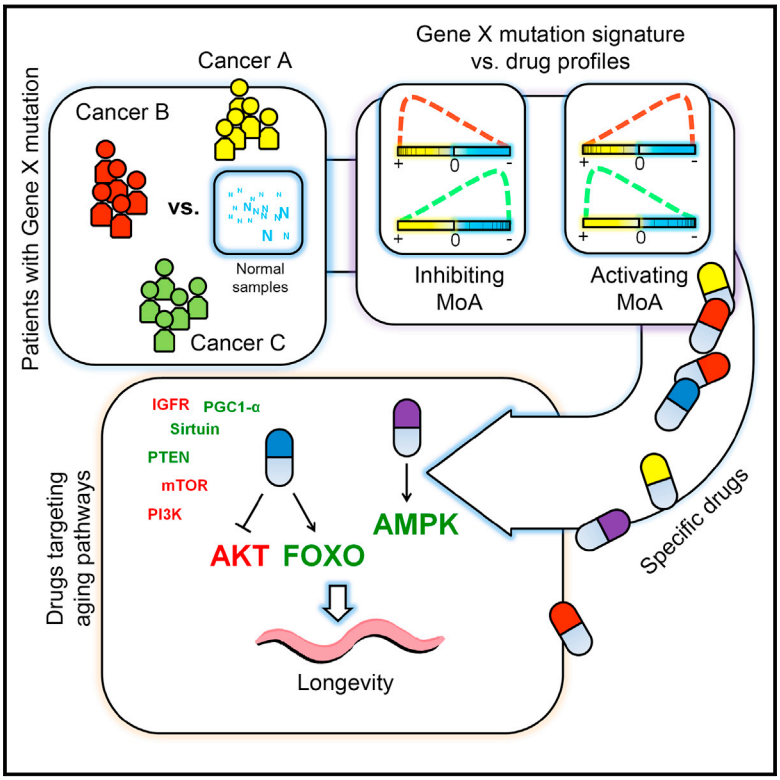
Cell Reports

Resource

Accurate Drug Repositioning through Non-tissue-Specific Core Signatures from Cancer Transcriptomes

Graphical Abstract



Highlights

- Non-tissue-specific core signatures (CSs) were generated for 8,476 human genes
- csD2G repositions drugs through drug-gene CS modules and specificity mapping
- Calcium blockers predicted to target AKT and AMPK are validated by medical records
- Seven drugs were repositioned to selectively target AKT and AMPK and extend lifespan

Authors

Chi Xu, Daosheng Ai, Dawei Shi, ..., Shiqiang Zhang, Yingying Zeng, Jing-Dong Jackie Han

Correspondence

jdhan@picb.ac.cn

In Brief

Xu et al. generated core transcriptome signatures representing single-gene mutations for 8,476 human genes from the cancer transcriptomes in the TCGA database and developed a drug-repositioning method to predict high-specificity candidate target genes for 1,938 drugs. Applied to targeting aging pathways, 7 lifespan-extending drugs are found and functionally validated.

Xu et al., 2018, Cell Reports *25*, 523–535
October 9, 2018
https://doi.org/10.1016/j.celrep.2018.09.031

CellPress

Cell Reports
Resource

Accurate Drug Repositioning through Non-tissue-Specific Core Signatures from Cancer Transcriptomes

Chi Xu,[1,2,3] Daosheng Ai,[1,3] Dawei Shi,[1] Shengbao Suo,[1] Xingwei Chen,[1,2] Yizhen Yan,[1,2] Yaqiang Cao,[1,2] Rui Zhang,[1] Na Sun,[1] Weizhong Chen,[1] Joseph McDermott,[1] Shiqiang Zhang,[1,2] Yingying Zeng,[1,2] and Jing-Dong Jackie Han[1,4,*]
[1]Key Laboratory of Computational Biology, Chinese Academy of Sciences Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: jdhan@picb.ac.cn
https://doi.org/10.1016/j.celrep.2018.09.031

SUMMARY

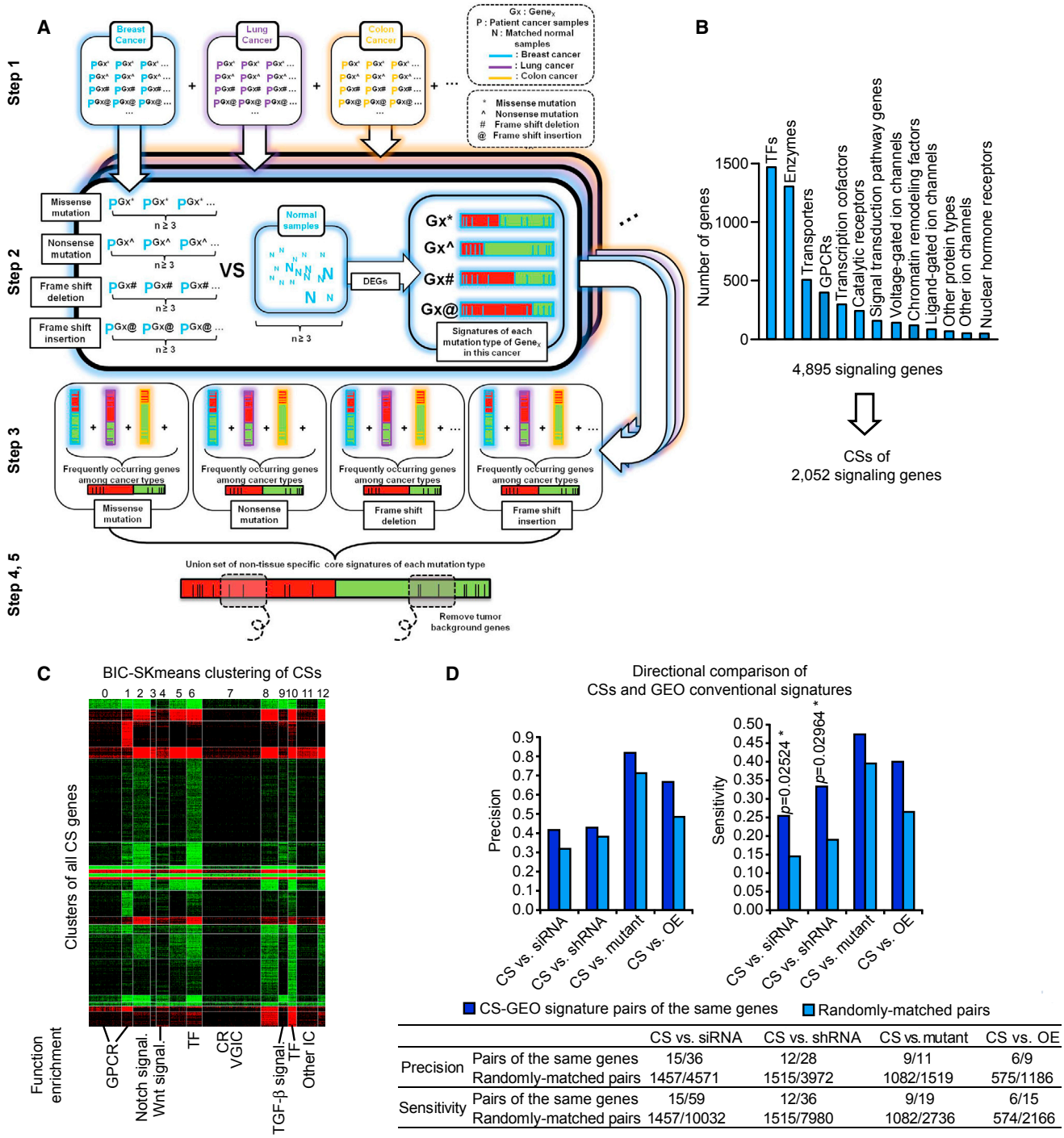
Experimental large-scale screens for drug repositioning are limited by restriction to *in vitro* conditions and lack of applicability to real human conditions. Here, we developed an *in silico* screen in human *in vivo* conditions using a reference of single gene mutations' non-tissue-specific "core transcriptome signatures" (CSs) of 8,476 genes generated from the TCGA database. We developed the core-signature drug-to-gene (csD2G) software to scan 3,546 drug treatment profiles against the reference signatures. csD2G significantly outperformed conventional cell line-based gene perturbation signatures and existing drug-repositioning methods in both coverage and specificity. We highlight this with 3 demonstrated applications: (1) repositioned category of psychiatric drugs to inhibit the TGF-β pathway; (2) antihypertensive calcium channel blockers predicted to activate AMPK and inhibit AKT pathways, and validated by clinical electronic medical records; and (3) 7 drugs predicted and validated to selectively target the AKT-FOXO and AMPK pathways and thus regulate worm lifespan.

INTRODUCTION

The biopharmaceutical industry faces 3 major challenges: (1) a productivity gap between enormous spending in research and development stages and the disproportionate output of novel drugs (Ashburn and Thor, 2004); (2) more effective drugs are needed for complex diseases and aging populations (Wu et al., 2013); and (3) drugs developed based on *in vitro* cell assays or mouse models often cannot be translated to human treatments, with the majority failing clinical trials for showing no efficacy or undesirable effects. To speed up drug development and reduce risks, drug repositioning has gained momentum. Previously approved drugs have already been tested through clinical trials and are well investigated regarding their safety, and many have other possible applications than those originally intended (Ashburn and Thor, 2004). Successful examples of drug repositioning include the use of sildenafil for erectile dysfunction, thalidomide for severe erythema nodosum leprosum (Ashburn and Thor, 2004), and all-*trans* retinoic acid for acute promyelocytic leukemia (Lo-Coco et al., 2013).

Computational repositioning predictions by integrative analysis of pharmacogenomic data have made significant progress because they are low cost and overcome many practical limitations of experimental high-throughput drug library screens (Dudley et al., 2011). Genome-wide expression profiles or cell sensitivity in response to drug treatment have been generated, including the Connectivity Map (CMap) (Lamb et al., 2006), the NCI-60 Human Tumor Cell Lines Screen, and the Genomics of Drug Sensitivity in Cancer (GDSC) profiles (Yang et al., 2013). To take advantage of these perturbation signatures, the idea of creating gene expression signatures to describe physiological, disease, or genetic perturbation states has been proposed (Dudley et al., 2011; Lamb et al., 2006). In this study, by drug repositioning, we aimed to reposition a drug to a new target gene but not to a new disease. Development of a high-throughput strategy to reposition drugs in humans with high precision is still challenging, perhaps due to the following obstacles: (1) batch effects, platform differences, and different tissue and cell backgrounds make it difficult for integrative analysis; (2) target gene signatures in existing resources are insufficient to cover known drug targets (e.g., only ~430 transcription factors (TFs) are included in the Encyclopedia of DNA Elements (ENCODE); and (3) genetic experiments can be performed only in cell lines or animal models, not in human models. In particular, pattern-matching tools cannot judge whether the similarities or differences stem from tissue-of-origin backgrounds or relevant or true drug-gene interactions (DGIs), while confining the searches to the same tissue or cell type severely limits the number of drugs, genes, or pathways that can be analyzed.

Here, to remove effects of the background of tissues of origin and to reposition drugs in a human environment, we constructed non-tissue-specific core signatures (CSs) from The Cancer Genome Atlas (TCGA) (Hudson et al., 2010) transcriptomes.



		CS vs. siRNA	CS vs. shRNA	CS vs. mutant	CS vs. OE
Precision	Pairs of the same genes	15/36	12/28	9/11	6/9
	Randomly-matched pairs	1457/4571	1515/3972	1082/1519	575/1186
Sensitivity	Pairs of the same genes	15/59	12/36	9/19	6/15
	Randomly-matched pairs	1457/10032	1515/7980	1082/2736	574/2166

Figure 1. The Pipeline to Generate CSs for Signaling Genes

(A) Pipeline to construct CSs. Step 1: collect patient samples with mutation of a select gene (gene *x*) in different cancer types. Step 2: generate mutation-type signatures of gene *x* in each cancer/tissue type. Step 3: generate CSs of each mutation type of gene *x*. Steps 4 and 5: merge CSs of each mutation type as the final CS for gene *x* and remove tumor background.

(B) Input of frequently mutated signaling genes and output of CSs.

(C) BIC-SKmeans clustering of CS matrix of signaling genes. Enriched signaling pathways and protein types are indicated at bottom. CR, catalytic receptor; GPCR, G protein-coupled receptor; IC, ion channel; signal, signaling pathways; TF, transcription factor; and VGIC, voltage-gated ion channel.

(legend continued on next page)

CSs are used to identify drugs whose perturbation signatures mimic a single gene mutation. Using 3,546 drug treatment profiles, we also developed a drug-specificity scoring system to precisely predict 83,745 DGIs with high targeting specificity. Among these DGIs, we predicted a psychiatric drug category to inhibit the transforming growth factor β (TGF-β) pathway, which comprised 7 drugs to specially target the conserved AKT-FOXO and 5′ adenosine monophosphate-activated protein kinase (AMPK) pathways, 6 of which were successfully validated in human cell lines and in *Caenorhabditis elegans* lifespan assays. We also predicted antihypertensive calcium channel blockers to inhibit the AKT pathway and activate the AMPK pathway, which are supported by electronic medical records showing a blood lipid-lowering effect—an expected consequence of AKT inhibition and AMPK activation. Our accurately *in silico* repositioned DGIs provide a rich resource for further exploration.

RESULTS

Generating CSs of Signaling Genes from Cancer Transcriptomes

Cell signaling pathways, which respond to an extracellular stimulus to regulate gene expression, cell metabolism, and development, have been recognized as essential drug targets in disease treatment (Behar et al., 2013). In different tissues or cell types, the core pathways and their immediate downstream target genes often share effectors, but the final output of these pathways can be very different (Akhurst and Hata, 2012; Johnson and Halder, 2014; Siegel and Massagué, 2003). To focus on and verify the core effects of pathways, we curated 4,895 signaling genes coding for receptors, enzymes, ion channels, TFs, and components of 8 major signaling pathways (Figure 1B).

While genome-wide reverse genetics is still difficult to perform in human tissues or cell lines, cancer genomes contain a naturally occurring pool of genetic mutations for many, if not all, human genes. TCGA provides thousands of high-quality transcriptomes from human cancer patients and tissue-matched normal controls in >20 cancer types. Cancer samples with somatic mutations in a gene can be used to define its transcriptome perturbation signature. Furthermore, the abundant variety of tissue origins of cancers provides an opportunity to search for the common differentially expressed genes (DEGs, or signature genes) shared by multiple tissues to represent the core effects of gene mutations without tissue-of-origin background.

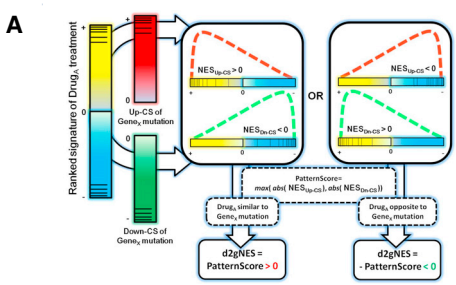
Here, we constructed a pipeline with the input of the 4,895 signaling genes to generate CSs using TCGA mutation and transcriptome data (Figure 1A). For each cancer type, patient samples carrying each type of mutation (e.g., missense mutations, nonsense mutations, frameshift deletion, frameshift insertion) in a gene is defined as a genetic perturbation group and compared with normal controls to detect mutation-type signatures in each cancer type (Figure 1A, steps 1 and 2). Different mutation types were analyzed separately because we observed that they often lead to different expression changes for the same gene, as shown by the low overlap among different mutation-type signatures; this suggests that they may probe various perturbation-type effects of the same gene (Figure S1A). We regard cancers of the same tissue as having the same tissue of origin and then generating CSs by enforcing overlap across multiple tissues (Figure 1A, steps 3 and 4; Method Details). These are followed by a CS trimming step, removing 87 DEGs shared by >50% of cancer types as common tumor background (Table S1). To reduce data complexity and to exclude undetermined or alternative explanations, we used only the single gene mutation CSs for analysis in our study, discarding CSs derived from samples in which multiple genes are commonly mutated (only 628 multigene signatures versus 2,052 single-gene signatures when using the 4,895 signaling genes as input, and 1,470 multigene signatures versus 8,476 single gene signatures when using all human coding genes as input; detailed description and distribution of these multiple hits can be found in Figures S1B and S1C). In the end, across different tissues, the expression fold changes of CS genes showed high consistency (Pearson correlation coefficient [PCC] ~0.85; Figure S1D), suggesting that they are indeed non-tissue specific.

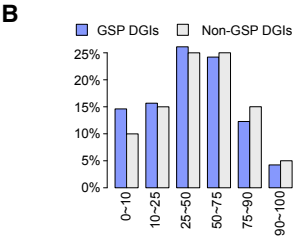
Using this pipeline, we obtained CSs for 2,052 of the 4,895 curated signaling genes. More than 98% of them were from missense mutations (Figure S1E), which mostly occurred in protein-coding regions, including functional domains, probably leading to loss of function. For example, according to TCGA mutation annotations, in breast-invasive carcinoma, 82,816/90,489 (92%) of mutations map to functional domains and 2,356/2,585 (91%) do so in acute myeloid leukemia. Only genes related to tumorigenesis and tumor progression (e.g., *TP53*) had enough samples carrying mutation types other than missense mutations for analysis (≥3 mutation samples and ≥3 normal samples).

Clustering analysis of CSs for the 2,052 signaling genes by an auto-optimizing Bayesian information criterion (BIC)-SKmeans algorithm (Zhang et al., 2013) showed 9/13 gene clusters significantly enriched for Notch, Wnt, TGF-β pathways, G protein-coupled receptor, catalytic receptor, TFs, and voltage-gated ion channels, respectively (Figure 1C). Genes in the same pathways or with the same function types were automatically grouped together. This indicates that CSs can identify and represent different signal pathways affected by human genetic perturbations and separate them from one another.

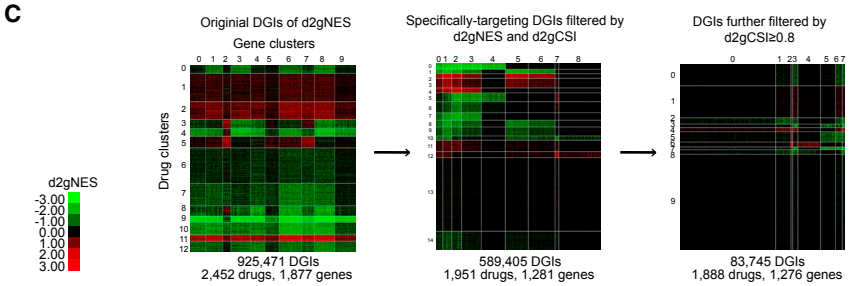
To evaluate the validity of CSs, we compared them with 196 conventional signatures generated by single-gene RNAi, genetic mutants, and overexpression (OE) of the same genes collected from GEO. A total of 78/114 (68%) of CSs showed significant overlap with at least one of the corresponding conventional signatures ($p < 0.05$ and false discovery rate [FDR]) <0.25, as determined by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) (Table S2). For all matched pairs of CSs and conventional signatures, compared to random background,

(D) Comparison in the directions of expression changes of CSs and GEO conventional signatures. Precision and sensitivity of CSs to capture GEO conventional perturbation signatures for the same genes in different types of perturbations, compared with random background. *, Binomial test p values < 0.05. Numbers of comparison cases for different types of perturbation are indicated at bottom. Precision and sensitivity are as defined in Method Details.
See also Figures S1 and S2 and Tables S1 and S2.



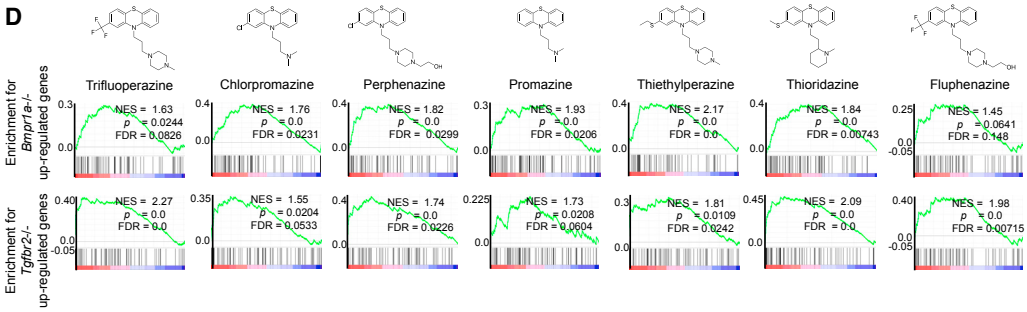


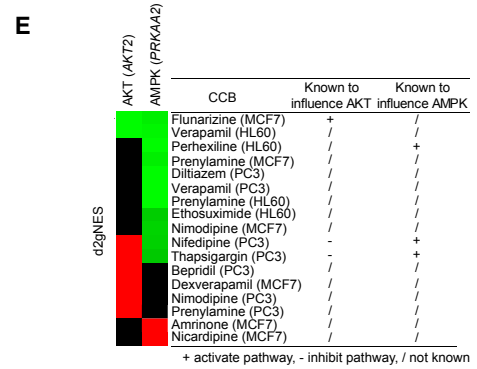
$$d2gCSI = 1 - \frac{n_G \text{ connected to } D' \text{ with } |PCC| \geq (|PCC_{D'G}| - 0.2)}{n_G}$$



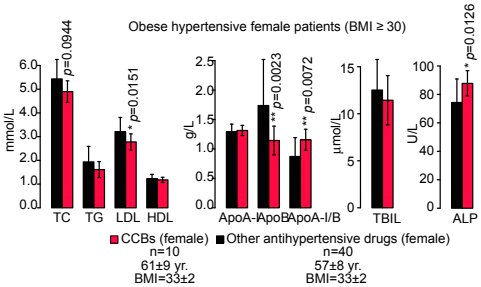
Drug cluster	Significantly enriched drug categories
0	Penicillins / Oxytocics / Androgen antagonists / Bisphosphonates
2	Progestins
3	*Norepinephrine-reuptake Inhi. / *Adrenergic uptake inhi. / Antiparasitic agents / Gout suppressants / Hormone antagonists
4	*Phenothiazines / *Dopamine antagonists / *Antipsychotics / Trypanothione disulfide reductase inhi. / Hsp90 inhi
5	Fibrinolytic agents / Microtubule destabilising agents / Selective serotonin reuptake inhi.
7	*Glucocorticoids
8	*Microtubule destabilising agents / Antihistamines / Na^+/K^+-transporting-ATPase inhi.
9	*Antimetabolites / *Nucleotide reverse transcriptase inhi./Antiviral / Anti-HIV / Antiprotozoal agents / Antimalarials / Dihydrofolate
12	Parasympathomimetics / Serotonin antagonists

Gene cluster	Significantly enriched GO terms
0	chromatin mod. / transmembrane receptor signal.
2	TGFβ signaling (R-SMAD binding)
3	cell cycle / immune effector process / cAMP, MAPK signal.
4	leukotriene biosynthesis / transcription
5	thymus develop. / vitamin transport / bone develop. / transcription
6	chromosome mod. / Ras signal. / urogenital system / response to hypoxia / NF-kappaB signal.
7	response to stimulus
8	CNS neuron develop. / defense to bacteria
9	transcription / mesoderm develop. / glucose homeostasis / NF-kappaB signal.





CCB	Known to influence AKT	Known to influence AMPK
Flunarizine (MCF7)	+	/
Verapamil (HL60)	/	/
Perhexiline (HL60)	/	+
Prenylamine (MCF7)	/	/
Diltiazem (PC3)	/	/
Verapamil (PC3)	/	/
Prenylamine (HL60)	/	/
Ethosuximide (HL60)	/	/
Nimodipine (MCF7)	/	/
Nifedipine (PC3)	-	+
Thapsigargin (PC3)	-	+
Bepridil (PC3)	/	/
Dexverapamil (MCF7)	/	/
Nimodipine (PC3)	/	/
Prenylamine (PC3)	/	/
Amrinone (MCF7)	/	/
Nicardipine (MCF7)	/	/



(legend on next page)

CSs consistently showed higher precision and sensitivity in capturing consistent (intersecting) expression change directions across different perturbations of the same gene (Method Details; Figures 1D and S2A for examples; Table S2). Similarly, CSs are also essential for achieving high accuracy in drug reposition (Figure S2B; Table S2; Method Details). These demonstrate that CSs can consistently represent transcriptome perturbations of a gene across different experimental methods.

We further examined the proportions of the CS mutations to be potentially inhibiting, activating a gene, or lacking a clear pattern based on the comparison to the GEO datasets. When using GEO mutant and OE data, 42% and 53% of the CSs do not have a clear direction (neither same nor opposite); but for those having a clear direction, 81% of them (47%/47% + 11%) showed the same directions as mutants and 85% (40%/40% + 7%) showed the opposite directions of OE, implying that these mutation CSs are consistent with inhibiting rather than activating target genes (Figure S2C).

Scanning Pharmacological Transcriptomes by CSs to Reposition Drugs

To use CSs for drug repositioning, we first generated 3,546 drug-treatment signatures (DTSs) based on CMap, which contains microarray data mainly for 3 human cancer cell lines treated with 1,309 drugs. Because our CSs are insensitive to tissue-specific transcriptome changes, they can be directly compared to treatment in any tissue or cell types, thus increasing repositioning coverage. Therefore, samples from different cell lines were analyzed separately for each drug. Here, we treat different dosages in each cell line as replicates to increase statistical power.

To evaluate the extent of signature overlap and potential mode of action of a drug on a gene set, we defined a drug-to-gene normalized enrichment score (d2gNES) with a plus-or-minus sign, determined by GSEA, to test CSs against the ranked list of DTS gene expression changes (Figure 2A; Method Details). A positive d2gNES score means drugs show the same patterns as gene mutations. Because most CS-producing gene mutations are probably loss-of-function mutations, a positive d2gNES score implies that the drug inhibits the activity of the gene. Conversely, a negative d2gNES implies activation of the activity of a gene by a drug. As background controls, we generated random CS gene sets for each signaling gene and repeated the analysis 1,000 times for each DTS, thereby identifying 924,357 significant interactions ($p < 0.05$ and FDR <0.25; 12.7% of all 7,276,392 possible interactions between 2,452 drugs and 1,877 genes).

Clustering of the d2gNES matrix by the BIC-SKmeans algorithm showed that although thousands of drugs and genes were analyzed, both drugs and genes were automatically clustered into a relatively small and finite number of gene and drug categories (Figure 2C). The 2,452 drugs were clustered into 13 DCs, with enrichment for antipsychiatric (DC-3 and -4), anti-infection (DC-0, -3, and -8), hormone control (DC-0, -2, and -7), and ion transport (DC-8) functions, and regulation of small metabolites (DC-9), respectively. The DC signatures overlap with gene clusters (GCs) enriched for central nervous systems (GC-8), defense against bacteria (GC-8), immune response (GC-3), and vitamin transport (GC-5), as well as classical pathways (e.g., TGF-β [GC-2], nuclear factor κB [NF-κB] [GC-9], and mitogen-activated protein kinase [MAPK] [GC-3]), signaling and transcription (GC-4, -5, and -9), chromatin modification (GC-0 and -6), and cell development (GC-9), respectively; these are suggestive of activation (Figure 2C, green) or inhibition (Figure 2C, red) of the GCs by the corresponding DCs. For example, we found 13 psychiatric drugs (DC-3 and -4) to potentially inhibit the TGF-β pathway genes (GC-2; Figure 2C). Two of them, bromocriptine and trifluoperazine, have been reported to inhibit TGF-β (Figure S3E) (Miyoshi et al., 2008; Zhuge and Cederbaum, 2006). Among the rest, 7 phenothiazine drugs showed a high chemical structure similarity with trifluoperazine (Figures 2D and S7). Including trifluoperazine, 6/8 of phenothiazine DTSs had DEG profiles similar to those of *Bmpr1a* knockout mouse embryonic stem cells and 7/8 to those of *Tgfbr2* knockout mouse palate tissues (Figure 2D). The above results suggest a signaling pathway target of the phenothiazine drug class and a potential molecular mechanism for their antipsychotic action.

DC 6 contains many calcium channel blockers (CCBs) showing AMPK activation (in GC 1) and AKT inhibition (GC 6) signatures. Among the 15 CCBs in CMap, 9 (60%) were predicted to activate the AMPK pathway and 6 (40%) were predicted to inhibit the AKT pathway according to the d2gNES of the 2 pathways' representative genes *PRKAA2* and *AKT2* (Figure 2E). AMPK activation or AKT inhibition are known to lower blood lipid levels (Hagiwara et al., 2012; Zang et al., 2004). Although 4 of them are known to regulate lipid levels (perhexiline, nifedipine, verapamil, thapsigargin; see Table S4 for details) in mouse or cell-line

Figure 2. Scanning CMap Using CSs for Drug Repositioning

(A) Pipeline to generate d2gNES by scanning drug treatment signatures (DTSs) using CSs.

(B) Percentage of GSP and non-GSP DGIs in the top ranking d2gCSI intervals. Formula of d2gCSI is shown at bottom. See also Figure S3A.

(C) Filtering steps for predicted DGI: original DGIs generated by d2gNES (left), specific DGIs filtered by greater-than-average d2gNES and d2gCSI across all gene clusters (middle), and with d2gCSI ≥ 0.8 (right). Top enriched drug category and gene function in each drug and gene clusters are shown at bottom. Red asterisks indicate $p < 0.05$ for drug clusters. All enrichment in gene clusters are significant at $p < 0.05$.

(D) GSEA results of phenothiazine drugs that show similar expression profiles with *Bmpr1a*-knockout mouse embryonic stem cells (mESCs) and *Tgfbr2*-knockout mouse palate tissues. Drug structures are shown at top.

(E) The predicted effects of calcium channel blockers (CCBs) on AKT and AMPK pathways and clinical effects on blood lipids. The d2gNES values of CCBs for the *AKT2* and *PRKAA2* genes visualized by a heatmap and previously reported functions of CCBs toward AKT and AMPK (left). Blood indices of obese hypertensive female patients in CCB or another antihypertensive drug group are shown at right. ALP, alkaline phosphatase activity; ApoA-I, apolipoprotein A-I; ApoA-I/B, apolipoprotein A-I/apolipoprotein B ratio; ApoB, apolipoprotein B; BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TBIL, total bilirubin level; TC, total cholesterol; and TG, triglyceride.

Student's t test p values are indicated. Data are represented as means ± SDs.

See also Figures S3 and S7.

experiments, whether the other CCBs have antilipid effects is not known and lack relevant clinical evidence. We therefore analyzed a series of Han Chinese electronic medical records (Method Details) and found that among the obese (body mass index [BMI] $\geq$30) hypertensive female patients without clinical manifestation of hyperglycemia and hyperlipidemia, and having never taken hypoglycemic or hypolipidemic drugs, the CCB treatment group (n = 10, ages 61 $\pm$ 9 years, BMI 33 $\pm$ 2) showed improved blood lipid profiles compared to those taking other antihypertensive drugs (n = 40, ages 57 $\pm$ 8 years, BMI 33 $\pm$ 2) (lower total cholesterol [TC], t test p = 0.0944; lower low-density lipoprotein [LDL], p = 0.0151; lower apolipoprotein B [ApoB], p = 0.0023; higher apolipoprotein A-I/apolipoprotein B ratio [ApoA-I/B], p = 0.0072; Figure 2E). The female CCB treatment group showed lower total bilirubin levels (TBILs) and higher alkaline phosphatase (ALP) activity (t test p = 0.0386) than the groups taking other antihypertensive drugs (Figure 2E), which are expected effects of activating AMPK and inhibiting AKT (Kanazawa et al., 2008; Mukherjee and Rotwein, 2009; Wegiel et al., 2009). Except for ApoA-I and TBIL, other indices did not show the same patterns in males (CCB group: n = 12, ages 55 $\pm$ 11 years, BMI 32 $\pm$ 1; other antihypertensive drug group: n = 40, ages 52 $\pm$ 11 years, BMI 32 $\pm$ 2; Figure S3F), suggesting a gender-specific anti-blood lipid effect of CCBs in females.

Drug Repositioning with High Specificity

Because the specificity of a drug for a particular pathway or gene is a desirable or necessary feature of a good drug, we tested 3 scores based on the connection specificity index (CSI) (Fuxman Bass et al., 2013) to assess the gene-level specificity of a drug and selected drug-to-gene CSI (d2gCSI; Figures 2B and S3A; Method Details), which can best distinguish gold standard-positive (GSP) DGIs (Method Details) and non-GSP DGIs to evaluate drug-target specificity. Drug repositioning aims to find alternative targets of existing drugs, making it by definition not necessarily specific to a single target gene, yet possibly gene module (pathway level) specific. Therefore, using d2gCSI, we only kept GCs (representing unbiased auto-assembled pathways) having above-average specificity for different DCs (Figures 2C, S3B, and S3C; Method Details), then further required each DGI to have a moderate gene-level specificity of d2gCSI $\geq$0.8 against the rest of the 2,052 signaling genes, which were optimally enriched for GSPs (Figure S3D). This left 83,745 DGIs among 1,888 drugs and 1,276 signaling genes—1.2% of all of the possible DGIs—as specific DGIs. For example, the predicted inhibition of TGF-β by phenothiazines is determined as specific, while inhibition to AKT and activation to AMPK of CCBs are determined as non-specific. We called our method (including both d2gNES and d2gCSI steps described above) core signature drug-to-gene (csD2G) prediction and used the derived 83,745 csD2G DGIs for following the analysis and performance evaluation (Figure 2C; Method Details).

Performance Comparison to Existing Drug Repositioning Methods

Currently, the ping-pong algorithm (PPA) (Kutalik et al., 2008) and multivariate analysis of variance (MANOVA) are the 2 best methods for drug repositioning. They associate drug responses with gene expression or mutations in different cell lines to infer targets. Because the data types used in these previous methods are different from CMap and CSs used by csD2G, we considered the combination of data and corresponding method as a holistic strategy for comparison (Method Details). Drug and gene sets shared between strategies were used for performance comparison. Receiver operating characteristic (ROC) curves showed that csD2G markedly outperformed both PPA (Method Details) and MANOVA in drug-target prediction (area under the curve [AUC]: 0.743 for csD2G and 0.676 and 0.659 for PPA; AUC: 0.685 for csD2G, 0.486 for MANOVA; DeLong test p = 0.0367) (Figure 3A). When focusing only on GSP drug targets, csD2G also ranked GSP interactions higher than PPA and MANOVA (t test p = 0.0234 for MANOVA; Figure 3B; Table S3). Our csD2G also outperformed DeMAND (Woo et al., 2015), which was designed to elucidate genome-wide modes of action (MoA) proteins through assessing the dysregulation of their molecular interactions following drug treatments. It also outperformed the recently published L1000 gene experimental drug perturbation results, which were derived from *in vitro* cell lines on 1,000 landmark genes (Subramanian et al., 2017), in both coverage and accuracy based on the shared drugs between DeMAND, L1000, and our analysis (Figures 3C and 3D; Table S3).

Functional Validation of Repositioned Drugs

AMPK and AKT-FOXO pathways play critical roles in cancer, aging, and metabolic diseases (López-Otín et al., 2013). Although highly desirable, there is no small molecule found to specifically inhibit the insulin-like growth factor 1 (IGF1) pathway or to specifically enhance the activity of the AMPK pathway (Fontana and Partridge, 2015). We therefore examined the csD2G results for 29 genes in these 2 pathways (Figures 4A and S4A). Clustering analysis showed that DC-24 and DC-32 had strong and specific activating signatures for 2 AMPK genes (an AMPK α-catalytic subunit gene *PRKAA2* and a non-catalytic subunit gamma 3 *PRKAG3*) (Figure 4C) and the *FOXO4* gene (with a weak inhibitory signature for *AKT2*), respectively (Figures 4D and S4B), among all 27 genes of the pathways examined.

Of the 12 predicted AMPK activators, 3 (25%) are known to activate AMPK and 1 (8%) is known to extend lifespan (Figures 4C and 4D). In addition, 7 (56%) significantly showed the same patterns as the known AMPK activators 5-aminoimidazole-4-carboxamide 1-β-D-ribofuranoside (AICAR) and sorafenib (Figure 4C; Table S4). For the 22 predicted AKT-FOXO targeting drugs, 41% are known to influence the pathway, 27% are known to extend lifespan, and 69% showed the same patterns as the known AKT inhibitor perifosine (Figure 4D; Table S4). Because AMPK and AKT are involved in many cancers, we further tested the potential anticancer effect of these drugs and found that 8/12 (67%) and 20/22 (91%) of the predicted AMPK activators and AKT inhibitors have significant overlapping signatures, with at least 1 of 21 TCGA cancer signatures. Among them, 7/8 (88%) of the AMPK drugs show opposite expression patterns to at least 1 cancer, and 14/20 (70%) of the AKT drugs show the same patterns (Figures S4C and S4D), suggesting that both categories of predicted drugs may have anticancer potential.

Inhibiting AKT or activating FOXO and AMPK extends lifespan significantly in *C. elegans* (Greer et al., 2007; Paradis and

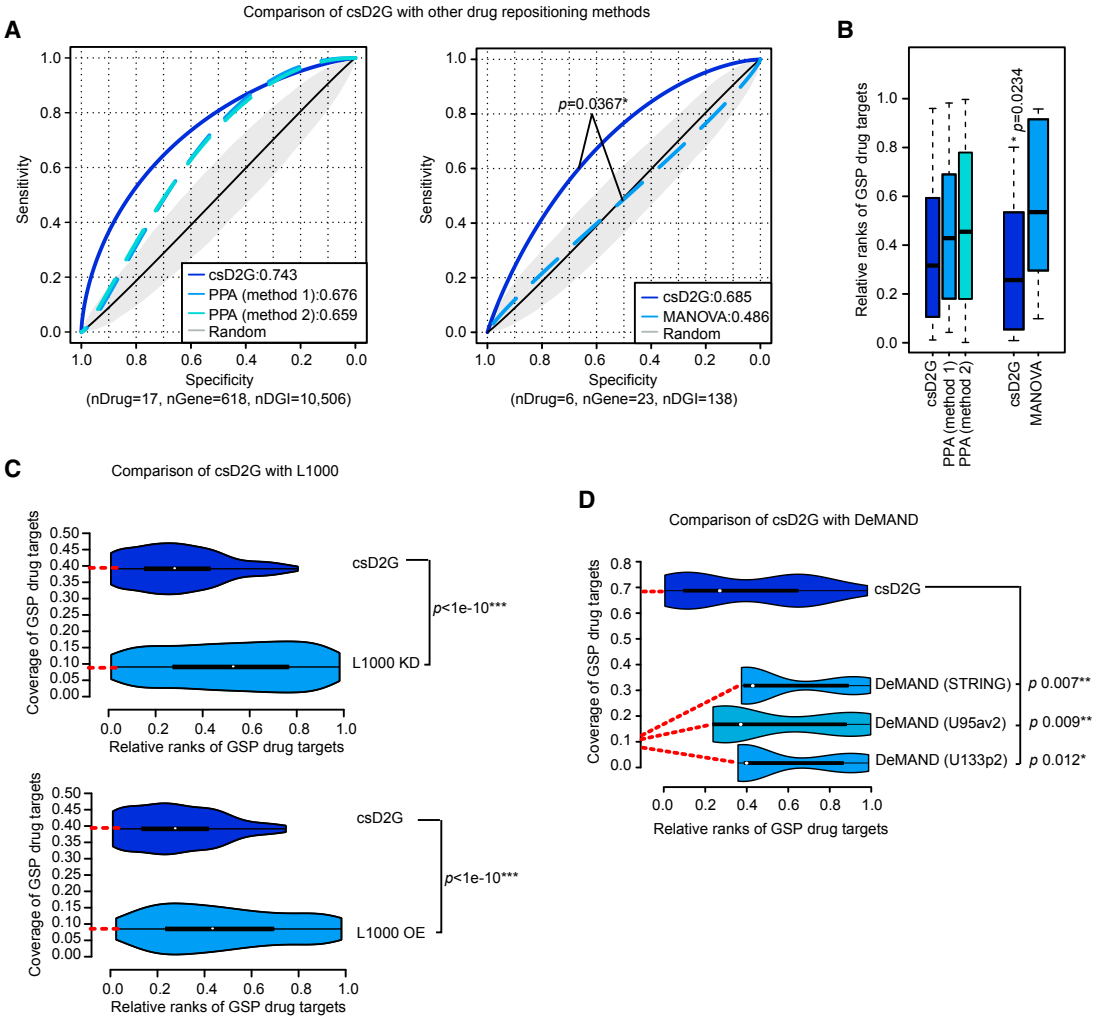


Figure 3. Performance Evaluation of csD2G and CSs

(A) ROC curves of csD2G, PPA, and MANOVA for drug-target prediction using shared drug and gene sets. Sensitivity and specificity are as defined in Method Details. DeLong test p values are indicated.

(B) Comparison of relative ranks of GSP targets predicted by csD2G, PPA, and MANOVA by GSP targets' relative ranks of shared drug and gene sets. Student's t test p values are indicated.

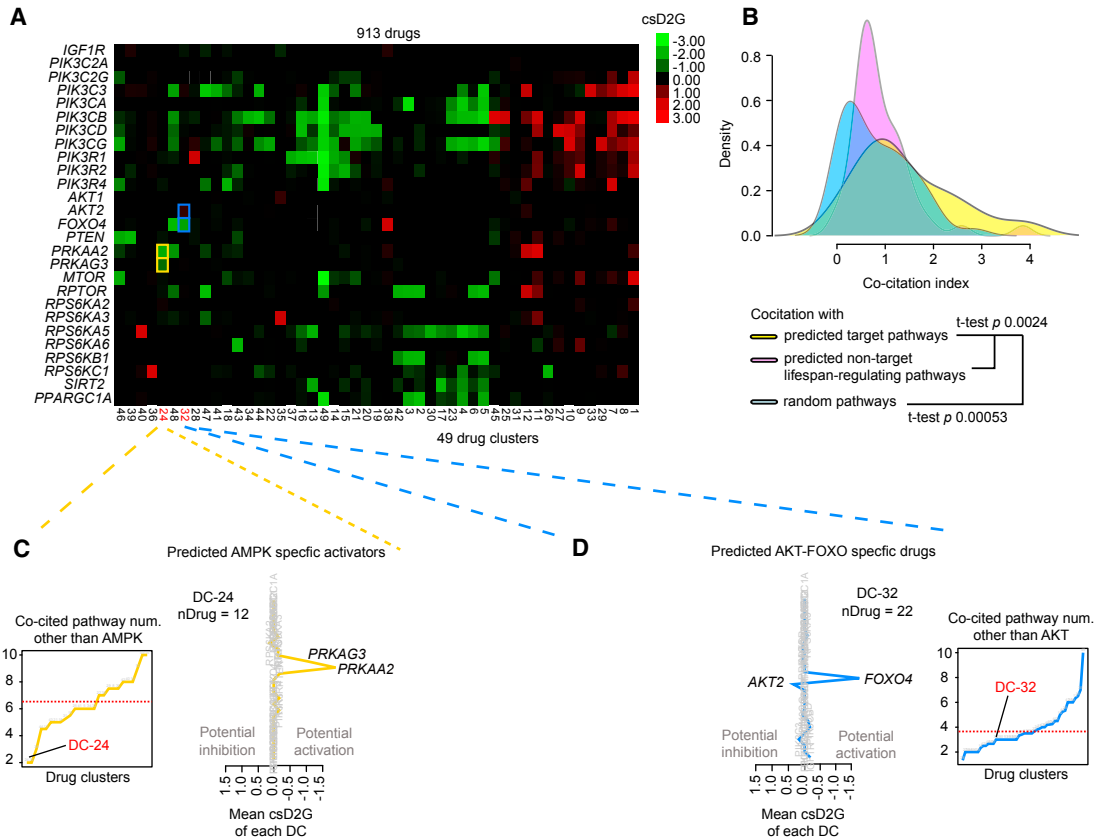
(C and D) Performance comparison of csD2G to L1000 (C) and DeMAND (D). Coverage (y axis) and relative ranks (x axis) of GSP targets predicted by these methods. For relative rank analysis, only predicted DGIs shared between csD2G and L1000 and between csD2G and DeMAND were used. For L1000, DGIs with score $\geq$80 in L1000 knockdown (KD) and overexpression (OE) experiments were used as L1000 predictions according to the original paper. For DeMAND, DGI predictions based on 3 protein-protein interaction datasets according to the original paper were used.

See also Figure S7 and Table S3.

Ruvkun, 1998). We reasoned that because lifespan extension cannot be expected to occur from random nonspecific effects, a lifespan assay is a good assessment for repositioned AKT-FOXO or AMPK modulators. Therefore, as functional validation of our predicted 12 specific AMPK activators and 24 AKT inhibitors, we randomly selected 7 drugs from these predictions (see Table S4 for details) and tested their lifespan effects in *C. elegans*. Six drugs (85.8%) significantly extended the lifespan (Figures S5A–S5C). The worm pharyngeal pumping rate was not changed after 3 days of drug treatment, excluding the possibility of lifespan extension by reduction in food intake (Figure S5E). We also tested 2 drugs not predicted to be involved in longevity as negative controls, and both showed no lifespan extension in the *C. elegans* lifespan assays (Figure S5D).

Consistent with the high specificity of these repositioned drugs, these DCs have significantly more literature co-citations (Qiao et al., 2013), with their predicted target pathways compared to randomly selected pathways or even to other lifespan-regulating pathways (Figure 4B). They also have relative fewer off-target pathways co-cited than the average expectation (Figures 4C and 4D).

To further test the specificity of the repositioned drugs, we tested their epistasis to their predicted target genes or pathways—we tested whether AKT-FOXO-axis (*daf-16*) and AMPK



Predicted drug	Known to perturb AMPK	Known to extend LS	Unknown drugs compared to AICAR or Sorafenib	d2gNES to *PRKAA2*	d2gNES to *PRKAG3*
Sirolimus	√	√	/	-1.910	-1.910
Perhexiline	√	×	/	-2.313	-2.313
Rolipram	√	×	/	-2.460	0
Procarbazine	×	×	sig. same	-1.888	0
Trihexyphenidyl	×	×	sig. same	-1.553	-1.553
Diperodon	×	×	sig. same	-1.781	-1.781
Etilefrine	×	×	sig. same	-1.985	0
Beclometasone	×	×	sig. same	-2.071	0
Verteporfin	×	×	sig. oppo.	-1.715	-1.715
AG-012559	×	×	sig. oppo.	-1.898	-1.898
Chlorzoxazone	×	×	not sig.	-1.377	-1.377
Midecamycin	×	×	not sig.	-1.751	0
	25%	8%	56%		

Predicted drug	Known to perturb AKT-FOXO	Known to extend LS	Unknown drugs compared to Perifosine	d2gNES to *AKT2*	d2gNES to *FOXO4*
Reserpine	√	√	/	0	-1.864
Tomatidine	√	√	/	0	-1.572
Cyproheptadine	√	√	/	1.667	-1.977
Fluoxetine	√	√	/	0	-1.814
Naringenin	√	√	/	0	-1.668
Mycophenolic acid	√	√	/	0	-1.927
Bupivacaine	√	×	/	0	-1.920
Beta escin	√	×	/	2.030	-1.641
Nocodazole	√	×	/	2.174	-1.657
Dicycloverine	×	×	sig. same	0	-1.629
Drofenine	×	×	sig. same	0	-1.850
Lasalocid	×	×	sig. same	1.761	-2.046
Lynestrenol	×	×	sig. same	0	-1.612
Zardaverine	×	×	sig. same	0	-1.707
Beclometasone	×	×	sig. same	0	-1.593
Isoetarine	×	×	sig. same	0	-1.456
Scoulerine	×	×	sig. same	0	-1.638
Benfluorex	×	×	sig. same	1.659	-1.640
Lactobionic acid	×	×	not sig.	0	-1.778
Fursultiamine	×	×	not sig.	0	-1.951
Nomegestrol	×	×	not sig.	0	-1.862
Pheneticillin	×	×	not sig.	0	-1.766
	41%	27%	69%		

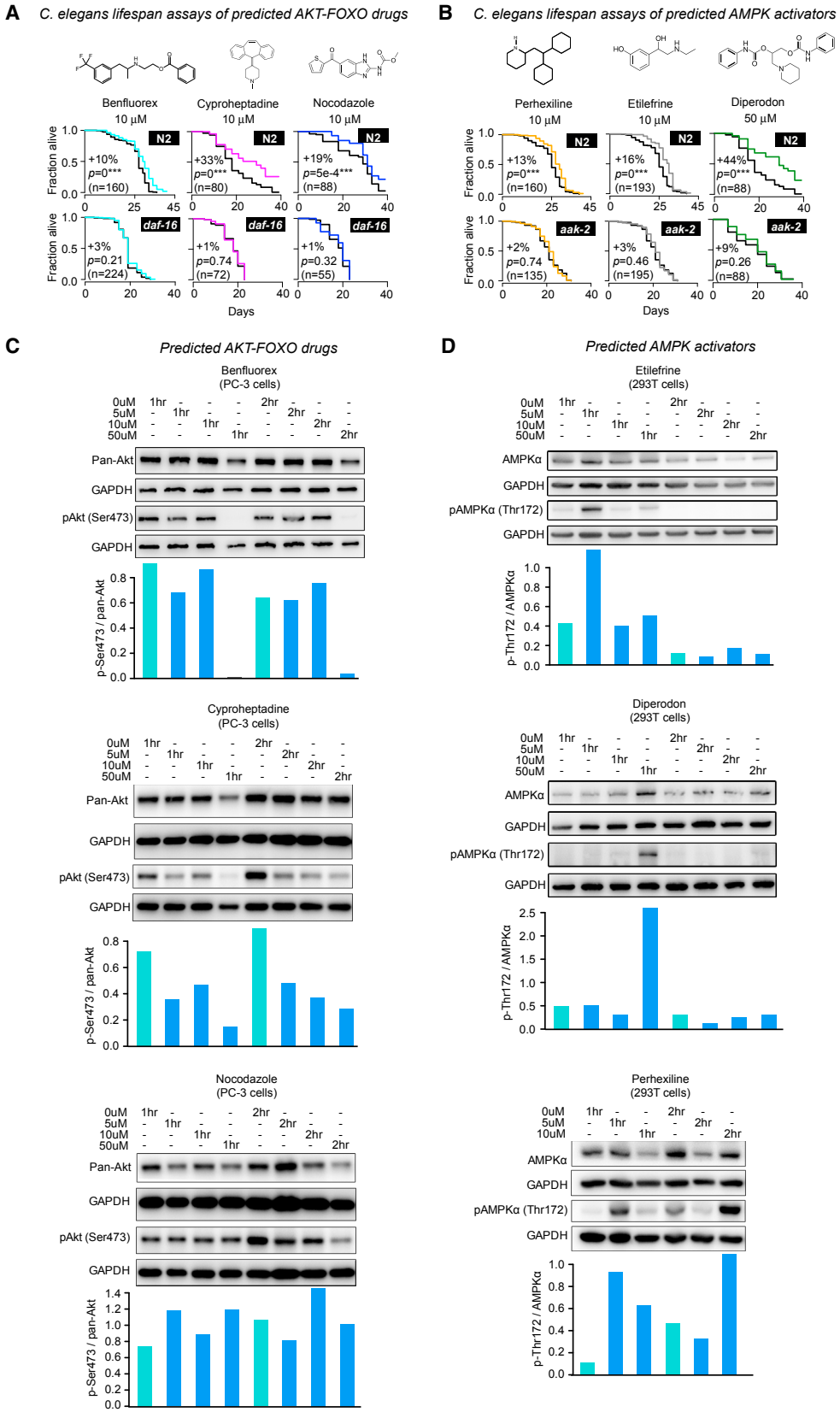
Figure 4. Using csD2G to Predict Drugs that Modulate AKT-FOXO and AMPK Pathways

(A) Heatmaps of csD2G values of drug clusters targeting lifespan-regulatory pathway genes.

(B) Density distributions of normalized co-citation index (CI) of drugs with predicted target pathways, non-target lifespan-regulating pathways, and random pathways. The 27 aging pathway genes with a csD2G value passing significance cutoff ($p < 0.05$ and false discovery rate [FDR] <0.25) to a drug were considered as a predicted target and otherwise as a predicted non-target. A total of 100 randomly selected pathway genes were used as basal background. For each drug, its CI to a pathway was normalized to correct for "star" pathways or drugs by first dividing it by the drug's mean CI to all of the genes in all of the pathways, and for each gene, dividing it by the gene's mean CI to all of the drugs.

(C and D) Drug clusters (DC-24 and DC-32) identified to specifically target AMPK (C) or AKT-FOXO (D) pathways. Average values of csD2G for each gene in DC-24 and DC-32 are shown at top. Numbers of co-cited pathways other than AMPK or AKT for each drug cluster are shown at right. Dotted red lines indicate mean of number of co-cited pathways. DC-24 and DC-32 are indicated in red. Bottom, previously reported targets and lifespan-modulating effects of these drugs, and expression profile comparison results by GSEA to known AKT or AMPK drugs (AICAR, sorafenib, and perifosine). csD2G values of the drugs to AMPK or AKT-FOXO pathway genes (*PRKAA2* and *PRKAG3*, *AKT2* and *FOXO4*) are marked on the right.

See also Figure S4 and Table S4.



(legend on next page)

(*aak-2*) mutants abolished their lifespan extension effects. As predicted, all 6 lifespan-extending drugs were indeed epistatic to *daf-16* or *aak-2* (Figures 5A and 5B). We also confirmed that all 6 directly regulated AKT and AMPK phosphorylation (Figures 5C, 5D, and S5F). In comparison, before removing the tissue-of-origin background for the 6 lifespan-extending drugs, we were unable to correctly predict their MoA, and some even showed the opposite results (Table S5). This further highlights the importance of removing tissue-of-origin signatures for reliable drug reposition.

To further experimentally test drug specificity, we tested the predicted AMPK activators and AKT inhibitors on the lifespan of both *aak-2* and *daf-16* mutants, as well as the *sir-2.1* mutant, representing an unrelated lifespan regulatory pathway. Probably due to the numerous cross-talks between the AKT and AMPK pathways, both *daf-16* and *aak-2* mutants can block the lifespan extension of 5 AKT or AMPK drugs (Figures 6 and S6), while 4 of the 5 tested drugs showed similar lifespan extension in the *sir-2.1* mutant in lifespan assays (Figures 6 and S6), which indicates their overall independence of the sirtuin pathway.

DISCUSSION

Conventional perturbation experiments performed in a particular cell or tissue capture the downstream responses that are often mostly context dependent. In contrast, CSs, which are naturally occurring signatures shared in multiple tissues, can reflect more upstream, core pathway changes. Thus, we developed a pipeline that takes advantage of the vast number of mutations in various tissue and cancer types to extract transcriptomic CSs without tissue-of-origin background, which enables efficient and accurate repositioning of drugs.

Compared to signatures derived from specific cells or tissues, CSs have the following advantages. First, by removing tissue-specific background, CSs can be integrated with many data resources from different tissues of origin. Second, while the 2,052 CSs we generated for signaling genes already cover a large quantity of druggable targets, our CS pipeline can be easily extended to any genes of interest. Overall, we generated 8,476 CSs for all TCGA mutated coding genes with sufficient mutation frequency and identified 179,004 specific DGIs (for 1,938 drugs) using these CSs (Figure S7A). Third, the *in vivo* naturally occurring mutations used from TCGA are more relevant to human disease contexts than are model organisms or artificial cell-based assays.

In addition to developing CS-based predictions, we further filtered these predictions by drug pathway level specificity through developing a CSI-based scoring system. These allow csD2Gs to outperform computational prediction tools PPA and MANOVA, as well as L1000 signature-based experimental results for drug target prediction, on either the 2,052 signaling genes or the total of 8,476 CS-bearing genes (Figures S7B and S7C).

We predicted and validated 7 drugs to specifically target the AKT-FOXO or AMPK pathway, which are well-known pathways regulating lifespan, but we await drugs to specifically target them. Using *C. elegans* lifespan assays, 6/7 of the predicted drug treatments were validated to affect this pathway by significantly extending lifespan, which was blocked by respective pathway mutants, and through modulating AKT and AMPK phosphoryalation.

For the purpose of forming a uniform analysis pipeline, we assumed the mutation signatures obtained from multiple samples tend to inhibit predicted target genes, which is likely to be correct for the majority of the cases but may be wrong in some cases. To ultimately resolve this, an experimental confirmation of the sign of prediction must be done before it is applied to drug development, as we did for the AKT inhibitors and AMPK activators and the calcium blockers.

A major conundrum in drug development is that the majority of drugs are developed based on *in vitro* cell assays and mouse models and consequently fail human clinical trials for either having no effect and/or undesirable effects in human subjects. This is exemplified by the abandoned Eli Lilly anti-β-amyloid drug, solanezumab, after hundreds of millions of dollars were spent in development (Carter and Lazar, 2018). Although TCGA data are generated in pathological conditions, the mutations still occurred in humans, and the transcriptome changes by the mutations are also *in vivo* human conditions. Furthermore, that the TCGA mutations are natural mutations from diverse genetic backgrounds makes our method applicable to a diverse human genetic population. More important, although not completely addressing the drug side effects, the drug specificity assessment method we developed here does reduce side effects by excluding drugs with non-specific targets. Therefore, in addition to the advantage of enabling the search for core perturbation signatures sans tissue-specific background, using signatures from *in vivo* human conditions leads to a higher probability of our repositioning predictions to be translatable human drug therapies, as compared to novel drug searches from *in vitro* cell line data or even mouse data.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Cell Lines
- METHOD DETAILS
  - Processing TCGA transcriptomes and generating CSs of signaling genes

Figure 5. Lifespan Effects of Drugs Predicted to Specifically Target AKT-FOXO or AMPK Pathway

(A and B) Representative survival curves of *C. elegans* under treatment of predicted AKT-FOXO targeting drugs (benfluorex, cyproheptadine, and nocodazole; A) or AMPK activators (perhexiline, etilefrine, and diperodon; B) for wild-type (N2) worms (upper) and pathway mutants (*daf-16* and *aak-2*). Corresponding solvents were used as blank controls for drugs (DMSO for benfluorex and nocodazole, water for cyproheptadine, perhexiline, etilefrine, and diperodon).

(C and D) Western blots of AKT or AMPK phosphorylation under treatment of predicted AKT-FOXO targeting drugs (C) or AMPK activators (D).

See also Figure S5 and Table S5.

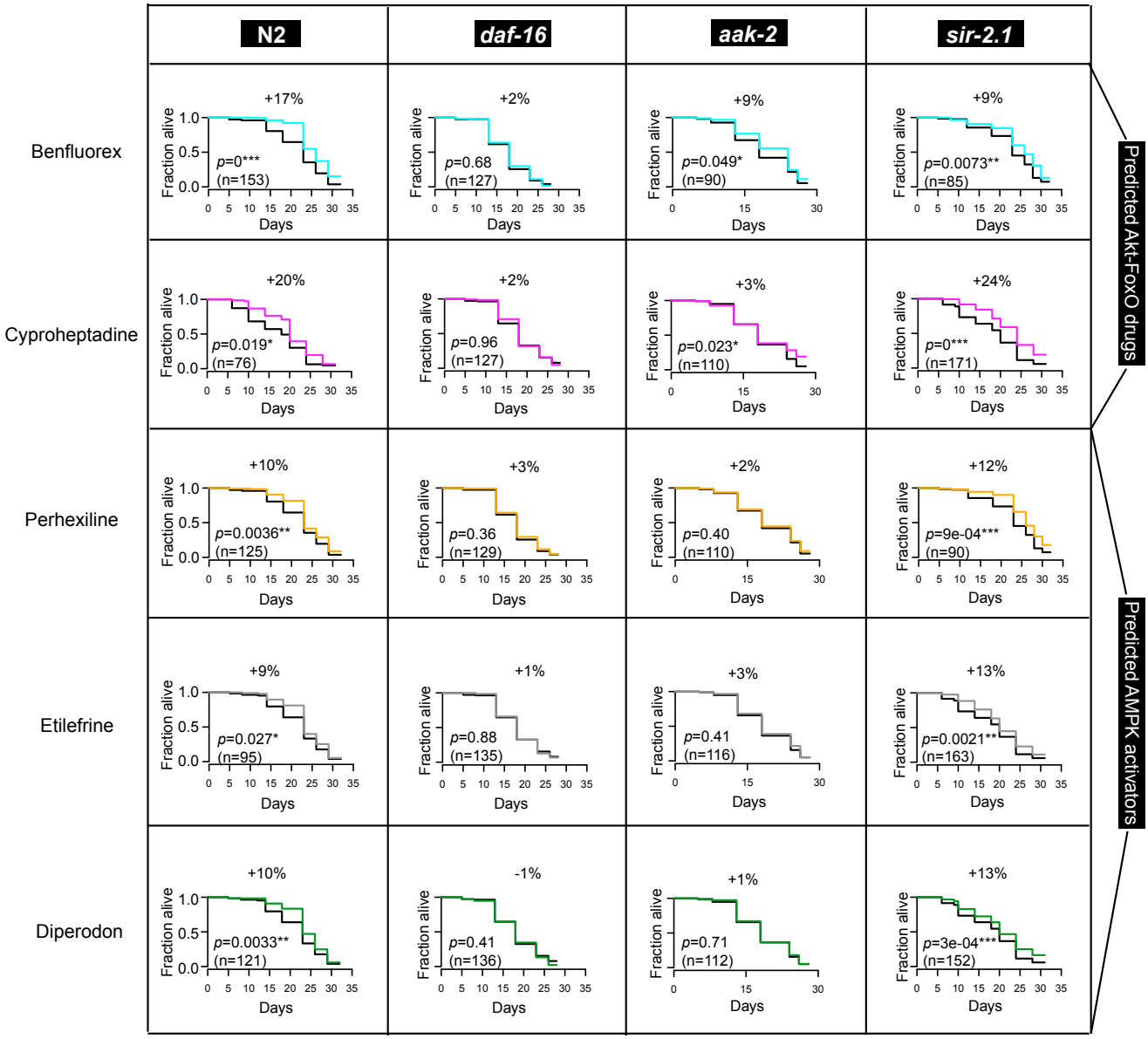


Figure 6. Independence of Predicted AKT-FOXO and AMPK Drugs on Sir-2.1

Survival curves of *C. elegans* under treatment of predicted AKT-FOXO targeting drugs or AMPK activators for wild-type (N2) worms, predicted pathway mutants *daf-16* and *aak-2*, and non-predicted pathway mutant *sir-2.1*. Corresponding solvents were used as blank controls for drugs (DMSO for benfluorex and perhexiline, water for cyproheptadine, etilefrine, and diperodon).

See also Figure S6.

- Generating DTSs, d2gNES and d2gCSI
- Drug-gene interaction specificity filter
- Evaluation of csD2G and CSs in drug-target prediction
- Directional comparison of CSs and conventional signatures
- Data analysis
- Worm lifespan assay
- Pharyngeal pumping
- Western blotting
- Electronic medical records

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.09.031.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (91749205, 31210103916, 91519330, and 91329302), China Ministry of Science and Technology (2015CB964803 and 2016YFE0108700), and the Chinese Academy of Sciences (XDB19020301 and XDA01010303, to J.-D.J.H.).

AUTHOR CONTRIBUTIONS

J.-D.J.H. conceived the project. C.X. performed the computational analysis and experimental validation with help from D.A., S.S., X.C., Y.Y., Y.C., N.S., W.C., S.Z., and Y.Z. J.-D.J.H. and C.X. wrote the manuscript with help from J.M.

DECLARATION OF INTERESTS

The authors declare no competing interests. Patent (application no. 201710208722.X) is pending.

Received: January 26, 2018
Revised: August 3, 2018
Accepted: September 10, 2018
Published: October 9, 2018; corrected online: October 15, 2019

REFERENCES

Akhurst, R.J., and Hata, A. (2012). Targeting the TGFβ signalling pathway in disease. Nat. Rev. Drug Discov. *11*, 790–811.

Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847.

Ashburn, T.T., and Thor, K.B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. Nat. Rev. Drug Discov. *3*, 673–683.

Behar, M., Barken, D., Werner, S.L., and Hoffmann, A. (2013). The dynamics of signaling as a pharmacological target. Cell *155*, 448–461.

Carter, P.J., and Lazar, G.A. (2018). Next generation antibody drugs: pursuit of the 'high-hanging fruit'. Nat. Rev. Drug Discov. *17*, 197–223.

Dudley, J.T., Sirota, M., Shenoy, M., Pai, R.K., Roedder, S., Chiang, A.P., Morgan, A.A., Sarwal, M.M., Pasricha, P.J., and Butte, A.J. (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci. Transl. Med. *3*, 96ra76.

Fontana, L., and Partridge, L. (2015). Promoting health and longevity through diet: from model organisms to humans. Cell *161*, 106–118.

Fuxman Bass, J.I., Diallo, A., Nelson, J., Soto, J.M., Myers, C.L., and Walhout, A.J. (2013). Using networks to measure similarity between genes: association index selection. Nat. Methods *10*, 1169–1176.

Gilson, M.K., Liu, T.Q., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. *44*, D1045–D1053.

Greer, E.L., Dowlatshahi, D., Banko, M.R., Villen, J., Hoang, K., Blanchard, D., Gygi, S.P., and Brunet, A. (2007). An AMPK-FOXO pathway mediates longevity induced by a novel method of dietary restriction in C. elegans. Curr. Biol. *17*, 1646–1656.

Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J.V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway, M.B., Eldred, J.M., et al. (2013). DGIdb: mining the druggable genome. Nat. Methods *10*, 1209–1210.

Hagiwara, A., Cornu, M., Cybulski, N., Polak, P., Betz, C., Trapani, F., Terracciano, L., Heim, M.H., Rüegg, M.A., and Hall, M.N. (2012). Hepatic mTORC2 activates glycolysis and lipogenesis through Akt, glucokinase, and SREBP1c. Cell Metab. *15*, 725–738.

Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabé, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S., et al.; International Cancer Genome Consortium (2010). International network of cancer genome projects. Nature *464*, 993–998.

Johnson, R., and Halder, G. (2014). The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. Nat. Rev. Drug Discov. *13*, 63–79.

Kanazawa, I., Yamaguchi, T., Yano, S., Yamauchi, M., and Sugimoto, T. (2008). Metformin enhances the differentiation and mineralization of osteoblastic MC3T3-E1 cells via AMP kinase activation as well as eNOS and BMP-2 expression. Biochem. Biophys. Res. Commun. *375*, 414–419.

Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T.H., von Mering, C., Jensen, L.J., and Bork, P. (2014). STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res. *42*, D401–D407.

Kutalik, Z., Beckmann, J.S., and Bergmann, S. (2008). A modular approach for integrative analysis of large-scale gene-expression and drug-response data. Nat. Biotechnol. *26*, 531–539.

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.

Lo-Coco, F., Avvisati, G., Vignetti, M., Thiede, C., Orlando, S.M., Iacobelli, S., Ferrara, F., Fazi, P., Cicconi, L., Di Bona, E., et al.; Gruppo Italiano Malattie Ematologiche dell'Adulto; German-Austrian Acute Myeloid Leukemia Study Group; Study Alliance Leukemia (2013). Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. N. Engl. J. Med. *369*, 111–121.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217.

Mavridis, L., and Mitchell, J.B.O. (2013). Predicting the protein targets for athletic performance-enhancing substances. J. Cheminformatics *5*, 31.

Miyoshi, T., Otsuka, F., Otani, H., Inagaki, K., Goto, J., Yamashita, M., Ogura, T., Iwasaki, Y., and Makino, H. (2008). Involvement of bone morphogenetic protein-4 in GH regulation by octreotide and bromocriptine in rat pituitary GH3 cells. J. Endocrinol. *197*, 159–169.

Mukherjee, A., and Rotwein, P. (2009). Akt promotes BMP2-mediated osteoblast differentiation and bone development. J. Cell Sci. *122*, 716–726.

Paradis, S., and Ruvkun, G. (1998). Caenorhabditis elegans Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. Genes Dev. *12*, 2488–2498.

Qiao, N., Huang, Y., Naveed, H., Green, C.D., and Han, J.D. (2013). CoCiter: an efficient tool to infer gene function by assessing the significance of literature co-citation. PLoS One *8*, e74074.

Siegel, P.M., and Massagué, J. (2003). Cytostatic and apoptotic actions of TGF-beta in homeostasis and cancer. Nat. Rev. Cancer *3*, 807–821.

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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

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.

Wegiel, B., Baty, C.J., Gallo, D., Csizmadia, E., Scott, J.R., Akhavan, A., Chin, B.Y., Kaczmarek, E., Alam, J., Bach, F.H., et al. (2009). Cell surface biliverdin reductase mediates biliverdin-induced anti-inflammatory effects via phosphatidylinositol 3-kinase and Akt. J. Biol. Chem. *284*, 21369–21378.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. *34*, D668–D672.

Woo, J.H., Shimoni, Y., Yang, W.S., Subramaniam, P., Iyer, A., Nicoletti, P., Martinez, M.R., Lopez, G., Mattioli, M., Realubit, R., et al. (2015). Elucidating Compound Mechanism of Action by Network Perturbation Analysis. Cell *162*, 441–451.

Wu, Z., Wang, Y., and Chen, L. (2013). Network-based drug repositioning. Mol. Biosyst. *9*, 1268–1281.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

Zang, M., Zuccollo, A., Hou, X., Nagata, D., Walsh, K., Herscovitz, H., Brecher, P., Ruderman, N.B., and Cohen, R.A. (2004). AMP-activated protein kinase is required for the lipid-lowering effect of metformin in insulin-resistant human HepG2 cells. J. Biol. Chem. *279*, 47898–47905.

Zhang, W., Liu, Y., Sun, N., Wang, D., Boyd-Kirkup, J., Dou, X., and Han, J.D. (2013). Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. Cell Rep. *4*, 542–553.

Zhuge, J., and Cederbaum, A.I. (2006). Increased toxicity by transforming growth factor-beta 1 in liver cells overexpressing CYP2E1. Free Radic. Biol. Med. *41*, 1100–1112.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
AKT (pan)	Cell Signaling Technology	Cat# 4691; RRID:AB_915783
phospho-AKT (Ser473)	Cell Signaling Technology	Cat# 4060; RRID:AB_2315049
AMPKα	Cell Signaling Technology	Cat# 2532; RRID:AB_330331
phospho-AMPKα (Thr172)	Cell Signaling Technology	Cat# 2535; RRID:AB_331250
GAPDH	Cell Signaling Technology	Cat# 5174; RRID:AB_10622025
Chemicals, Peptides, and Recombinant Proteins		
Benfluorex	Sigma	B7522-5G
Cyproheptadine	Sigma	279072-5G
Nocodazole	Sigma	M1404-2MG
Perhexiline	Sigma	SML0120-10MG
Etilefrine	Sigma	E2451000
Verteporfin	Sigma	SML0534-5MG
Diperodon	Sigma	D8536-5G
Deposited Data		
Non-tissue specific core signatures of human coding gene mutations	This paper; Mendeley Data	https://data.mendeley.com/datasets/8wznn46d43; http://www.picb.ac.cn/hanlab/csD2G
csD2G predictions of drug repositioning candidates	This paper; Mendeley Data	https://data.mendeley.com/datasets/8wznn46d43; http://www.picb.ac.cn/hanlab/csD2G
Experimental Models: Cell Lines		
HEK293T	SIBS Cellbank	GNHu44
PC-3	SIBS Cellbank	SCSP-532
Experimental Models: Organisms/Strains		
N2 (*C. elegans*)	Caenorhabditis Genetics Center	N/A
daf-16 (*C. elegans*)	Caenorhabditis Genetics Center	Cat# CF1038; RRID:WB-STRAIN:CF1038
aak-2 (*C. elegans*)	Caenorhabditis Genetics Center	RA8205
sir-2.1 (*C. elegans*)	Caenorhabditis Genetics Center	Cat# VC199; RRID:WB-STRAIN:VC199
eat-2 (*C. elegans*)	Caenorhabditis Genetics Center	Cat# DA1116; RRID:WB-STRAIN:DA1116
Software and Algorithms		
Non-tissue specific core signature processing and signature generation code	This paper	http://www.picb.ac.cn/hanlab/csD2G
Generation code of d2gNES	This paper	http://www.picb.ac.cn/hanlab/csD2G
Generation code of d2gCSI	This paper	http://www.picb.ac.cn/hanlab/csD2G

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jing-Dong Jackie Han (jdhan@picb.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

C. elegans wild-type N2 strain and *daf-16*, *aak-2*, *sir-2.1* mutant strains were used in lifespan assay and *eat-2* mutant strain was used in pharyngeal pumping assay. All animals are hermaphrodites. They are cultured in Nematode Growth Media agar plates in 20°C from

eggs to young adult stages. When drug treatment starts, animals are cultured in S-medium in 384-well plates with floxuridine in 20°C. *E. coli* OP50 is used to feed animals in both culture conditions. All mutants are backcrossed with at least two times.

Cell Lines

PC-3 and for HEK293T cells were used in western blotting experiments. PC-3 is a human prostate cancer cell line (male), cultured using F-12 medium (GIBCO, #21700075). HEK293T is a human embryonic kidney cell (the sex is unknown; lack of Y chromosome sequence suggesting the source was female), cultured using MEM medium (GIBCO, #41500034). Both cells are cultured in 37°C, with 5% CO2 in O2.

METHOD DETAILS

Processing TCGA transcriptomes and generating CSs of signaling genes

Signaling genes included: 1) human signal transduction pathway gene sets downloaded from KEGG pathway (https://www.genome.jp/kegg/pathway.html) and Reactome (https://www.reactome.org/) databases; 2) human transcription factors, enzymes, transporters, receptors and ion channels downloaded from Animal Transcription Factor Database (http://www.bioguo.org/AnimalTFDB/), the human DEPhOsphorylation Database (http://www.depod.bioss.uni-freiburg.de/) and IUPHAR/BPS database (http://www.guidetopharmacology.org/).

Processed RNA-seq data of 7,216 cancer patients and 653 normal samples (), and all "Level 3" mutation annotation files of 22 cancer types were downloaded from TCGA (https://cancergenome.nih.gov/). Expression profiles of cancer patients and cancer-type-matched normal samples were $\log_2$-transformed and quantile normalized together. Samples carrying silent mutations were discarded.

In each cancer, DEGs of a gene somatic mutation were first detected between normal samples and cancer samples that carry a specific mutation type (missense mutation, nonsense mutation, frameshift deletion or frameshift insertion, etc.) in the gene by the R package "limma" (Benjamini-Hochberg (BH)-corrected-$p < 0.05$) and filtered by expression fold changes ($\log_2$fold-change ≥ 2) as the mutation-type signature of the gene in a cancer type.

Then overlapping genes of the gene's mutation-type signatures in multiple cancers/tissues were fetched to eliminate tissue-of-origin background, as the CS for the gene with a mutation type. We used the following criteria: if a type of mutation for a gene occurred in only two tissues, we define the overlapping DEGs between the two tissues as a CS; if a mutation occurs in N tissue types where $N \geq 2$, we define the overlapping DEGs in $\geq$ N/2 tissue types as CSs (formula 1).

$$CS_i = \left\{ DEG \mid \sum_j \left(1\{DEG \in S_{ij}\}\right) \geq \frac{N_i}{2} \right\} \quad (1)$$

where CS_i defines the CS of the *ith* mutation type; S_{ij} defines the *jth* signature set of the *ith* mutation type, and N_i defines the number of cancers for the *ith* mutation type. Samples of kidney chromophobe cell carcinoma (KICH) and kidney renal clear cell carcinoma (KIRC) both belonging to kidney tissues, of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) both belonging to lung tissues as well as of uterine carcinosarcoma (UCS) and uterine corpus endometrial Carcinoma (UCEC) both belonging to uterine tissues were designated as from the same tissue of origin.

To capture the full spectrum of downstream effects from a gene, the union of CSs for all mutation types was defined as the CS for the gene (formula 2), where CS and i are as defined above.

$$CS = \mathrm{U}_i CS_i \quad (2)$$

$$CS_{final} = CS - DEG_{BG} \quad (3)$$

DEG_{BG} are tumor-background genes differentially expressed in more than 50% of cancer types and removed them from CSs (formula 3).

We also discarded a CS if samples used to derive the CS contained > 1 overlapping somatically mutated gene, to ensure one CS only represents a single gene's effect. Expression fold changes of CS genes were calculated in each tissue first and then averaged in all tissues included.

Generating DTSs, d2gNES and d2gCSI

Drug treatment microarray data were downloaded from CMap (https://portals.broadinstitute.org/cmap/) and processed from raw data to get expression profiles by the R package "affy." Samples of the same drugs in certain cell lines were compared with corresponding blank controls to obtain DEGs by "RankProd" ($p < 0.05$) as DTSs.

To generate d2gNES, CSs were separated into up- and downregulated gene sets, and scanned by the ranked lists of gene expression changes of all DTSs simultaneously through GSEA, to retain significant overlapping DGIs (nominal-$p < 0.05$). Then random gene sets were generated for each CS and the analysis was repeated 1,000 times for each DTS to obtain significant DGIs (FDR < 0.25).

DGIs were classified into two overlap patterns to infer direction of drug action based on GSEA NES of up- and downregulated CSs: 1) same pattern, where NESs of DTSs against upregulated CSs were positive and against downregulated CSs were negative; 2) opposite pattern, where NESs of CSs against upregulated DTSs were negative and against downregulated DTSs were positive. DGIs with conflicting directions were discarded. The maximum of absolute NESs against up- and downregulated CSs were defined as the score for an overlap. When a DTS only had significant overlap with either up- or downregulated CS gene but not both, the direction and score for the overlap were determined by NES of the more significant one. A d2gNES was assigned as the overlap score with a plus sign for a DGI with same pattern, while with a minus sign for opposite patterns.

CSI was transformed into 3 scores for selection to measure the specificity of DGIs: 1) d2gCSI for finding the specific target genes of a drug; 2) gene-drug CSI (g2dCsi) for finding the specific drugs of a target gene; 3) mean value of d2gCSI and g2dCsi. The one that best distinguished GSP and non-GSP DGIs was chosen. To generate these scores, PCCs were first calculated using expression fold changes of CS signature genes and corresponding values in DTSs, and then performed using the indicated formulas.

Drug-gene interaction specificity filter

We clustered DGIs based on d2gNES and d2gCSI values by BIC-SKmeans algorithm respectively to modularize DGIs. For DCs with sparse values, they were fetched and re-clustered. We only kept DC-GC pairs with > average d2gNES and d2gCSI across all GCs. DGIs in d2gNES and d2gCSI DC-GC pairs are merged if they are the same or if the genes are in the same complexes as annotated having "input," "reaction," "binding," "catalyze" or "complex" relationships in Reactome. As d2gCSI $\geq$ 0.8 enriched for the most GSPs, we further required DGIs in the remaining DC-GC pairs to have d2gCSI $\geq$ 0.8.

Evaluation of csD2G and CSs in drug-target prediction

GSP DGI sets were integrated from DrugBank (Wishart et al., 2006), DGIdb (Griffith et al., 2013), STITCH (Kuhn et al., 2014) and BindingDB (Gilson et al., 2016). For STITCH, only DGIs that passed a high-confidence score threshold (combined score $\geq$ 900) were used. For BindingDB, only DGIs that passed the rule-based thresholds of bioactivity indices (IC_{50}, Ki, Kd, EC_{50} or ED, etc.) according to a previous publication (Mavridis and Mitchell, 2013) were used.

The input data of gene expression and drug responses for PPA were downloaded from NCI-60 (https://dtp.cancer.gov/discovery_development/nci-60/) and used with the PPA's R package "isa2" in two ways. PPA (method 1): run PPA using all NCI-60 data and then fetch results containing shared drug and gene sets with csD2G's input. PPA (method 2): direct run PPA using NCI-60 data of the shared drug and gene sets with csD2G's input. MANOVA results were downloaded from GDSC (https://www.cancerrxgene.org/). P values of associations between drug responses and gene expression given by MANOVA were used to evaluate the strength of DGIs for MANOVA. As no drug action modes are predicted by PPA and MANOVA, only the targeting strength of csD2G, represented by the absolute values of prediction scores, was compared to these methods. Although a drug had DTSs in multiple cell lines, among significant connections predicted by csD2G, 94% DGIs occurred in only one cell line, or in multiple cell lines, but with the same direction in prediction score. Thus, to make csD2G comparable to PPA and MANOVA sans cell line information, we fetched the maximum of the absolute values of csD2G prediction scores of multiple cell lines for a DGI that occurred in multiple cell lines. Performance comparison between methods was based on shared drugs that are GSP and shared genes between the input of PPA and csD2G, or between MANOVA and csD2G. All drug-gene combination pairs of the shared drugs and genes were considered and pairs that were missed by a method were given a value of 0. To make normalized prediction power for different drugs, each drug's DGI prediction were first rank transformed, then the relative ranks of all drugs' DGIs were merged for ROC analysis for each method. Here, sensitivity is defined as the ratio of GSP DGIs' occurrence number versus that of all DGIs at various thresholds according to the rank-transformed prediction scores. As there is no available true negative DGI set, specificity is defined as the ratio of non-GSP DGIs' occurrence number versus that of all DGIs. The DeLong test was used to judge the significance of difference between prediction AUCs. For comparison with L1000, we queried 255 drugs prediction scores, a random selection from 606 shared drugs between L1000 and csD2G, in gene knockdown and overexpression perturbation sets.

We also compared CSs to conventional signatures from GEO, ENCODE ChIP-seq and regulator-regulon interaction networks. 860 microarray datasets of single gene perturbation were downloaded from GEO and processed from raw data to get expression profiles with the R package "affy." Samples of treatment and control groups were classified manually to detect DEGs by the R package "RankProd" ($p < 0.05$), as GEO single gene perturbation signatures. Outlier samples that were not clustered within the assigned groups were discarded. GEO signatures were separated into up- and downregulated gene sets to scan the ranked lists of gene expression changes of DTSs by GSEA to obtain GEO-signature-generated d2gNES (nominal-$p < 0.05$ and FDR < 0.25). Human TF ChIP-seq peaks were downloaded from ENCODE (https://www.encodeproject.org/comparative/regulation/) and assigned to nearest genes by Homer (http://homer.ucsd.edu/homer/) to get TF-target sets. Target sets of TFs were used to scan the ranked lists of gene expression changes of DTSs by GSEA to calculate ENCODE-generated d2gNES (nominal-$p < 0.05$ and FDR < 0.25). Protein activity changes under drug treatments inferred by VIPER were fetched from its original publication(Alvarez et al., 2016). The activity changing scores were used to evaluate the influence of drugs on genes of corresponding DGIs. For each drug, relative ranks of GSP targets were calculated for comparison between CSs and other signatures.

Directional comparison of CSs and conventional signatures

For comparison of CSs and conventional GEO signatures in expression change directions, CSs were separated into up- and down-regulated to scan the ranked list of gene expression changes of GEO signatures by GSEA to obtain CS-GEO signature pairs (C-G pairs) that significantly shared overlapping signature genes (nominal-p < 0.05 and FDR < 0.25). Direction of C-G pairs was defined as consistent when for mutation and RNAi treatments: NES > 0 for up versus upregulated genes and NES < 0 for down versus down-regulated genes, and opposite for OE treatment. Precision and sensitivity for CS capturing expression change directions of conventional signatures were calculated as below.

$$Precision = \frac{\sum True\ positive}{\sum Predictive\ condition\ positive}$$

$$Sensitivity = \frac{\sum True\ positive}{\sum Condition\ positive}$$

where *predictiv condition positive* are C-G pairs with consistent directions, *ture positive* are C-G pairs of the same genes and with consistent directions, and *condition positive* are all C-S pairs of the same genes being tested. Randomly-matched C-S pairs were used as random background.

Data analysis

The BIC-SKmeans clustering method was used to automatically determine optimal cluster numbers and then perform clustering analysis, as described previously(Zhang et al., 2013). Heatmaps were visualized by Java Treeview (http://jtreeview.sourceforge.net/) and the R package "pheatmap." Enrichment analyses of gene function and drug category were performed by hypergeometric tests for clusters of genes and drugs respectively. Gene sets of KEGG pathway and Gene Ontology (http://geneontology.org/) as well as drug category sets of DrugBank, ChEBI (https://www.ebi.ac.uk/chebi/) and USP drug classification system (http://www.usp.org/) were used for enrichment analyses. The Tanimoto coefficient between drugs was calculated with the Python module "Indigo" for chemical similarity analysis. Drug chemical structure strings of Simplified Molecular Input Line Entry System (SMILES) were downloaded from DrugBank. ROC curves were generated and smoothed by R package "pROC." Student's t test, binomial test, hypergeometric test, Fisher's exact test, log-rank test and the DeLong test were performed in R for statistical analysis.

Worm lifespan assay

Lifespans were carried out in liquid medium (S-complete medium) at 20°C in 384-well plates containing 6-15 *C. elegans* worms (N2, *daf-16*, *aak-2* or *sir-2.1* strain) in 150 μl, 10 mg ml^{-1} freshly prepared *E. coli* OP50 and 0.1 mg ml^{-1} 5-fluoro-2′-deoxyuridine (floxuridine) per well. Age-synchronized nematodes were seeded as L1 larvae in solid plates containing floxuridine and transferred to 384-well plates at adult day 0. Drug treatments started from adult day 1. Benfluorex, cyproheptadine, nocodazole, perhexiline, etilefrine, verteporfin and diperodon (Sigma) were dissolved in dimethyl sulfoxide (DMSO) at 500 × or water at 50 × final concentrations as stocks before use. DMSO or water with the same volumes of drugs was used as blank controls. Freshly prepared *E. coli* OP50 and drugs with working concentrations were added in 384-well plates every other day. The fraction of animals alive was based on movement in video recordings. Video capture was performed by microscope (Nikon) automatically every other day for each well. Plates were shaken by rotator for 3 min and exposed to bright light for 3 min before video capture. For the drugs' dosages, since most of them have never been tested on lifespan before, we checked their cell line experiment concentrations and tested using those concentrations first (Figure S5A). Then since all the reported concentrations are around 10∼50μM and showed some lifespan effects, we expanded the concentration gradients to 10, 100 and 1000μM to perform lifespan assays further. Each lifespan assay of a drug was arranged with 10 or 20 wells per experiment.

Pharyngeal pumping

Worms were grown in liquid medium (384-well plates) at 20°C from adult day 0 and treated with drugs or blank controls from adult day 1. Pumping rate was tested after 3-day treatment on adult days 4 and 5. Worms were transferred to bacteria-containing agar plates for 30 min and then the grinder movement within 10 s was counted. The *eat-2* strain was used as a control.

Western blotting

Cellular lysates were prepared by suspending 1 × 10^6 cells in 100 μL of 2 × SDS loading buffer for PC-3 cells and 40 μL for HEK293T cells after 12 hr serum-free starvation followed by drug treatment in indicated concentrations. Then cells were disrupted by sonication and extracted at 4°C for 20 min. Antibodies for AKT (pan) (#4691), phospho-AKT (Ser473) (#4060), AMPKα (#2532), phospho-AMPKα (Thr172) (#2535) and GAPDH (#5174) were purchased from Cell Signaling Technology. AKT western blotting was performed in PC-3 cells and insulin were added to medium for 10 min before cell extraction. AMPK western blotting was performed in HEK293T cells. ImageJ software was used for quantification.

Electronic medical records

Electronic medical records of Han Chinese patients were collected from RecData Technology Co, Ltd. To focus on the relationship between blood lipids and antihypertensive drugs, we selected hypertensive patients without medical history of hyperglycemia and hyperlipidemia and never administrated hypoglycemic or hypolipidemic drugs. All recorded blood indices were exacted from the records.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis

Statistical analyses were performed by R version 3.4.1. One-tailed Student's t test was performed for Figure 2E, Figure 3B and 3C, Figure S1C, Figure S2B, Figure S3F and Figure S7B and S7C; one-tailed binomial test was performed for Figure 1D; hypergeometric test was performed for Figure 2C; one-tailed DeLong test was performed for Figure 3A; log-rank test was performed for Figure 5A and 5B, Figure S5A-D and Figure S6. All significance p values were indicated in panels where performed. P value 0.05 was the significance cutoff. Number, age and gender of patients of electronic medical records were indicated in Figure 2E and Figure S3F. Number of *C. elegans* used for life assays was indicated in Figure 5A and 5B, Figure S5A-D and Figure S6.

DATA AND SOFTWARE AVAILABILITY

Non-tissue specific core signatures of human coding gene mutations and csD2G predictions of drug repositioning candidates can be found at https://data.mendeley.com/datasets/8wznn46d43 and http://www.picb.ac.cn/hanlab/csD2G.

Codes of generating core signatures, d2gNES and d2gCSI can be found at http://www.picb.ac.cn/hanlab/csD2G.

Update

Cell Reports

Volume 29, Issue 4, 22 October 2019, Page 1055

DOI: https://doi.org/10.1016/j.celrep.2019.10.023

Cell Reports
Correction

Accurate Drug Repositioning through Non-tissue-Specific Core Signatures from Cancer Transcriptomes

Chi Xu, Daosheng Ai, Dawei Shi, Shengbao Suo, Xingwei Chen, Yizhen Yan, Yaqiang Cao, Rui Zhang, Na Sun, Weizhong Chen, Joseph McDermott, Shiqiang Zhang, Yingying Zeng, and Jing-Dong Jackie Han*
*Correspondence: jdhan@picb.ac.cn
https://doi.org/10.1016/j.celrep.2019.10.023

(Cell Reports *25*, 523–535.e1–e5; October 9, 2018)
In the originally published version of this article, Dawei Shi (shidawei@picb.ac.cn) and Rui Zhang (zhangrui@picb.ac.cn) were mistakenly omitted from the author list. The corrected author list appears here and with the paper online.

The authors regret this error.