7*, 10690.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

Rosati, A., Graziano, V., De Laurenzi, V., Pascale, M., and Turco, M.C. (2011). BAG3: a multifaceted protein that regulates major cell pathways. Cell Death Dis. *2*, e141.

Ryall, K.A., and Tan, A.C. (2015). Systems biology approaches for advancing the discovery of effective drug combinations. J. Cheminform. *7*, 7.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. Nat. Rev. Drug Discov. *16*, 19–34.

Saunders, N.A., Simpson, F., Thompson, E.W., Hill, M.M., Endo-Munoz, L., Leggatt, G., Minchin, R.F., and Guminski, A. (2012). Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. EMBO Mol. Med. *4*, 675–684.

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.

Sun, X., Vilar, S., and Tatonetti, N.P. (2013). High-throughput methods for combinatorial drug discovery. Sci. Transl. Med. *5*, 205rv201.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Vu, L.P., Prieto, C., Amin, E.M., Chhangawala, S., Krivtsov, A., Calvo-Vidal, M.N., Chou, T., Chow, A., Minuesa, G., Park, S.M., et al. (2017). Functional screen of MSI2 interactors identifies an essential role for SYNCRIP in myeloid leukemia stem cells. Nat. Genet. *49*, 866–875.

Wang, B., Zhu, X., Kim, Y., Li, J., Huang, S., Saleem, S., Li, R.C., Xu, Y., Dore, S., and Cao, W. (2012). Histone deacetylase inhibition activates transcription factor Nrf2 and protects against cerebral ischemic damage. Free Radic. Biol. Med. *52*, 928–936.

Wang, E., Zaman, N., McGee, S., Milanese, J.S., Masoudi-Nejad, A., and O'Connor-McCourt, M. (2015). Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. Semin. Cancer Biol. *30*, 4–12.

Waring, M.J., Arrowsmith, J., Leach, A.R., Leeson, P.D., Mandrell, S., Owen, R.M., Pairaudeau, G., Pennie, W.D., Pickett, S.D., Wang, J., et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat. Rev. Drug Discov. *14*, 475–486.

Webster, R.M. (2016). Combination therapies in oncology. Nat. Rev. Drug Discov. *15*, 81–82.

Wiklund, E.D., Catts, V.S., Catts, S.V., Ng, T.F., Whitaker, N.J., Brown, A.J., and Lutze-Mann, L.H. (2010). Cytotoxic effects of antipsychotic drugs implicate cholesterol homeostasis as a novel chemotherapeutic target. Int. J. Cancer *126*, 28–40.

Yadav, B., Wennerberg, K., Aittokallio, T., and Tang, J. (2015). Searching for drug synergy in complex dose-response landscapes using an interaction potency model. Comput. Struct. Biotechnol. J. *13*, 504–513.

Zaman, N., Li, L., Jaramillo, M.L., Sun, Z., Tibiche, C., Banville, M., Collins, C., Trifiro, M., Paliouras, M., Nantel, A., et al. (2013). Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. Cell Rep. *5*, 216–223.

Zimmermann, J., Erdmann, D., Lalande, I., Grossenbacher, R., Noorani, M., and Furst, P. (2000). Proteasome inhibitor induced gene expression profiles reveal overexpression of transcriptional regulators ATF3, GADD153 and MAD1. Oncogene *19*, 2913–2920.

ISCI, Volume *15*

Supplemental Information

Perturbational Gene-Expression Signatures for Combinatorial Drug Discovery

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

SUPPLEMENTAL FIGURES

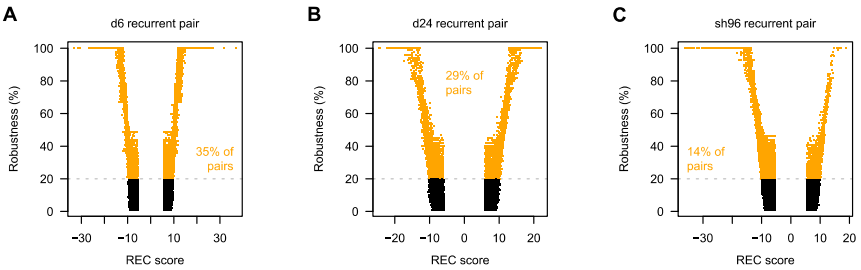


Figure S1. Resampling analysis for the recurrent perturbation–transcript pairs, related to Figure 2.

For recurrent perturbation–transcript 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 REC scores (i.e., those from the analysis when 100% of cell types were used; *x*-axis). For each plot, the robustness at 20% recovery is indicated by a horizontal dash gray line, above which the proportion of pairs (orange) is shown.

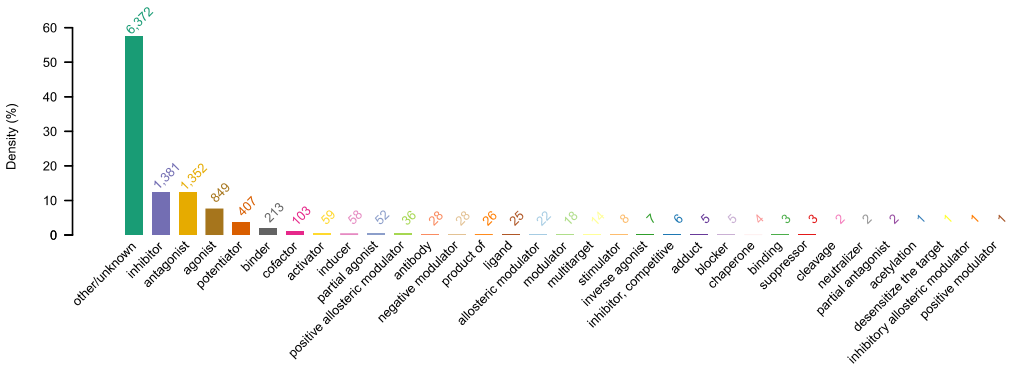


Figure S2. Distribution of small-molecule binding action types in DrugBank, related to Figure 3.
The number of compound–target annotations for each binding action type is shown above the bar.

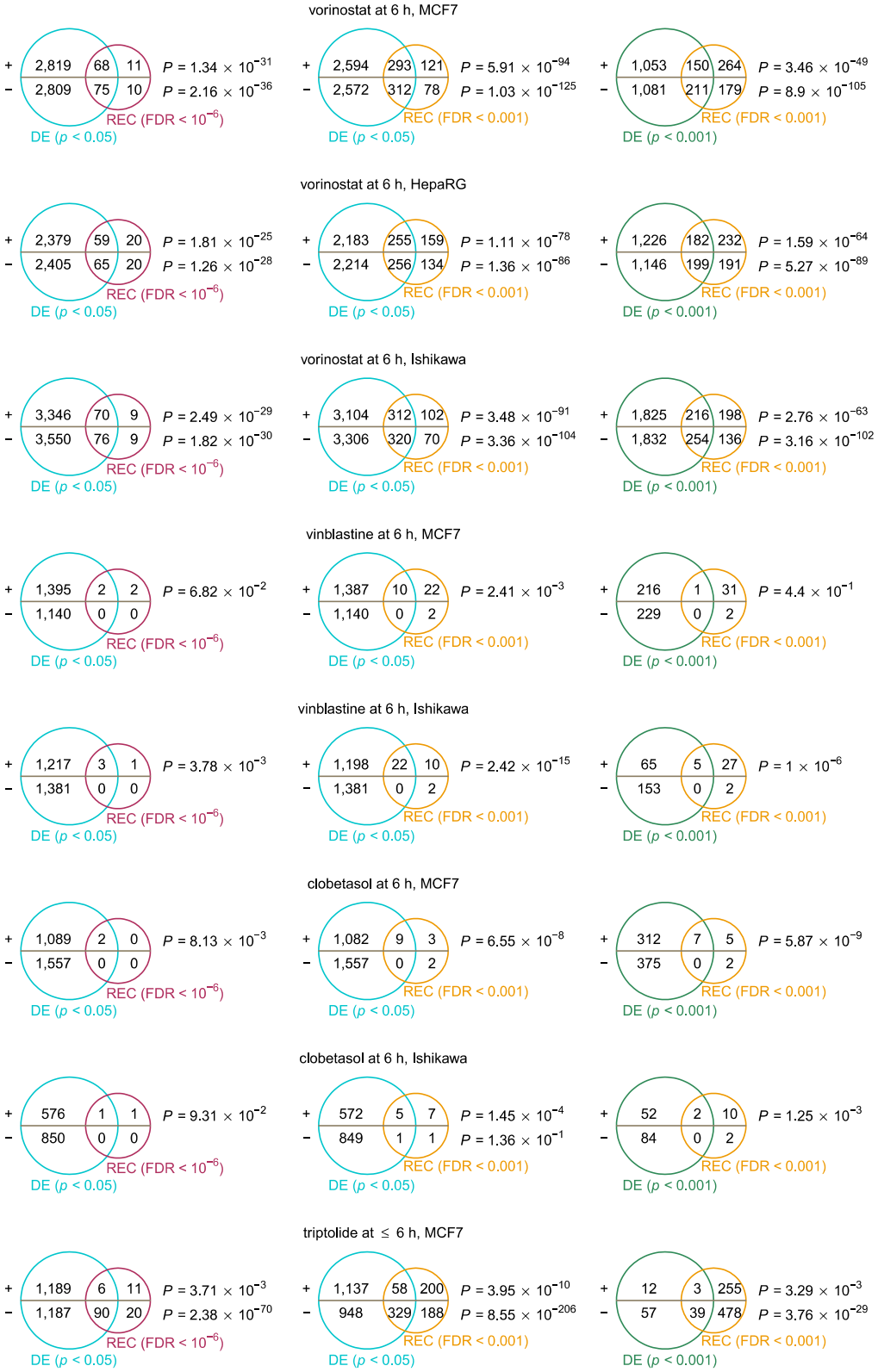


Figure S3. Comparison of transcript-level recurrences with independently derived perturbation profiles, related to Figure 4. For each perturbation dataset, the extent of overlap between differentially expressed (DE) genes (two-sided unpaired Student's *t*-test, Benjamini–Hochberg (BH)-corrected $P < 0.05$ or 0.001) and corresponding transcript-level recurrences (FDR < 0.001 or 10^{-6}) were assessed by the hypergeometric test for upregulated and downregulated genes separately. Figure 4 shows the representative results from this part.

Figure S4. Occurrences of primary MoAs of the compounds analyzed in LINCS, related to Figure 5.
For detailed descriptions of primary MoAs, see Table S1.

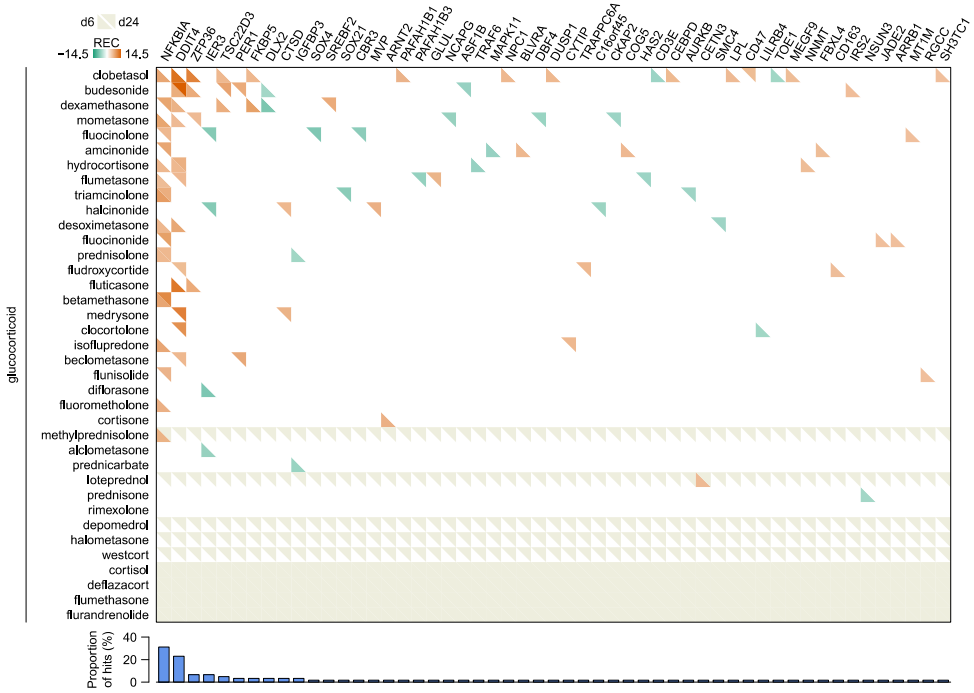


Figure S5. Recurrently regulated genes by glucocorticoids, related to Figure 5.

Shown in Figure 5A are genes with more than one hit by the tested compounds belonging to this class.

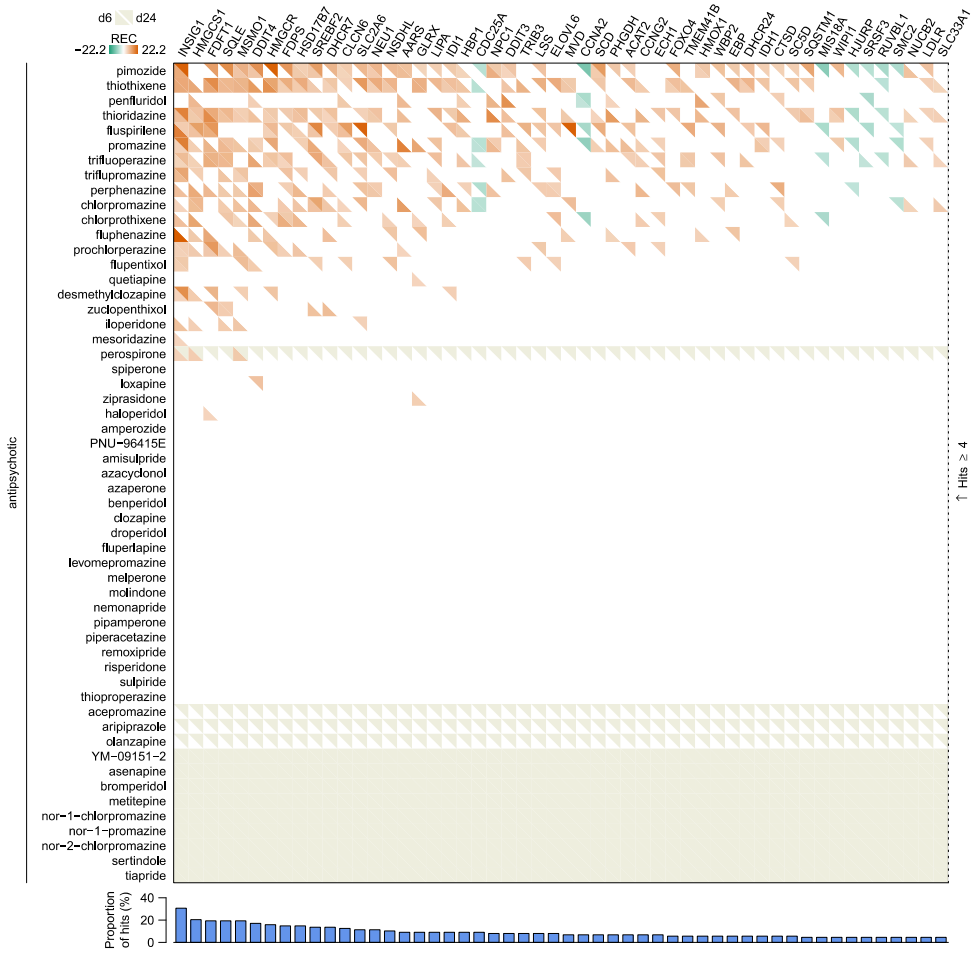


Figure S6. Recurrently regulated genes by antipsychotics, related to Figure 5.
Shown are genes with at least four recurrent hits by the tested compounds belonging to this class.

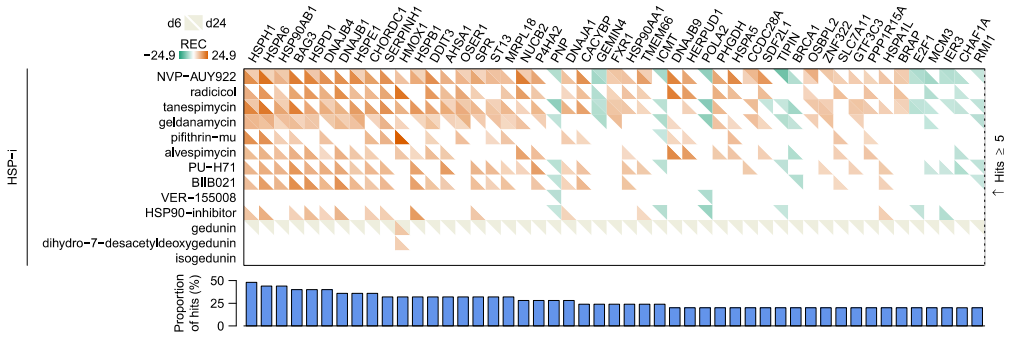


Figure S7. Recurrently regulated genes by heat shock protein (HSP) inhibitors, related to Figure 5.
Shown are genes with at least five recurrent hits by the tested compounds belonging to this class.

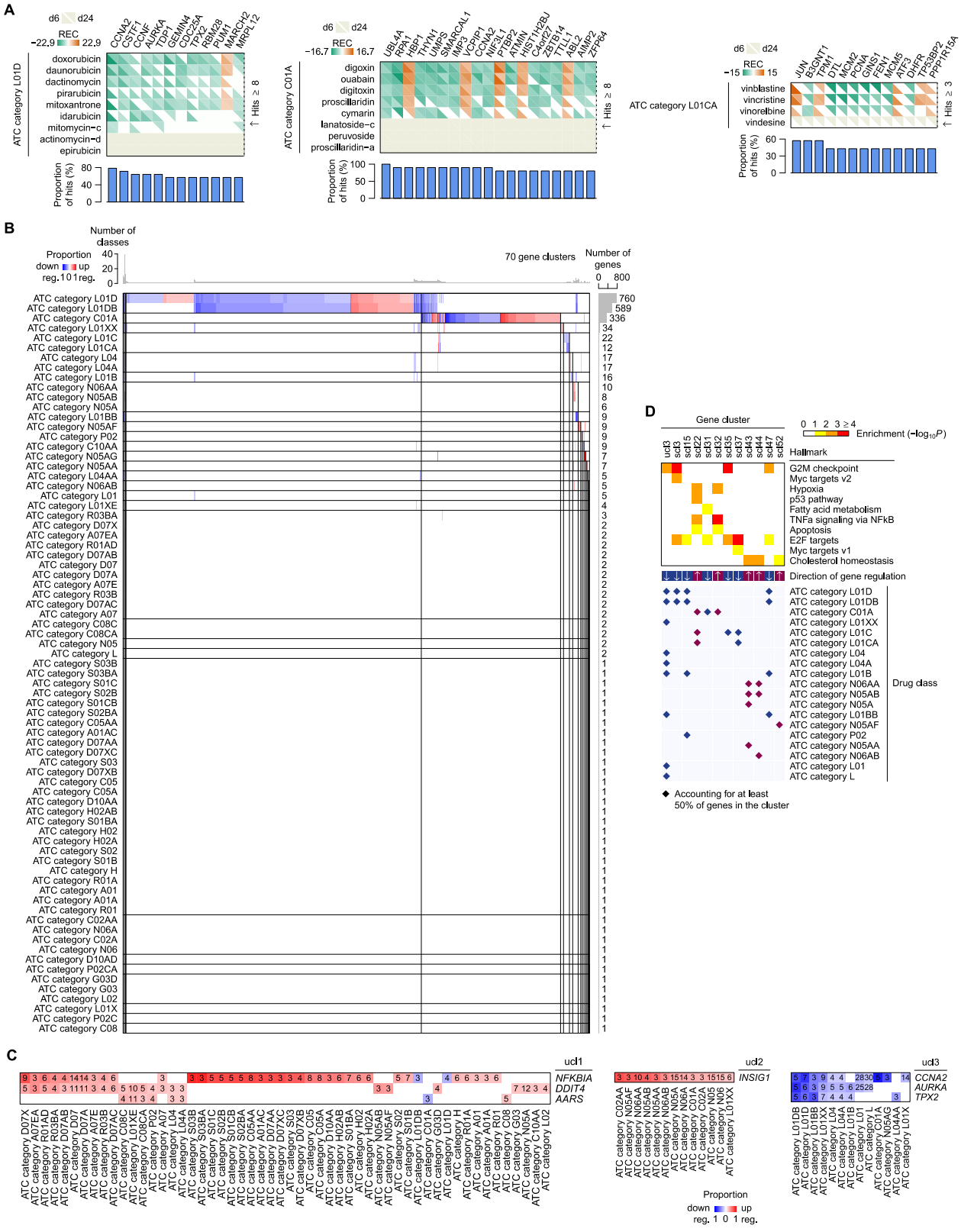


Figure S8. Regulatory gene modules across therapeutic areas, related to Figure 5.

(A) Examples of recurrently regulated genes by approved drugs belonging to the WHO Anatomical Therapeutic Chemical Classification System (ATC) codes L01D (cytotoxic antibiotics and related substances; left panel), C01A (cardiac glycosides; middle panel), or L01CA (vinca alkaloids and analogues; right panel). For detailed descriptions of ATC codes, see Table S1.

(B) Transcriptional modules across ATC categories.

(C) Ubiquitous gene clusters identified in (B). For each ATC-class hit, the number of within-class drugs accounting for the hit is shown.

(D) Enrichments of gene regulatory modules for cancer hallmarks.

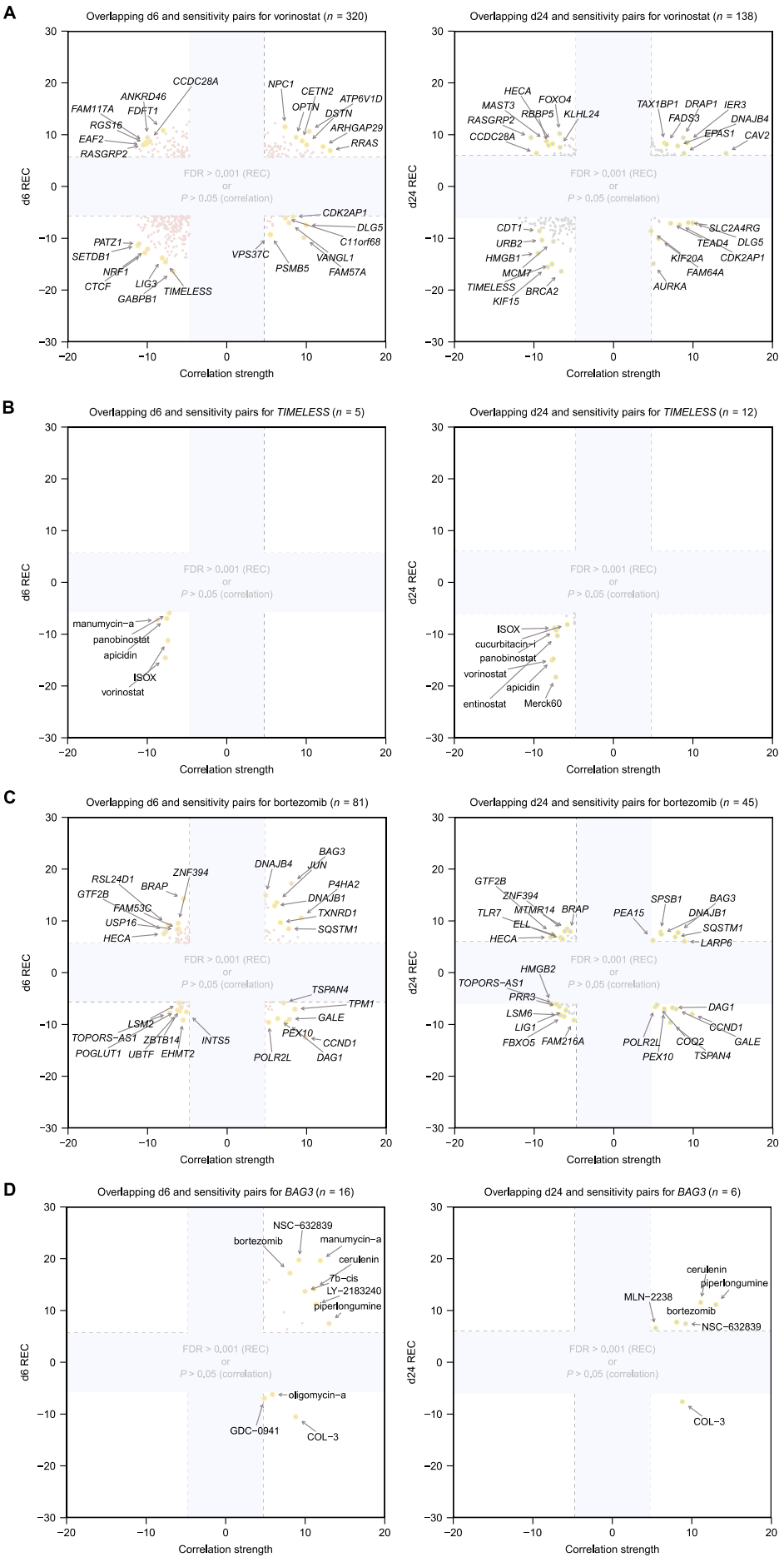


Figure S9. Overlapping compound–transcript and sensitivity–expression relationships for selected compounds and transcripts, related to Figure 6. Shown are overlapping relationships for the compounds vorinostat (A) and bortezomib (C) and the transcripts *TIMELESS* (B) and *BAG3* (D). For each plot, up to seven relationships are indicated in each quadrant.

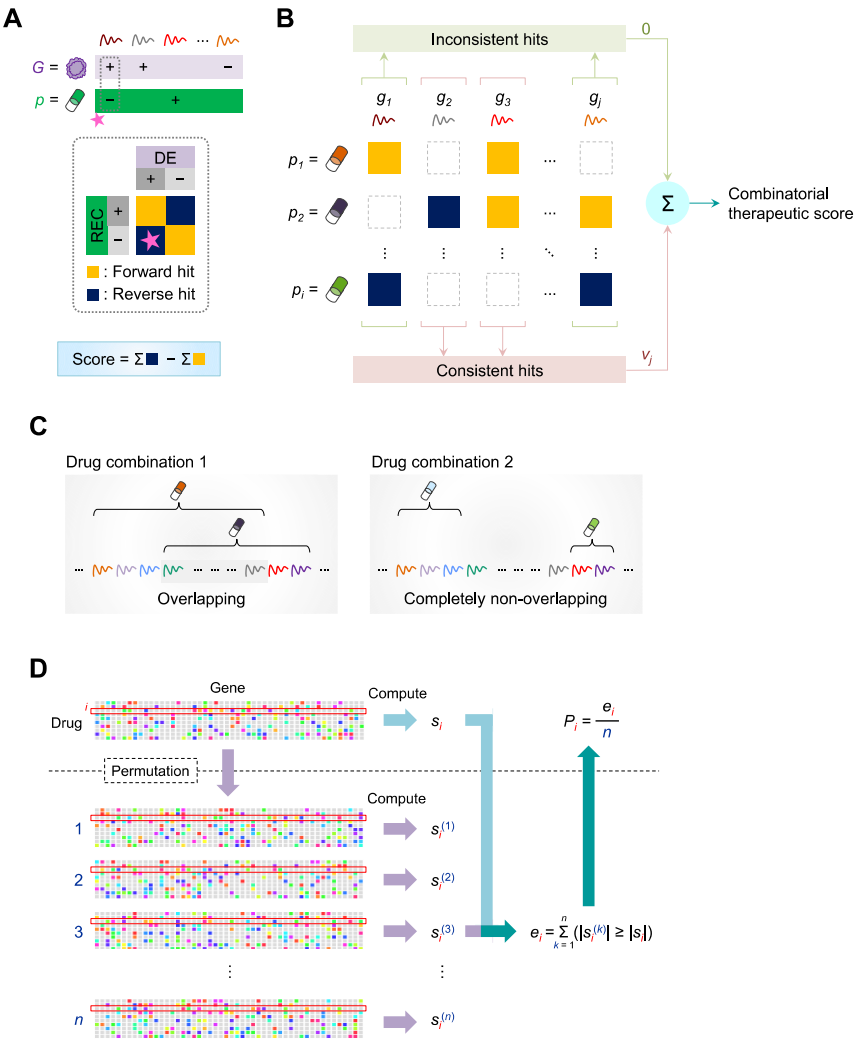


Figure S10. Schematic of our approach for combinatorial drug discovery, related to Figure 7.

(A) Expression-based therapeutic score for a perturbation *p* given a disease signature *G*. The pink star icon indicates an instance of reverse hit.

(B) Combinatorial therapeutic score.

(C) Illustrative examples of an overlapping (left) and a completely non-overlapping (right) pattern of gene reversal achieved by two different drug combinations.

(D) Empirical *P*-value for an expression-based therapeutic. For detailed formal descriptions, see Methods.

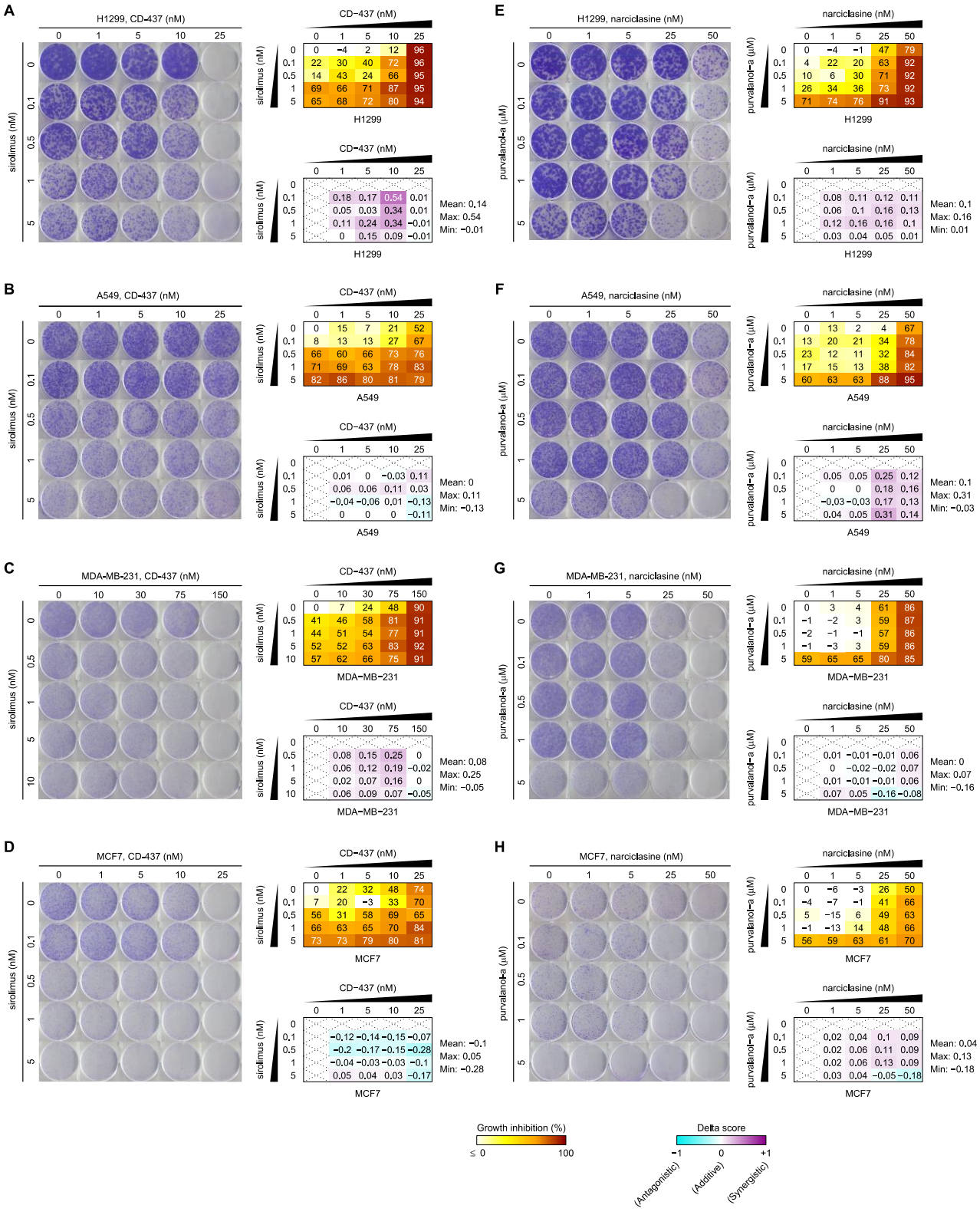


Figure S11. Validation of the selected drug combinations in cancer cell lines, related to Figure 7.
Clonogenic assays were performed for the combination of CD-437 and sirolimus (A–D) and the combination of narciclasine and purvalanol A (E–H) in H1299 (lung cancer; A, E), A549 (lung cancer; B, F), MDA-MB-231 (breast cancer; C, G), and MCF7 (breast cancer; D, H) cells. For each tested condition, an image from a representative experiment (left), percentage of cell growth inhibition (top right), and delta scores (as synergistic indices; bottom right) are shown. Figure 7D shows the representative results from this part.

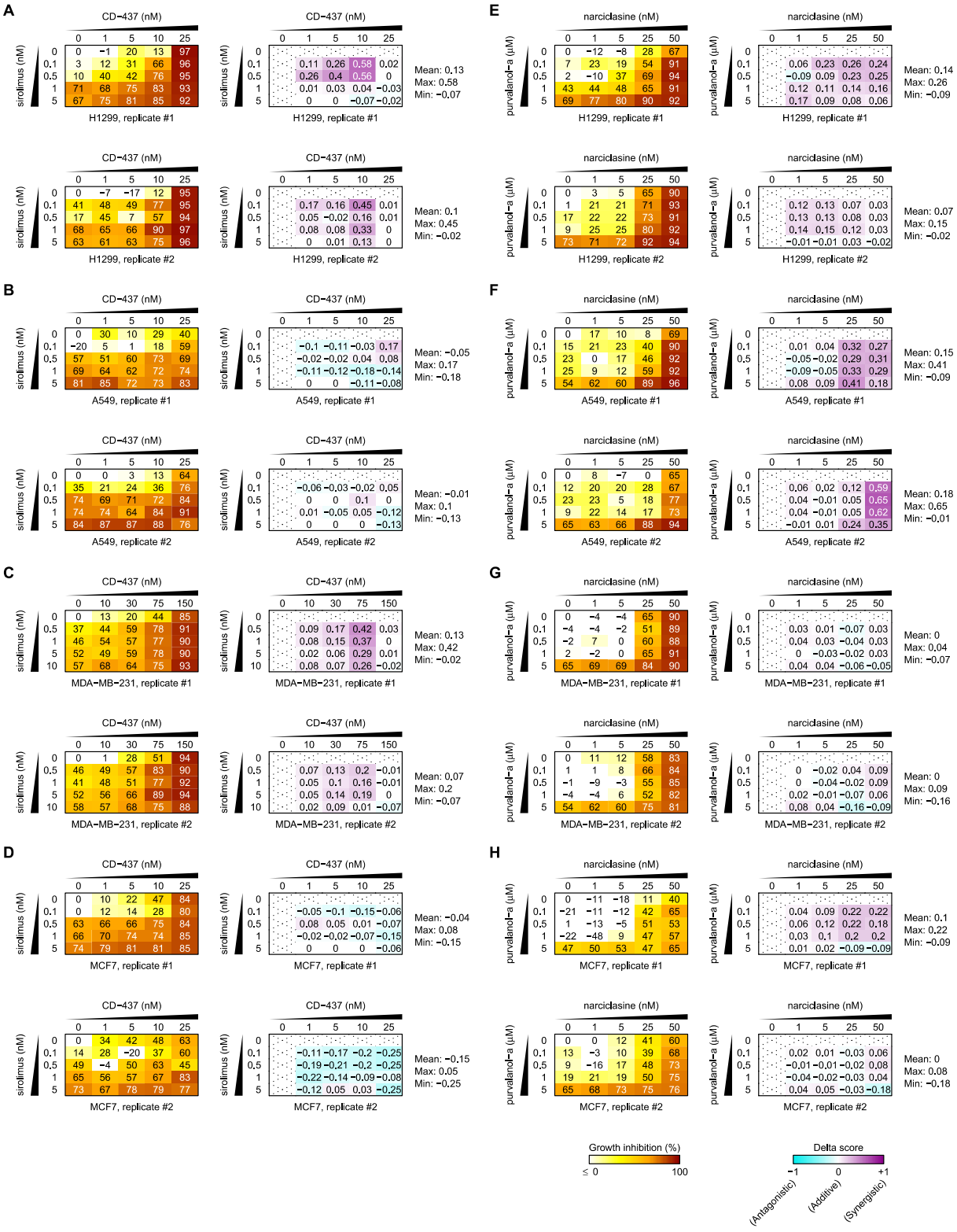


Figure S12. Replicate data of the validated drug combinations, related to Figure 7.

Clonogenic assays were performed in two biological replicates for the combination of CD-437 and sirolimus (A–D) and the combination of narciclasine and purvalanol A (E–H) in H1299 (lung cancer; A, E), A549 (lung cancer; B, F), MDA-MB-231 (breast cancer; C, G), and MCF7 (breast cancer; D, H) cells.

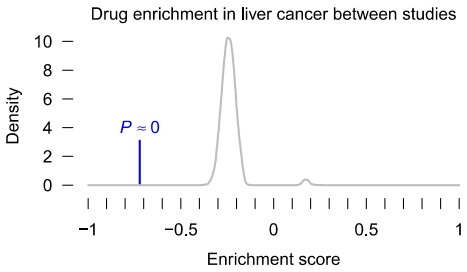


Figure S13. Enrichment of effective compounds for liver cancer identified by our approach in the top compound list generated by Chen et al., 2017, related to Figure 7.

Density of enrichment scores generated from 10,000 permutations (gray line), with the observed enrichment score as indicated (blue line). See Methods.

TRANSPARENT METHODS

Large-scale perturbation data. We used the same procedures for data preprocessing and for manual curation of primary small-molecule mechanisms of action as previously described (Huang et al., 2018). In brief, 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 with accession number GSE92742; 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 using the 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 S1, corresponding to a union of 3,332 chemical and 3,934 genetic perturbagens. For chemical perturbagens, we manually curated primary mechanisms of action (MoAs) and, if available, the WHO Anatomical Therapeutic Chemical Classification codes (ATC codes) with associated LINCS metadata and MoA tag descriptions (Figure S4 and Table S1).

Discovery of recurrent perturbation–transcript regulatory associations. To mitigate cross-plate batch effects in LINCS L1000 data, in which most perturbation experiments were performed in only one biological replicate with no more than three technical replicates in a cell type, we followed the procedures for normalization as described below. For each of detection instances corresponding to a given perturbation *p* (of d6, d24, or sh96) treated in a given cell type *c* (one detection instance indicates one well of measurement in a 384-well plate for fluorescence detection), we first subtracted the

average expression values for all corresponding control experiments (of d6, d24, or sh96 in cell type *c*) on the same detection plate from the expression values for the perturbation instance. Next, for each detection plate, we subtracted the average expression values for all corresponding control experiments from each of the control instances. These processed values corresponding to the treatment (of perturbation *p* in cell type *c*) and the relative control group were quantile-normalized separately to dissipate within-group deviation effects. Based on these normalized values, we computed the signal-to-noise ratios (SNRs) to rank genes in this context (namely, the difference of two means divided by the sum of two standard deviations from the 'perturbed' and 'control' groups) and combined them with a rank-based method (Jacobsen et al., 2013) modified *ad hoc* to infer recurrent perturbation−transcript regulatory relationships.

First, for a given perturbation *p* treated in a cell type *c*, we defined a relative rank for gene *g*

$$rr_{g,p,c} = \frac{r_{g,p,c}}{\left|L_{p,c}\right|} - \frac{1}{2\left|L_{p,c}\right|} \tag{1}$$

where $L_{p,c}$ is the ordered (from negative to positive) vector of SNRs involving perturbation *p* in cell type *c* and $r_{g,p,c}$ is the rank for gene *g* in the vector $L_{p,c}$. Note that $rr_{g,p,c}$ is uniformly distributed on the interval (0, 1). Under the null hypothesis that no negative score exists between gene *g* and perturbation *p* 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

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

In this way, the chi-squared test gave a one-sided *P*-value called p_-. Alternatively, we could test the hypothesis at another side using the inverted rank ($|L_{p,c}| - r_{g,p,c} + 1$), 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 *p* with measurements of gene *g* in at least half of the cell types (≥5 in this case), we could define a recurrence score REC as the signed logarithm of these *P*-values

$$REC_{g,p} = \begin{cases} \log_{10}\left(2 \times p_-\right), \text{if } p_- < p_+ \\ -\log_{10}\left(2 \times p_+\right), \text{if } p_+ < p_- \\ 0 \qquad\quad , \text{if } p_- = p_+ \end{cases} \tag{3}$$

To resolve the multiple testing problem, we used the Benjamini−Hochberg (BH) method to correct *P*-values as false discovery rates (FDRs) for all perturbation−gene pairs. The inference analysis was performed for each perturbation type (d6, d24, or sh96) separately, and those recurrent relationships with FDR < 0.001 are summarized in Figure 2A and Table S2.

To assess the robustness of recurrent perturbation−gene pairs against cell types, we resampled 60% of cell types and repeated the analysis for 100 times for each perturbation type (d6, d24, or sh96). A recurrent perturbation−gene pair is recovered in one round of the resampling analysis if FDR < 0.001. The robustness of a given recurrent perturbation−gene pair is then defined as the ratio of the number of cases in which they are recovered over the number of tested cases (Figure S1).

Comparison with compound–target annotations. To allow comparison between transcript-level recurrences and compound–target interactions, we used the publicly available drug information resource DrugBank (Law et al., 2014) (version 4.3). We considered only the human targets, for which the action types 'pending', 'unknown', and 'other' were incorporated into 'other/unknown', corresponding to a total of 11,045 drug−target pairs of miscellaneous action types (Figure S2). The mapping of compound names between DrugBank and LINCS was achieved using PubChem Compound Identifiers (CIDs) or compound aliases if available (Table S2), and the mapping between protein targets (DrugBank) and transcripts (LINCS) was made by gene symbols together with manual check (Table S2). This process by which to resolve many-to-many relationships for both compound and gene names between DrugBank and LINCS yielded 11,022 non-redundant compound–target annotations for use in Figure 3.

External validation of transcript-level recurrences. We performed differential gene expression analysis on relevant perturbation datasets available in Gene Expression Omnibus (GEO) to examine whether the inferred transcript-level recurrences within LINCS are adequately reproducible externally. We used the datasets with GEO accession numbers listed as follows: GSE28662 (triptolide at 2, 4, and 6 h in MCF7 cells; Affymetrix Human Genome U133A Array), GSE69844 (vorinostat at 6 h in HepaRG cells; Affymetrix Human Genome U219 Array), GSE69845 (clobetasol, vinblastine, or vorinostat at 6 h in MCF7 cells; Affymetrix Human Genome U219 Array), and GSE69849 (vinblastine or vorinostat in Ishikawa cells; Affymetrix Human Genome U219 Array). 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. For each perturbation experiment, the differentially expressed (DE) genes (two-sided unpaired Student's *t*-test, BH-corrected *P*-value < 0.05 or 0.001) were compared with corresponding transcript-level recurrences (FDR < 0.001 or 10^{-6}) for upregulated and downregulated genes separately using the hypergeometric test (Figures 4 and S3).

Analysis of small-molecule transcriptional regulatory modules. For each primary MoA or ATC code at any level (1–4), we first computed for each gene the ratio of recurrent hits (i.e., the number of REC scores with FDR < 0.001, considering both d6 and d24 relationships) over all recurrent and non-recurrent hits (i.e., the number of available REC scores) as the proportion of hits among the tested compounds belonging to the class. Examples of recurrently regulated genes for primary MoA glucocorticoid, antipsychotic, and heat shock protein (HSP) inhibitor and for ATC codes L01D (cytotoxic antibiotics and related substances), C01A (cardiac glycosides), and L01CA (vinca alkaloids and analogues) are shown in Figures 5A, S5–S7, and 7A.

To identify transcriptional regulatory modules across small-molecule classes, we then defined a gene with proportion of hits ≥ 20% (with consistent hits by ≥3 compounds belonging to a given class) to indicate a positive (+1) or negative (−1) regulation by a given class label (primary MoA or ATC code), or no (0) regulation otherwise. Small-molecule classes with at least one gene hit (i.e., not all = 0) were included for analysis, and genes with ≥ 10% hits by the considered small-molecule classes were marked out as 'ubiquitously' regulated genes. The values of small-molecule-class gene regulation (−1, 0, or +1) were used to define the similarity between any two genes **x** and **y** (across primary MoAs or ATC codes separately) by an element-wise Jaccard index (i.e., $|\mathbf{x} \cap \mathbf{y}|/|\mathbf{x} \cup \mathbf{y}|$), in which, for any class j for which two values x_j and y_j are elements of **x** and **y**, respectively, $(x_j, y_j) = (+1, +1)$ or $(-1, -1)$ falls within $\mathbf{x} \cap \mathbf{y}$, whereas any $(x_j, y_j) \neq (0, 0)$ falls within $\mathbf{x} \cup \mathbf{y}$. Using these Jaccard-index-based distances (one minus the Jaccard similarity), we clustered genes (ubiquitously and non-ubiquitously regulated genes separately) or small-molecule classes by WPGMA (weighted pair group method with arithmetic mean, using hclust in R package stats version 3.4.3 with method = "mcquitty") and selected the number of clusters k across a full range of possible values ($k = 2,..., n - 1$,where n is the number of genes or small-molecule classes) that yielded the highest silhouette score (Wiwie et al., 2015), a quality measure that judges a clustering by its internal statistical properties (reaching the best score at +1 and the worst at −1). For gene clusters, this process resulted in 'ubiquitous' and 'non-ubiquitous' clusters. To visualize these clusters, we showed the ubiquitous clusters first and then ordered the non-ubiquitous clusters according to which small-molecule class has the highest proportion of hits, 'diagonalizing' the results when possible (Figure 5B for primary MoAs and Figure S8B for ATC categories). The ubiquitous clusters are shown with the number of compounds accounting for consistent hits in each available small-molecule class (Figure 5C for primary MoAs and Figure S8C for ATC categories).

We next tested whether the identified gene regulatory modules could be enriched for specific cancer hallmarks, using the hallmark gene sets (H) in the Molecular Signature Database (MSigDB (Subramanian et al., 2005); version 5.0). We performed hypergeometric tests on enrichments of gene modules (of primary MoAs and ATC categories separately) for each of hallmark gene sets and adjusted P-values using the BH correction. The enrichment results were visualized by selecting gene modules with at least one significant enrichment (corrected $P < 0.05$) while grouping the enriched cancer hallmarks according to which gene module has the most significant P-value in a nearly 'diagonal' fashion. For each selected gene module, we indicated small-molecule classes that

regulated at least 50% of genes in the module (Figure 5D for primary MoAs and Figure S8D for ATC categories).

Comparison with small-molecule sensitivity. We compared transcript-level regulatory recurrences with small-molecule sensitivity patterns that related an area under the dose–response curve (AUC) metric for each of 481 tested compounds to 18,543 basal (unperturbed) transcripts across 823 cancer cell lines (Rees et al., 2016), publicly accessible in the Cancer Therapeutics Response Portal (CTRP, version 2; http://portals.broadinstitute.org/ctrp.v2.2/). We used significant sensitivity–expression relationships with Bonferroni-corrected two-sided *P*-values < 0.05 (corresponding to *z*-scored AUC correlation strength $|z| > 4.69$), in which a negative correlation indicates high expression of a given basal transcript correlates with sensitivity of a compound of interest whereas a positive correlation indicates low expression correlates with sensitivity. We overlapped perturbation–transcript regulatory interactions (in this study) with sensitivity–expression relationships (in CTRP) (Figure 6A and Table S6), and showed significantly correlated relationships for the compounds vorinostat and bortezomib and the transcripts *TIMELESS* and *BAG3* in Figure S9. To examine overlapping relationships at the MoA level, we mapped the compound names in CTRP to our primary MoAs and counted occurrences of significant MoA–transcript overlapping pairs (including both d6 and d24 recurrent relationships) over twice the number of available compounds for each MoA in a given MoA–transcript pair (Table S7). The significant MoA–transcript overlapping pairs with occurrences ≥ 3 and the proportion of ≥ 50%, or with occurrences ≥ 8 alone, are shown in Figure 6B.

Expression-based therapeutic score. For each compound available in LINCS, we computed an expression-based therapeutic score as the extent of reversal relationships between small-molecule transcript-level recurrences (d6 and d24 separately) and a given disease gene signature (Figure S10A). In this study, we used the gene-expression signatures between primary tumors and tumor-adjacent normal tissues among eight cancer types from The Cancer Genome Atlas (TCGA) (Aran et al., 2017) (Bonferroni-corrected *P*-value < 0.05, $|\log_2(\text{fold change})| > 1$, and $\log_2(\text{count per million}) > 3$), corresponding to $n = 1{,}312$ differentially expressed (DE) genes in bladder cancer (BLCA), $n = 1{,}417$ in breast cancer (BRCA), $n = 1{,}503$ in colon cancer (COAD), $n = 962$ in live cancer (LIHC), $n = 1{,}446$ in lung cancer (LUAD), $n = 568$ in prostate cancer (PRAD), $n = 759$ in thyroid cancer (THCA), and $n = 1{,}876$ in uterine cancer (UCEC) (Table S8). We restricted our analysis to genes involved in both the inference of transcript-level recurrences and the derivation of disease gene signatures.

Formally, given a disease signature G and a perturbation p, we defined a therapeutic score

$$score(p,G) = -\left(\frac{\sum_{g \in G} w(g) h(p,g)}{\lambda \sum_{g \in G} |w(g)|} \right) \tag{4}$$

where λ is the magnitude of the REC score for which FDR = 0.001, $w(g)$ determines the weight of gene g (in this study, $w(g) = \log_2 fc(g)$ where $fc(g)$ is the fold change of gene g), and $h(p, g)$ summarizes the contribution from the transcript-level recurrence between perturbation p and gene g as

$$h(p,g) = [\![rec(p,g) > \lambda]\!] \times sgn(rec(p,g)) \times \log_2 |rec(p,g)| \tag{5}$$

in which $[\![\cdot]\!]$ returns 1 if the expression inside gives a TRUE value and returns 0 otherwise, $rec(p, g)$ provides the REC score between perturbation p and gene g, and $sgn(\cdot)$ is the sign function. The therapeutic score provides a rough interpretation of the gene reversal such that suppose there are 1,000 disease genes and a given perturbation can reverse exactly 100 genes with REC scores for which all FDRs were 0.001, then the therapeutic score is 0.1. Note that the therapeutic score in equation (4) can be rewritten as

$$score(p,G) = rev(p,G) + fwd(p,G) \tag{6}$$

with

$$rev(p,G) = score(p, G^{r}) \geq 0 \tag{7}$$

$$fwd(p,G) = score(p, G^{f}) \leq 0 \tag{8}$$

such that $w(g)h(p, g) < 0$ for all g's in G^r (corresponding to reverse hits), $w(g)h(p, g) > 0$ for all g's in G^f (corresponding to forward hits), and G^f and G^r are disjoint subsets of G. To assign statistical significance to the therapeutic score for each perturbation p given a disease signature G, we shuffled the genes in the perturbation signature and re-computed a permuted therapeutic score, and this procedure was repeated 1,000 times. We then computed an empirical permutation P-value (P_{ecdf}) as the ratio of

the number of permuted therapeutic scores whose absolute values were greater than or equal to that of the actual therapeutic score over the number of permutations (n = 1,000; Figure S10D). If $P_{ecdf} < 1/100$, we used the generalized Pareto distribution (GPD) to model the P-value distribution (Knijnenburg et al., 2009) (the number of exceedances used for fitting GPD was 250, and the parameters were estimated by the probability weighted moment (PWM) method using the gpdFit function in the R package fExtremes (version 3042.82) with the arguments type = "pwm" and information = "expected") and computed an improved estimate (P_{gpd}) for this low P_{ecdf}. Finally, the permutation P-value for each perturbation p given a disease signature G was defined as P_{gpd} if $P_{ecdf} < 1/100$, or P_{ecdf} otherwise. To address the multiple testing problem, we used the BH method to correct permutation P-values for all perturbations p's across the tested disease signatures G's (n = 8). For visualization in Figure 7A, we selected chemical perturbagens with at least one significant therapeutic score ≥ 0.05 (d6 or d24; corrected P-value < 0.05) across the tested disease signatures (Table S9) and, if available, computed the median dose–response AUC across cancer cell lines of each corresponding cancer type for each selected chemical perturbagen (Rees et al., 2016), using the mapping information available in the Genomics of Drug Sensitivity in Cancer (GDSC) (Iorio et al., 2016).

For computing the therapeutic score for a given drug combination, we used a strategy that checks the consistency of hits (reverse, forward, or neither) among the compounds involved in the combination (Figure S10B). Formally, given a disease signature $G = \{g_1, \ldots, g_m\}$ and a combination Π involving n perturbations $p_1, \ldots, p_n$, we defined a combinatorial therapeutic score

$$score(\Pi, G) = -\left(\frac{\sum_{j=1}^{m} v_j}{\lambda \sum_{j=1}^{m} \left| w(g_j) \right|} \right) \tag{9}$$

with

$$v_j = \begin{cases} \min\limits_{p_i \in \Pi} w(g_j) h(p_i, g_j), \text{ if } w(g_j) h(p_i, g_j) \le 0, \ \forall p_i \in \Pi \\ \max\limits_{p_i \in \Pi} w(g_j) h(p_i, g_j), \text{ if } w(g_j) h(p_i, g_j) \ge 0, \ \forall p_i \in \Pi \\ 0 \qquad\qquad , \text{otherwise} \end{cases} \tag{10}$$

Note that when n = 1 (i.e., a single perturbation), equation (9) is reduced to equation (4). In this study, we calculated the combinatorial therapeutic scores for two-way combinations of chemical perturbations whose therapeutic scores ≥ 0.01 with corrected P-values < 0.05. We adopted a similar procedure

as done for single therapeutics to assess statistical significance of the combinatorial therapeutic score for each given combination Π given a disease signature G by shuffling the genes in each perturbation signature of the combination independently, except that the number of permutations was 100 and the number of exceedances used for fitting GPD was 50.

Based on individual and combinatorial therapeutic scores for the chemical perturbations and the combinations of them respectively, we predicted synergistic drug combinations for cancer (Figure S10C). Formally, for a given combination Π involving perturbations $p_1, \ldots, p_n$ with therapeutic scores $s_1, \ldots, s_n$, respectively, where $s_1 \geq \ldots \geq s_n \geq 0$, we defined a threshold τ_Π

$$\tau_\Pi = s_1 + \alpha \sum_{k=2}^{n} s_k \tag{11}$$

where we set $\alpha = 0.75$. Let T_Π be the actual combinatorial therapeutic score for the combination Π, then Π is predicted to be synergistic if $T_\Pi > \tau_\Pi$.

For visualizing the predicted synergistic drug combinations across cancer types in Figure 7B, drug combinations were ordered by the sum of magnitudes of synergistic combinatorial therapeutic scores with corrected P-values < 0.05 (Table S10), with the median dose–response AUC value across cell lines of each corresponding cancer type for each drug in the combination shown if available. The occurrences of the combinations of primary MoAs among the top 100 synergistic drug pairs across cancer types are summarized in Figure 7C.

We validated our approach first by determining the enrichment of effective compounds we identified for liver cancer in the top compound list generated by Chen et al. (Chen et al., 2017) using the technique of gene set enrichment analysis. The effective compounds identified by our approach were those with d6 or d24 therapeutic scores > 0 and corrected P-value < 0.05 in the liver cancer signature (n = 260; Table S9). The compound list was obtained from Chen et al. but restricted to the compounds with definable therapeutic scores (Table S9), with sRGES scores used for enrichment analysis. The statistical significance was evaluated by shuffling the compounds in the list for 10,000 times to obtain a distribution of permuted enrichment scores, from which we used the GPD model as described above to approximate the permutation P-value (Figure S13). We also compared the effective compounds for cancer in this study with those compounds identified by OncoTreat for their ability to target metastatic master-regulator protein activities in neuroendocrine tumors (Table S11) (Alvarez et al., 2018).

Drug-combination experimental validation. *Cell culture and chemicals.* MDA-MB-231, MCF7, H1299, and A549 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), incubated at 37 °C in humidified atmosphere with 5% CO_2, and routinely passaged when

90–95% confluent. Cells tested negative for mycoplasma. CD-437 (1549) was purchased from Tocris biosciences. Sirolimus (SLKS1039) was ordered from Selleckchem. Narciclasine (MCEHY-16563) and purvalanol A (MCEHY-18299A) were purchased from MedChem Express. DMEM (12800), FBS (A3160601), and trypsin-EDTA (15400054) were obtained from Thermo Fisher Scientific.

Clonogenic assay. Cells (1,000 per well for H1299 and A549 cells and 5,000 per well for MDA-MB-231 and MCF7 cells) were seeded in six-well plates and allowed to adhere overnight in culture media. Cells were then treated with drugs as indicated for 12 days during which the culture media with or without the drugs were replaced every three days. The remaining cells were fixed with 100% methanol, stained with 1% crystal violet, and recorded by a digital scanner. Crystal violet was extracted from the stained cells by 10% acetic acid, and growth inhibition was quantified by the following steps: (1) the absorbance at 595 nm (A_{595}) was measured using a ELISA reader (Bio-Rad); (2) relative growth was represented by A_{590} values with background correction and normalized with the corresponding control group as 100%; and (3) growth inhibition was calculated as one minus relative growth. Two technical replicates were performed for each of two independent biological replicates.

Analysis of drug synergy. Drug synergy was assessed by the delta score (Yadav et al., 2015), ranging from −1 (antagonism) through 0 (additive effect) to +1 (synergy), which is based on the zero interaction potency (ZIP) model and has been shown to produce superior accuracy for classifying drug combinations with known interaction modes compared with those based on the Loewe additivity or Bliss independence model. For each tested drug combination in a given cell line, we took the average (Figures 7D and S11) or individual values (Figure S12) of clonognic growth inhibition across biological replicates to calculate delta scores, using the ZIP function in the R package synergyfinder (version 1.4.2) with the arguments correction = TRUE and nan.handle = "L4".

SUPPLEMENTAL REFERENCES

Alvarez, M.J., Subramaniam, P.S., Tang, L.H., Grunn, A., Aburi, M., Rieckhof, G., Komissarova, E.V., Hagan, E.A., Bodei, L., Clemons, P.A., *et al.* (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nat. Genet. *50*, 979-989.

Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M., and Butte, A.J. (2017). Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat. Commun. *8*, 1077.

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.

Huang, C.T., Hsieh, C.H, Oyang, Y.J., Huang, H.C., and Juan, H.F. (2018). A large-scale gene expression intensity-based similarity metric for drug repositioning. iScience *7*, 40-52.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Goncalves, E., Barthorpe, S., Lightfoot, H., *et al.* (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740-754.

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.

Knijnenburg, T.A., Wessels, L.F., Reinders, M.J., and Shmulevich, I. (2009). Fewer permutations, more accurate P-values. Bioinformatics *25*, i161-168.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., *et al.* (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. *42*, D1091-1097.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., *et al.* (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109-116.

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 e1417.

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.

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

Yadav, B., Wennerberg, K., Aittokallio, T., and Tang, J. (2015). Searching for Drug Synergy in Complex Dose-Response Landscapes Using an Interaction Potency Model. Comput. Struct. Biotechnol. J. *13*, 504-513.