A gene essentiality signature for studying the mechanism of action of drugs

Wenyu Wang[1]*, Jie Bao[1], Shuyu Zheng[1], Shan Huang[2], Jehad Aldahdooh[1], Yinyin Wang[1], Johanna Eriksson[1], Ziaurrehman Tanoli[1], Jing Tang[1*]

1. **Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki 00290, Finland**
2. **Department of Radiation Oncology, Second Affiliated Hospital, Xi'an Jiaotong University, No.157, Xi Wu Road, Xi'an, 710004, Shaanxi, China**

***Corresponding author**

Abstract

Background:

Cancer drugs often kill cells independent of their putative targets, suggesting the limitation of existing drug target information. The lack of understanding of a drug's mechanism of action may prevent biomarker identification and ultimately lead to attrition in clinical trials. Current experimental strategies, such as binding affinity assays provide limited coverage at the proteome scale. In this study, we explored whether the integration of loss-of-function genetic and drug sensitivity screening data could define a novel signature to better understand the mechanisms of action of drugs.

Methods:

Loss-of-function genetic screening data was collected from the DepMap database, while drug sensitivity data were collected from three extensive screening studies, namely CTRP (n = 545), GDSC (n = 198), and PRISM (n = 1448). An L1 penalized regression model using the gene essentiality features was constructed for each drug to predict its sensitivity on multiple cell lines. The optimized model coefficients were then considered as the gene essentiality signature of the drug. We compared the gene essentiality signature with structure-based fingerprints and the gene expression signature of cancer drugs in predictions of their known targets. Finally, we applied the gene essentiality signature to predict the novel targets for a panel of noncancer drugs with potential anticancer efficacy.

Results:

We showed that the gene essentiality signature can predict drug targets and their downstream signaling pathways. Both supervised and unsupervised prediction accuracies were higher than those using chemical fingerprints and gene expression signatures. Pathway analyses of these gene essentiality signatures confirmed key mechanisms previously reported, including the EGFR signaling network for lapatinib, and DNA mismatch repair drugs. Finally, we showed that the gene essentiality signature of noncancer drugs can discover novel targets.

Conclusions:

Integrating drug sensitivity data and loss-of-function genetic data enables the construction of gene essentiality signatures that help discover drug targets and their downstream signaling pathways. We found novel targets for noncancer drugs that explain their anticancer efficacy, paving the way for the rational design of drug repurposing.

Introduction

Drugs affect cancer cells by binding to their target proteins, which trigger a cascade of downstream molecular modulations, leading to phenotypic changes such as viability. Although many candidate drugs are tested in oncological clinical trials, the full knowledge of their mechanisms of action is often lacking. For example, a recent study has shown that multiple drugs kill cancer cells even when their putative targets are depleted, suggesting the limitations of existing drug target information[1]. The incomplete drug target information also prevents successful biomarker identification and, in the end, leads to treatment failure[2].

Drug discovery process mainly falls into two categories, target-based or phenotype-based[3]. The target-based drug discovery starts with a target gene or a disease pathway, followed by the identification of proper chemical hits, with a minimal capacity to interrogate polypharmacological effects beyond the initially proposed targets[4,5]. In comparison, phenotype-based drug discovery is initialized with phenotypic screening to select candidate compounds, which are then forwarded for target deconvolution analysis[6]. Direct target deconvolution approaches, i.e., cell-free affinity experiments, are typically conducted in a hypothesis-driven manner with purified candidate proteins[3]. Recent advances such as CESTA[7] and PISA[8] allow interrogating proteome-wide drug-target interactions using cell lysate or intact cells. However, these cell-based target deconvolutional techniques, often time- and effort-consuming, have been conducted only for a few compounds that are biased towards certain cell lines. To our knowledge, an unbiased proteome-wide drug target deconvolution on a large panel of cell lines is currently unfeasible, even for individual drugs[9]. For example, a recent large-scale phenotype-based drug screening study showed that many approved noncancer drugs can efficiently kill cancer cells[10]. Despite the initial evidence, why and how these drugs affected cancer cells remains poorly understood due to insufficient mechanistic studies, partly due to the limited capacity of experimental approaches that are unlikely to meet the accumulating high demands of target identification. Therefore, the need to develop computational system medicine models to understand the mechanisms of actions is strong, especially for repurposing of noncancer drugs.

One of the common strategies to study the mechanisms of action of drugs is gene expression signature analysis. For example, the L1000 assay[11] or the PLATE-Seq assay[12] allows measuring drug-induced transcriptomics changes simultaneously for hundreds of genes or even the whole transcriptome. The gene expression signatures greatly enhanced the identification of drug targets and their downstream effectors, providing critical information to understand drug mechanisms at the pathway level[13-16]. Despite the success of many machine learning models in exploiting the gene expression signatures, it should be noted that the gene expression signature data has limited coverage. For example, the commonly used LINCS-L1000 dataset includes only 978 consensus genes for dozens of cell lines. PANACEA is a more recent study expected to increase the coverage of gene expressions, albeit still derived from a limited tumor context consisting of 25 cell lines[14].

On the other hand, loss-of-function genetic screens have recently been applied to study genetic dependencies in cancer, termed gene essentiality profiles[17]. In contrast to the drug-induced gene expression assays, the gene essentiality screening allows rapid genetic perturbations for a large number of genes and cell lines, thanks to the pooled CRISPR and RNAi techniques[18-21]. For example, the dependency mapping (DepMap) study has profiled

the gene essentiality profiles of more than 17k genes for a panel of 1070 cancer cell lines with CRISPR loss-of-function screening. Meanwhile, the same panel of cell lines has been tested with drug sensitivity screening in the major pharmacogenomics studies (i.e., CTRP and GDSC), which motivated several machine learning models to predict drug sensitivities with gene essentiality profiles[10,22]. Furthermore, the gene essentiality scores were also leveraged for drug target prediction. For example, Gonçalves et. al. tested the associations between gene essentiality with drug sensitivity and reported interesting overlaps between drug-associated genes and their putative targets[23]. Despite the initial evidence of using gene essentiality for drug-target prediction, several questions remain unanswered. First, the gene essentiality data was derived solely using CRISPR-based perturbations, while the RNAi-based genetic screening data was not evaluated. Second, the potential of multivariate machine learning method remains unexplored in essentiality-based drug target study. Thirdly, there is a lack of systematic comparison of the gene essentiality-based features against conventional drug features, such as gene expression signatures, or chemical fingerprints. Furthermore, it was unclear whether the essentiality-based drug target prediction approaches are valid for noncancer drugs.

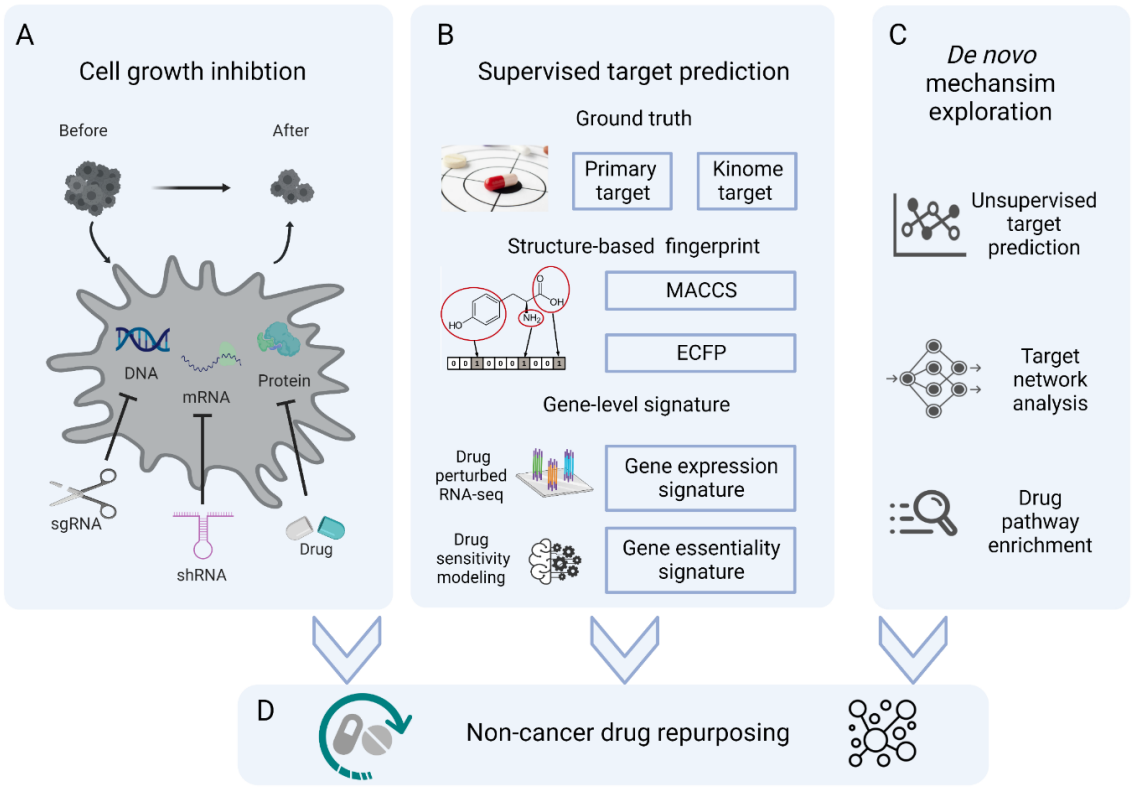


Figure 1. Schematic of the study design.
A. Gene essentiality and drug sensitivity are determined by cell growth inhibition after perturbations, which enables the establishment of gene essentiality features to predict drug sensitivity. sgRNA: single guide RNA; shRNA: short hairpin RNA.
B. The gene essentiality signatures are compared with chemical fingerprints and gene expression signatures in terms of their accuracy in supervised drug target prediction. MACCS: Molecular ACCess Systems keys fingerprint; ECFP: Extended Connectivity Fingerprint.

C. The gene essentiality signatures can also be used for *de novo* exploration of mechanisms of action, including unsupervised target prediction, target network analysis, and drug pathway association analysis.
D. Application of the gene essentiality signatures to study drug mechanism and rationalize noncancer drug repurposing.

In this study, we aimed to solve abovementioned limitations by deriving a novel gene essentiality signature for a drug. Our study was motivated by the complementarity of drug sensitivity and gene essentiality screens. Despite targeting molecules at different levels, both screens are designed to quantify the growth inhibition effects of perturbations (**Figure 1A**). We first developed a machine learning model to deconvolute drug sensitivity into its gene-level effects by leveraging CRISPR- and RNAi-based genetic perturbation data. The optimized model coefficients were derived as the gene essentiality signatures of the drugs. We evaluated the gene essentiality signatures as compared to gene expression signatures and chemical fingerprints, in both supervised drug target prediction and *de novo* mechanisms of action prediction (**Figure 1B-C**). Finally, we applied the gene essentiality signatures for noncancer drugs and determined their potential targets that may explain their anti-cancer efficacies (**Figure 1D**).

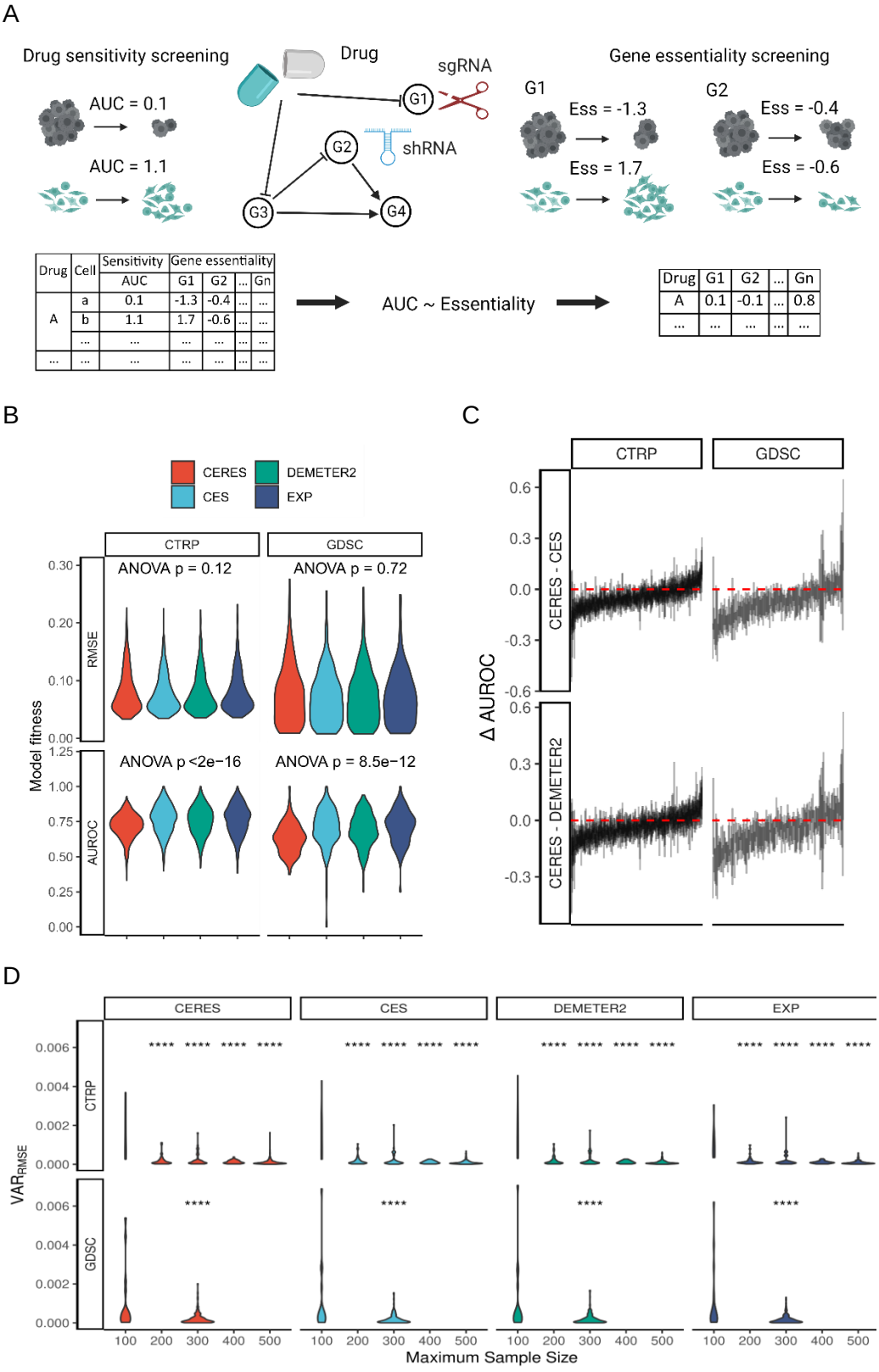
Results

Establishing the gene essentiality signatures for drugs

For each drug, we established the link between drug sensitivity and gene essentiality of the cell lines, based on the rationale that drug or genetic perturbations that affect the same cancer signaling pathways lead to similar growth inhibition effects (**Figure 2A**). To minimize the model complexity due to data noise, we employed an L1 penalized ridge regression to optimize the coefficients for all the gene essentiality features. We first evaluated whether the drug sensitivity regression models built on the gene essentiality features were generally predictive. We considered three variants of the gene essentiality features, namely CERES, DEMETER2, and CES (**Methods section**).

As shown in **Figure 2B and Supplementary Figure 1**, all three types of gene essentiality features yield comparable model fitness performance, to the same level of using the gene expression feature, which is known to be informative for drug sensitivity prediction[24,25]. The RMSE did not show significant differences among the three gene essentiality feature types (one-way ANOVA test P-value = 0.12 for CTRP and GDSC). For the CERES signature, we observed a decrease in AUROC compared to the gene expression feature ($\overline{\Delta AUROC}$ = -0.046, Tukey's test P-value = 6.8×10^{-11} for CTRP, and $\overline{\Delta AUROC}$ = -0.079, Tukey's test P-value < 2×10^{-16} for GDSC), while the CES and DEMETER2 gene essentiality features were equally predictable (all the Tukey's test P-values > 0.05, **Supplementary Table 1**). Furthermore, CERES and CES showed no difference when predicting 91.6% of the CTRP drugs and 91.4% of the GDSC drugs, while CERES and DEMETER2 showed no difference when predicting 95.4% of the CTRP and 95.5% of the GDSC drugs (**Figure 2C and Supplementary Tables 2-3**). Therefore, we considered that all three types of gene essentiality features have similar performance and took the average of their gene essentiality coefficients to derive a consensus gene essentiality signature for each drug, consisting of 10624 genes.
Furthermore, we found that the drugs with smaller sample sizes (i.e., number of cell lines) tend to yield a larger variance of RMSE in the cross-validation (**Figure 2D**), especially if

sample size is less than 100 (all Dunnett test P-value < 0.001, **Supplementary Table 4**). To ensure the robustness of the gene essentiality signatures, we considered only the drugs with more than 100 cell lines for the subsequent analyses.



Drug	Cell	Sensitivity	Gene essentiality			
		AUC	G1	G2	...	Gn
A	a	0.1	-1.3	-0.4	...	...
	b	1.1	1.7	-0.6	...	...
	...	...	...	...	...	...
...	...	...	...	...	...	...

Drug	G1	G2	...	Gn
A	0.1	-0.1	...	0.8
...	...	...	...	...

Figure 2. Establishing gene essentiality signatures of drugs.

A. Rationale for modeling of drug sensitivity with gene essentiality profiles. In drug sensitivity screening, a drug's effect on cell growth starts from its target binding, which triggers a cascade of downstream signaling alterations. In gene essentiality screening, genetic perturbations are made by either shRNA-mediated knocking down or sgRNA-mediated knocking out, after which the cell growth inhibition is measured. For each drug, a drug sensitivity prediction model across multiple cell lines can be built by considering their gene essentiality profiles as features.
B. Model fitness of gene essentiality feature types, including CERES, CES, DEMETER2, as compared to that of gene expression features (EXP).
C. Differences in AUROC when comparing CERES, CES, and DEMETER2. Each vertical line represents the 95% confidence interval of the difference for a drug.
D. Variance of RMSE for each feature type grouped by drugs with different sample sizes. ****: $p < 0.0001$ by Dunnett's test using maximum sample size = 100 as the reference.

The gene essentiality signatures predict targets of cancer drugs

We explored how the gene essentiality signatures can predict drug targets in a supervised setting (**Figure 1B**). We employed a similarity-based method that showed superior performance in a recent DREAM Challenge competition for drug target predictions (**Methods section**)[14]. The performance of the gene essentiality signature was compared with that of the gene expression signatures, which are the transcriptomics changes after drug treatment. We also evaluated the MACCS and ECFP fingerprints generated using the chemical structures of the drugs.

We observed that the gene essentiality signature achieved top performance across CTRP and GDSC datasets (**Figure 3A and Supplementary Table 5**). For example, for the GDSC dataset, when considering the primary targets of drugs as the ground truth, the gene essentiality signature achieved an average AUC of 0.61, as compared to 0.55, 0.47 and 0.43 for the gene expression signature, ECFP and MACCS fingerprints, respectively (**Figure 3A, upper panel**). In addition, we also examined the performance in predicting a broader spectrum of secondary targets. We considered the kinome-wide binding affinity scores from the DTC database as the ground truth and examined the prediction performance by binarizing the binding affinity scores with a threshold of 0.4. We observed a similar result for both the CTRP and GDSC datasets, where the gene essentiality signatures generally outperformed the other types of drug signatures (**Figure 3A, lower panel**).

We are also interested in whether the gene essentiality signatures can prioritize drug pairs with shared drug targets and sensitivity profiles. We reasoned that such drug pairs are potential candidates for drug repurposing, as their drug sensitivity similarities are well supported by the shared mechanisms of action. We determined such drug pairs as the ground truth, and then examined whether they can be prioritized by the gene essentiality signatures. As shown in **Supplementary Figure 2** and **Supplementary Table 6**, we found that the gene essentiality signatures successfully prioritized the ground truth drug pairs in both CTRP (median percentile = 93.4% for n = 150 drug pairs when considering the primary targets and median percentile = 97.6% for n = 70 drug pairs when considering the kinome-wide targets), and GDSC datasets (median percentile = 90.8% for n = 29 dug pairs when considering the primary targets and median similarity percentile = 95.4% for n = 23 drug pairs when considering kinome-wide targets).

The prioritization of the top 16 ground truth drug pairs by their predicted targets using different drug signatures is shown in Figure 3B. The gene essentiality signature was able to predict many of the top ground truth drug pairs, which the other signature features failed to predict, indicating that the gene essentiality signature could be an informative new feature to improve drug target prediction. For example, the gene essentiality feature was able to prioritize (similarity percentiles ≥ 97.5%) all the mTOR inhibitor (AZD8055, KU-0063794, and sirolimus) drug pairs, while the chemical fingerprints were able to predict only the AZD8055-KU-0063794 pair, that share the same chemical pyridopyrimidine structure[26], which is absent in sirolimus. In contrast, the gene expression signature prioritized the sirolimus-AZD8055 pair (similarity percentile 93.5%) but not sirolimus-KU-0063794 or AZD8055-KU-0063794 pairs (similarity percentiles 74% and 84%, respectively).

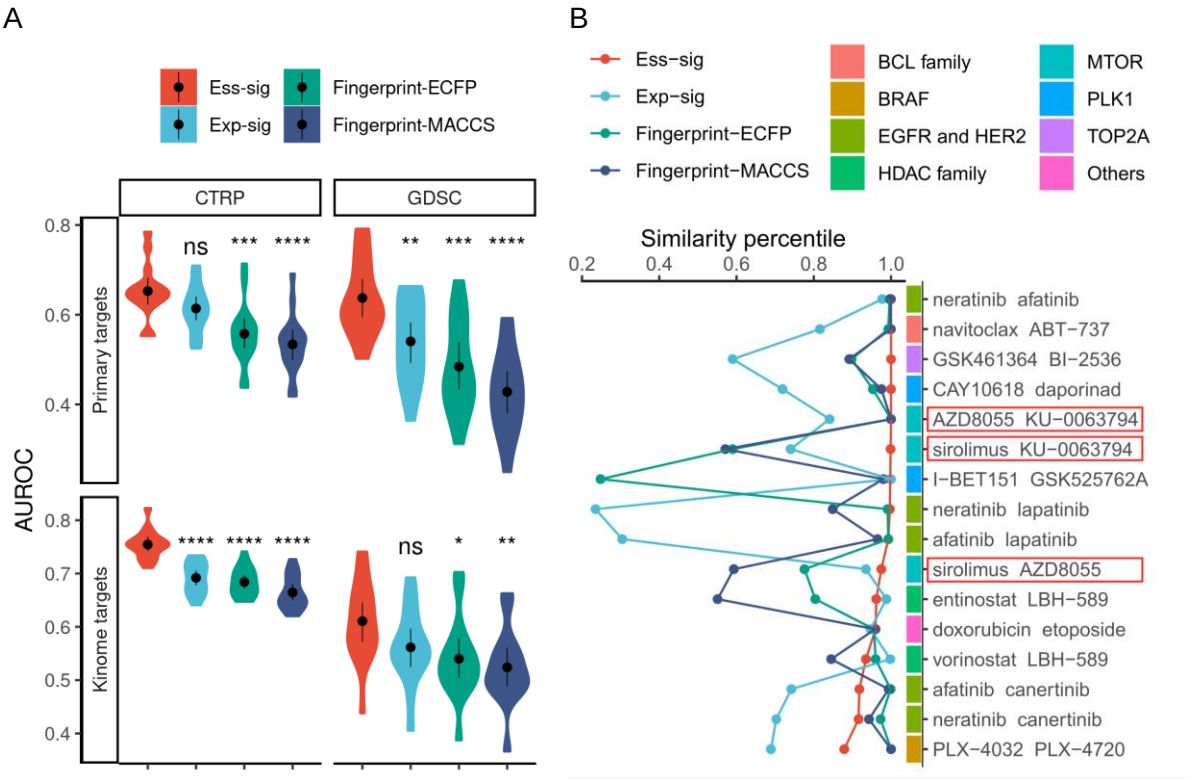


Figure 3. Gene essentiality signatures improve drug target prediction.
A. Accuracy of supervised drug target prediction.
ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001 by Dunnett's test with the gene essentiality signature as the reference. Ess-sig: gene essentiality signature, Exp-sig: gene expression signature.

B. Rank of similarity for the top repurposable drug pairs.

Gene essentiality signature facilitates *de novo* exploration of mechanisms of action
After confirming the improved performance of gene essentiality signatures in the supervised target prediction, we explored if the signatures could predict drug targets *de novo* without training data, even though the gene essentiality signatures were determined without leveraging drug target information.

Note that the gene essentiality signature is a vector of genes with feature importance values, so we asked whether the targets of drugs can be identified as top genes in the signature. As shown in **Figure 4A**, many drugs had their primary targets enriched in their top essentiality signatures genes (Fisher's test $P = 1.2\times10^{-7}$). For example, the primary targets for 30.4% of CTRP drugs and 57.1% of GDSC drugs were found within the top 50 genes in the gene essentiality signatures. The primary targets showed no enrichment in the bottom genes (Fisher's test $P > 0.05$) of the gene essentiality signatures. This suggests that most drugs induced inhibitory effects on their targets, consistent with the inhibition effects induced by the gene essentiality screens. In contrast, drug targets were much less identified by the top or the bottom genes of the gene expression signatures (Fisher test P-value $< 2.2\times10^{-16}$). For example, only 3.7% and 3.1% of the CTRP drugs had their primary targets identified in the top 50 and bottom 50 genes, respectively. Likewise, the gene essentiality signature also showed higher accuracy when predicting drug targets, with an average AUC of 0.66 (CTRP) and 0.78 (GDSC), compared to an AUC of 0.47 (CTRP) and 0.53 (GDSC) for gene expression signatures (t-test P-value = 6.0×10^{-5} and 9.0×10^{-4} for CTRP and GDSC, respectively, **Figure 4B-4C and Supplementary Figure 4**).

We next checked whether the gene essentiality signatures can shed more light on the downstream pathways affected by drugs. The top and the bottom genes (n = 50) recapture not only primary targets but also their neighboring genes in the PPI network (Supplementary Figure 4). For example, the top 50 and bottom 50 genes together identify the first-degree neighbors of targets for 33.5% of CTRP drugs (18.6% and 14.9% in the top and the bottom 50 genes, respectively). Take an EGFR inhibitor lapatinib for example, the top 10 genes from the gene essentiality signature not only identified the primary targets, namely ERBB1(EGFR) and ERBB2 (HER2), but also revealed several important downstream genes in the EGFR signaling pathway, including ERBB3[27], GAB1[28,29], and SOS1[29] (**Figure 4D**). This showed the gene essentiality signature's ability to reveal both binding targets and downstream changes upon initial target gene inhibition. In comparison, the top and bottom genes in the gene expression signatures identify mostly the 1st-degree neighbors (52.6%) instead of the direct targets (6.7%) of the drugs, suggesting that drugs often do not affect the expression of their target gene but only the downstream factors.

To further map the gene essentiality signatures to biological processes, we implemented a gene set enrichment analysis using the gene essentiality signatures as input. For each drug, we determined the pathways that are enriched in the gene essentiality signature, with absolute normalized enrichment score (NES) > 1.5 and $P_{adj} < 0.05$ (**Supplementary Tables 7 - 8**). Here we highlighted the drugs for which the DNA mismatch repair pathways were enriched (**Table 1**), as the pathway may regulate tumor mutation burden and immune cell infiltration[30,31]. Interestingly, many drugs in Table 1 are known to impact cell cycle and genome integrity. In particular, we found that all three approved drugs in Table 1 (talazoparib, cytarabine, and irinotecan) are under investigation in combination with immunotherapy in recent clinical trials (**Supplementary Table 9**). For example, the PARP inhibitor talazoparib is combined with immune checkpoint inhibitors in multiple cancers, including urothelial carcinoma (NCT04678362), lung cancer (NCT04173507), as well as breast and ovarian cancer (NCT03330405). The other compounds enriched in the DNA mismatch repair pathway may also be worthy of further exploration as sensitizers for immune checkpoint inhibitors.

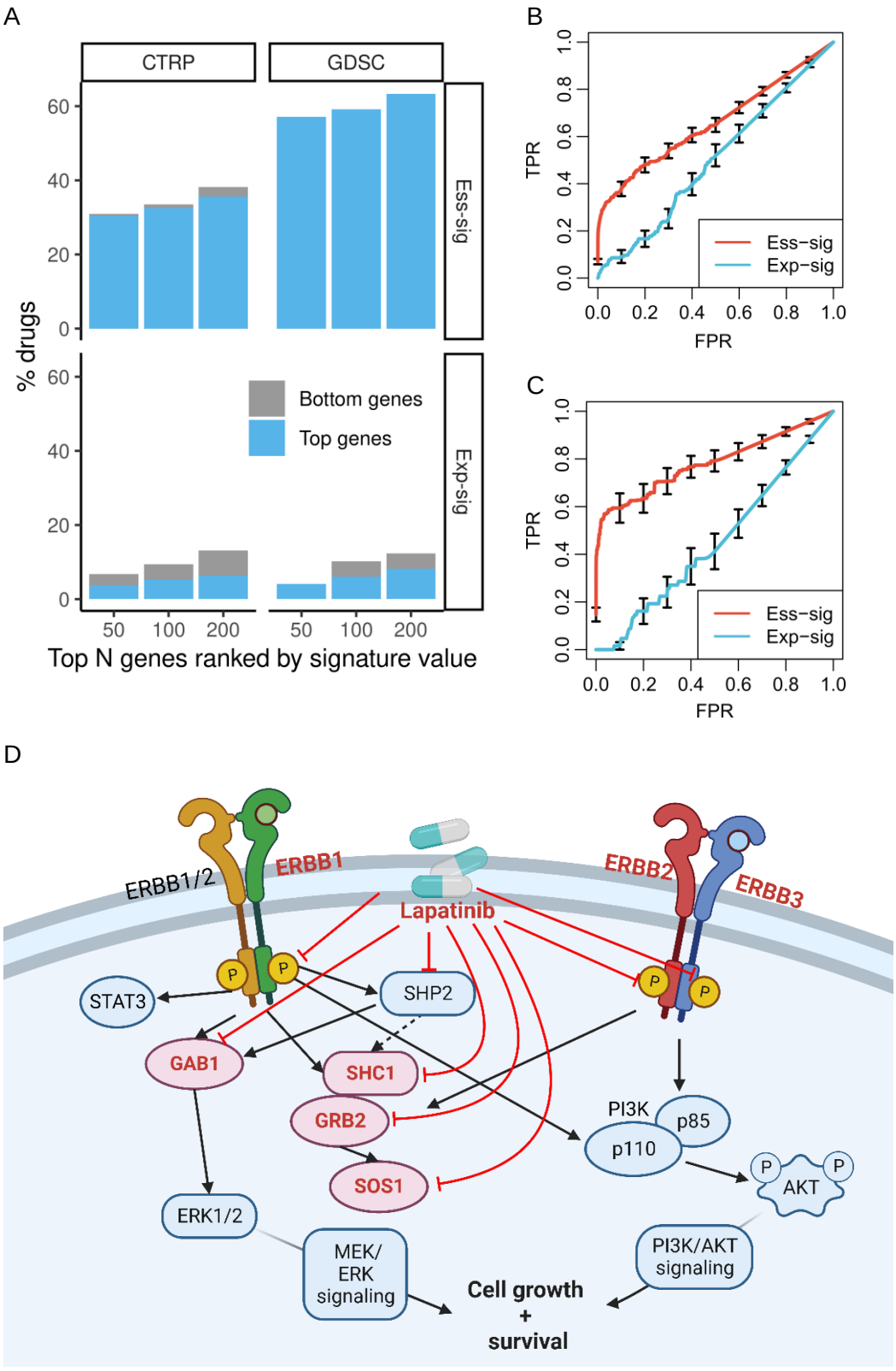


Figure 4. *De novo* target prediction and pathway analysis using the gene essentiality signature.

A. Percentage of drugs with putative targets recovered by the top and bottom genes in their gene signatures.

B-C. ROC curves of *de novo* prediction of primary targets for CTRP and GDSC, respectively.

D. Top genes in the gene essentiality signature of lapatinib are highlighted in the EGFR pathway.

Table 1 List of drugs for which the DNA mismatch repair pathway was enriched in their gene essentiality signatures.

DRUG NAME	P_{adj}	ES	NES	REPORTED MECHANISM	DRUG TYPE
AZD7762	6.11×10^{-5}	0.75	2.40	Cell cycle	Targeted
MK-8776	4.95×10^{-4}	0.71	2.21	Cell cycle	Targeted
Wee1 Inhibitor	1.24×10^{-3}	0.71	2.26	Cell cycle	Targeted
MK-1775	2.97×10^{-3}	0.67	2.07	Cell cycle	Targeted
Talazoparib	3.40×10^{-3}	0.70	2.17	Genome integrity	Targeted
AZD6738	8.53×10^{-3}	0.64	2.05	Genome integrity	Targeted
VE-822	8.54×10^{-3}	0.66	2.04	Genome integrity	Targeted
Cytarabine	3.51×10^{-2}	0.61	1.96	Other	Chemotherapeutic
Irinotecan	4.18×10^{-2}	0.62	1.95	DNA replication	Chemotherapeutic

Gene essentiality signatures help target prediction for the PRISM drugs

Profiling Relative Inhibition Simultaneously in Mixture (PRISM) is a recent technology that increases the throughput of drug sensitivity screening to thousands of drugs across hundreds of cell lines[32]. With this technology researchers have discovered multiple drugs with efficacy in killing cancer cells in addition to their initial disease indications[10]. Over half (53.4%) of the 1448 compounds probed in their multi-dose secondary screen are developed for indications other than cancer, mostly already approved[10] (n = 406). However, these noncancer drugs, such as antibiotics and anti-parasite drugs, often lack relevant cancer target information (**Figure 5A and Supplementary Table 9**). Therefore, we tested whether the gene essentiality signature can be applied to the PRISM data to elucidate the mechanisms of action of these promising drugs for cancer treatment.

We first evaluated whether a reliable gene essentiality signature could be determined from the PRISM drug sensitivity profiles. Comparing the CTRP and GDSC datasets, we observed, on average lower model fitness in the PRISM dataset (Dunnett's test, all P-values < 0.001, **Supplementary Table 10**). The decreased model fitness may be partly due to the inclusion of noncancer drugs in the screening library. As shown in **Figure 5B**, noncancer drugs were more likely to have lower drug sensitivities (median AUC of 1.1 compared to 0.71 and 0.90 for chemotherapeutic and targeted drugs, Dunnett's test P-value = 1.6×10^{-15} and 8.9×10^{-16}), which was associated with lower model fitness (Spearman correlation = -0.46, P-value < 2.2×10^{-16}). In addition, compared to CTRP and GDSC, we observed that the model fitness decreased, even for the same set of cancer drugs (n = 44, **Supplementary Table 11**). Moreover, the gene essentiality signatures derived from PRISM were less consistent with those derived from CTRP and GDSC ($\overline{\mathrm{COR}}(CTRP, PRISM) = 0.17, \overline{\mathrm{COR}}(GDSC,\ PRISM) = 0.13$, $\overline{\mathrm{COR}}(CTRP, GDSC) = 0.33$, Dunnett's test P-value = 9.4×10^{-8} and 1.6×10^{-9} for CTRP-PRISM and GDSC-PRISM, respectively).

We next evaluated the target prediction accuracy of the gene essentiality signatures, using the existing target annotation of the PRISM drugs as the ground truth (**Figure 5** and **Supplementary Table 12**). We observed similar performance of the different signatures in

predicting the targets, with median AUC of 0.72, 0.72, 0.71, and 0.69 for gene essentiality signatures, gene expression signatures, ECFP and MACCS fingerprints, respectively (**Figure 5C,** left panel, n = 805). The lack of difference between gene essentiality signatures and ECFP fingerprints may be partly due to the drugs for which the model fitness is poor (**Figure 5C**, right panel, n = 200). Indeed, when we evaluated only the top 200 drugs with higher model fitness, we observed more significant differences between the gene essentiality signatures and the chemical fingerprints (**Figure 5C**, middle panel, n = 200.

Likewise, the quality of the gene essentiality signature also had an impact on the *de novo* target predictions. As shown in **Figure 5D**, we observed that the gene essentiality signature outperformed the gene expression signatures only in drugs with better model fitness (median AUC of 0.72 versus 0.61, t-test P-value = 1.7×10^{-3}). For the drugs with poor model fitness, the *de novo* target prediction accuracy dropped to random, similar to the gene expression signature (median AUC 0.50). Together our results support the validity of gene essentiality signatures for PRISM drugs with top model fitness.

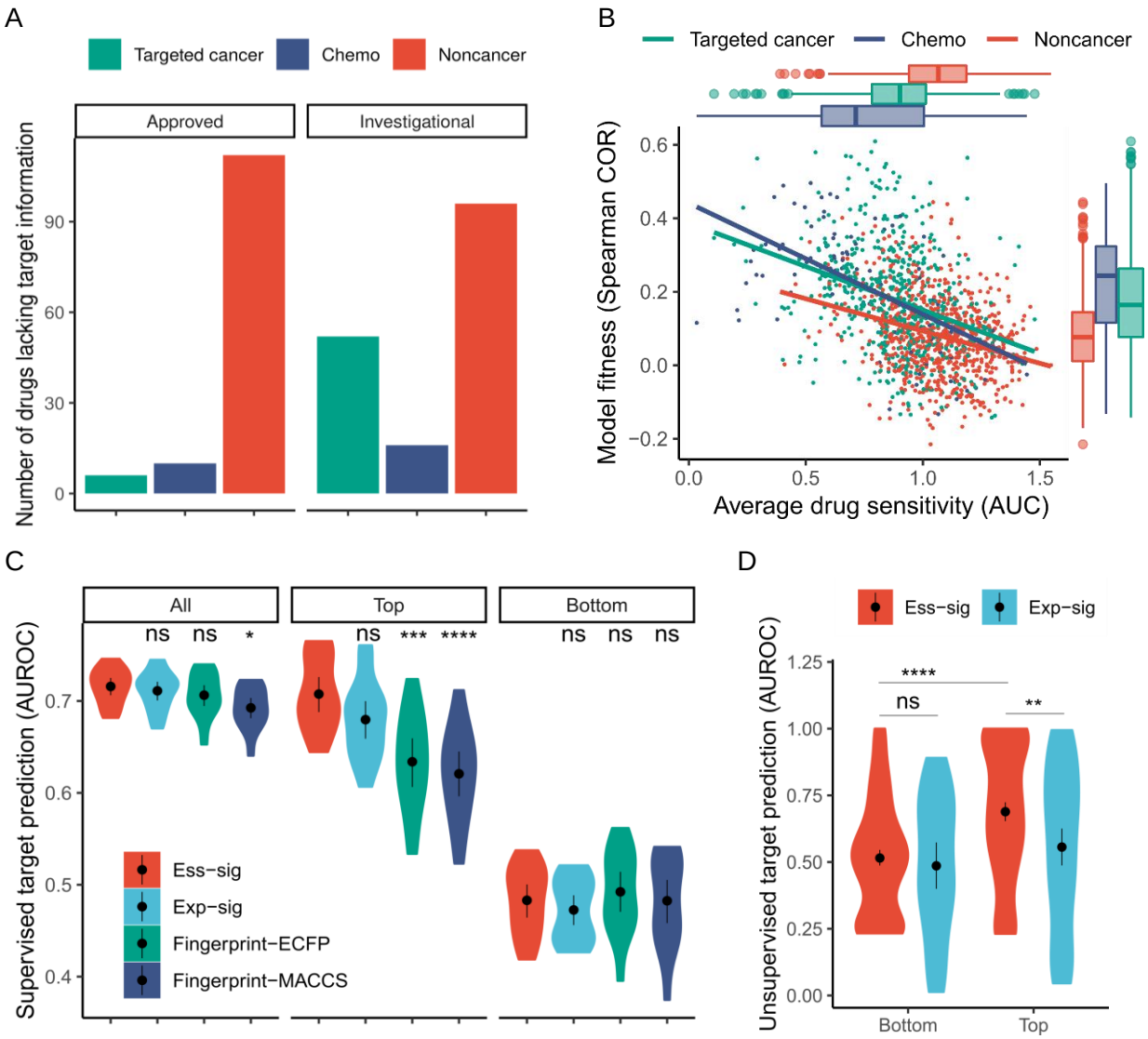


Figure 5. Application of gene essentiality signatures to predict targets of PRISM drugs
A. Numbers of drugs with missing target information in different categories

B. Model fitness (spearman correlation of the drug sensitivity prediction versus actual values) is negatively associated with the average drug sensitivity (AUC). Noncancer drugs tend to have lower model fitness and higher AUC (lower drug sensitivity).
C-D. Supervised (C) and Unsupervised (D) prediction accuracy of different representations of PRISM drugs in predicting DRH annotated drug target.

Gene essentiality signatures help mechanism exploration for noncancer drugs

To minimize false predictions for the PRISM drugs, we considered the 312 drugs with top model fitness with a Spearman COR > 0.20 (**Supplementary Figure 6**). Furthermore, only the drugs with target information were retained, resulting in 222 cancer and 46 noncancer drugs. We found that the supervised target prediction accuracy is much higher for cancer drugs than noncancer drugs (median AUC 0.97 and 0.26, t-test P-value = 3.8×10^{-7}, **Figure 6A**). Likewise, the gene essentiality signatures of cancer drugs performed better in the *de novo* target prediction (median AUC 0.73 vs 0.64, two-sided t-test P-value = 0.04). We also observed that a noncancer drug's putative targets were less likely to be shared with other drugs. **(Figure 6B)**. As shown in **Figure 6C**, supervised target prediction was significantly improved for drugs having a target(s) shared by drugs in the training set (Dunnett test, all P-values < 0.001). There might be two explanations for such results. First, given that the putative targets of noncancer drugs indeed account for the drug's killing effect, the low frequencies of the target genes in the training set prevented efficient model training. In this case, the *de novo* target prediction, which does not rely on putative target information of the training drugs, could be adopted. For example, the known target for a noncancer drug digoxin is ATP1A1, which is not targeted by any other drug in the training set and hence was not well predicted. In comparison, ATP1A1 was accurately predicted in the *de novo* target prediction as the top1 essentiality gene. Secondly, using putative targets as the ground truth may lead to an underestimation of the prediction accuracy, especially for noncancer drugs as cancer is not the targeted disease in the original development of noncancer drugs. The putative targets alone, therefore, are insufficient in explaining the growth inhibition effect of noncancer drugs.

Given that the putative targets for noncancer drugs often cannot explain the mechanisms of action in cancer treatment, we further explored the similarity of gene essentiality signatures between cancer and noncancer drugs. We projected the gene essentiality signatures using the Uniform Manifold Approximation and Projection (UMAP). As shown in **Figure 6D**, cancer drugs with common mechanisms mostly clustered together, consistent with the clustering results determined by their drug sensitivity profiles[10]. Moreover, the gene essentiality signatures recovered putative mechanisms for multiple noncancer drugs. For example, sirolimus, an mTOR inhibitor initially approved for preventing renal translation rejection, clusters well with other cancer drugs inhibiting mTOR. Likewise, three anti-parasite drugs known to inhibit tubulin polymerization are also clustered into cancer drugs classified as tubulin polymerization inhibitors (parbendazole, albendazole, and podophyllotoxin). Detecting these clusters of well-studied cancer and noncancer drugs supports the validity of using gene essentiality signatures to recover putative mechanisms for non-cancer drugs in the treatment of cancer.

Furthermore, we observed that noncancer drugs tended to distribute diversely into clusters of different cancer drug categories. For example, a FAAH inhibitor LY2183240 fell into the

tubulin polymerization inhibitor cluster and adrenergic receptor antagonists (CGS-15943 and MRS-1220) are close to the cluster of EGFR inhibitors. Interestingly, dopamine receptor antagonists were more diverse, with domperidone clustering with mTOR inhibitors whereas pardoprunox clusters with tubulin polymerization inhibitors. While the majority of noncancer drugs do not cluster clearly into specific cancer drug categories, four HMGCR inhibitors (pitavastation, atorvastatin, mevastatin and fluvastatin) formed an isolated cluster that is clearly separable from the other drugs, suggesting their unique mode-of-action in killing cancer cells. Together these results showed the capability of gene essentiality signatures to explore the mechanisms of noncancer drugs to rationalize drug repurposing.

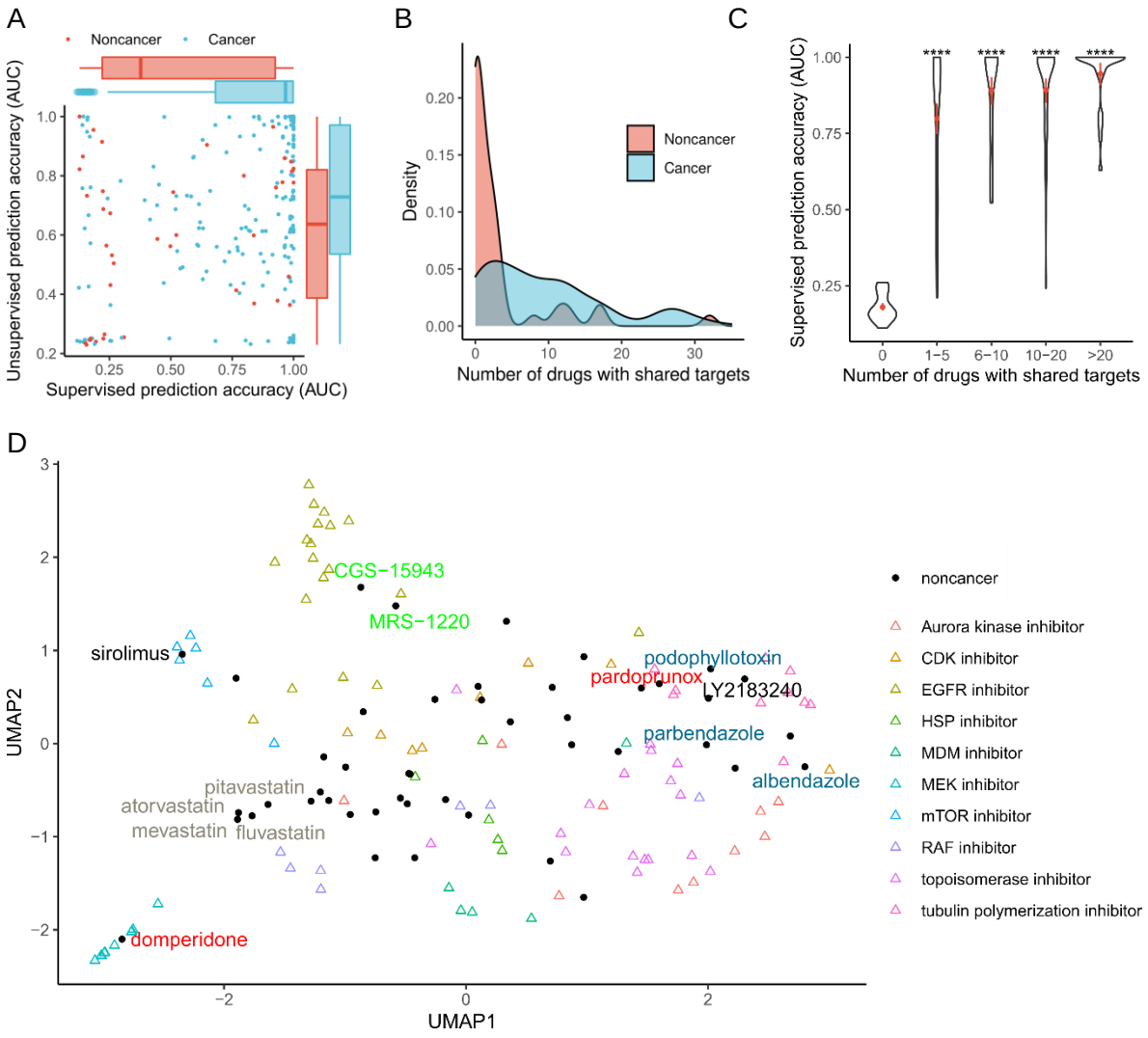


Figure 6. Target identification for noncancer drugs.

A. Supervised and *de novo* target prediction accuracy on the putative targets of the PRISM drugs.

B. Kernel density estimation of drugs having shared targets with other drugs

C. Supervised target prediction accuracy for drugs with different numbers of shared targets.

D. UMAP projection of PRISM drugs by their gene essentiality signatures.

Discussion

Characterizing the full spectrum of drug targets is fundamental for drug development[33]. In this study, we developed a gene essentiality signaturing method to characterize drug targets. Using the GDSC and CTRP data sets, we showed that our gene essentiality signatures outperformed gene expression signatures and conventional chemical fingerprints, in both supervised and *de novo* drug mechanism discovery. The gene essentiality signature enables drug-specific pathway analysis, helping uncover the mechanisms of action beyond drug target binding. We further determined the gene essentiality signatures for drugs that were tested in a drug sensitivity screening[10] based on the PRISM technology[32]. We showed the feasibility of our method to predict the drug targets. More importantly, we predicted the cancer targets and their mechanisms of action for noncancer drugs.

Meanwhile, through our analysis of gene essentiality signatures of CTRP and GDSC drugs, we identified drugs that block DNA repair, which may help improve the efficacy of immune checkpoint inhibitors[31]. Immune checkpoint inhibitors have become standard-of-care in the first-line treatment of multiple cancers[34,35]. Despite their popularity, the responsiveness to immune checkpoint inhibitors varies greatly across individuals[36]. For example, we and others have identified that higher tumor mutation burden (TMB) serves as an important prognosis biomarker for immune checkpoint inhibitor treatment[37] (https://www.synapse.org/#!Synapse:syn24241625/wiki/608681). Using the gene essentiality signatures, we identified multiple compounds that may inhibit the DNA mismatch repair process of tumor cells (**Table 1**). Indeed, we have successfully predicted all the approved drugs that are under phase 2 clinical trials in combination with immune checkpoint inhibitors. Therefore, our gene essentiality signatures may support drug combination discovery to circumvent the limitation of immunotherapy.

We used both supervised and *de novo* drug target prediction to evaluate the performance of gene essentiality signatures. When predicting the primary targets for cancer drugs, the supervised prediction performs better than the *de novo* prediction (**Figure 6A**). The performance difference, however, should be marginal when predicting new targets, as supervised prediction methods rely heavily on the coverage of training data. Recent advances in machine learning using prior information such as protein similarity to improve the prediction accuracy on the less studied target space[38]. Applying such a method with our gene essentiality signatures may improve the drug target prediction accuracy. In addition, we leveraged existing knowledge about the drug targets as the ground truth. However, such information is likely to be incomplete, especially when evaluating noncancer drugs, as the primary targets previously identified for noncancer drugs may be irrelevant to their cancer-killing efficacy. Indeed, for noncancer drugs, we observed lower concordance between predicted and putative labels. It should be noted that for cancer drugs, focusing on primary targets still induce a risk of missing additional targets that account for the treatment effecacy[1,39]. In the future, we expect that the proteome-wide drug target binding experiments could generate more unbiased data to improve computational prediction methods for drug target prediction[8,40].

Our method may be extended to study the mechanisms of action of drug combinations, as a large amount of drug combination screening data is also available for the modelling[41-43]. Secondly, we assumed that a drug inhibits its targets. However, drug target binding may not always lead to an inhibition but rather a stabilizing or activating effect[44]. Such an assumption

may contribute to false negative drug target predictions. In addition, we relied on the drug sensitivity data and gene essentiality data on cancer cell lines, which may not capture the mechanisms of action of drugs in the tumor microenvironment. Drug screening based on emerging in-vitro models, such as organoids and 3D cell culture[45,46], may further rationalize drug discovery and drug repositioning[47,48].

Methods

Gene essentiality data

We used the DepMap data portal[49] to collect the gene essentiality profiles. Specifically, we collected the CERES gene essentiality scores derived from CRISPR screens[19] and the DEMETER2 gene essentiality scores derived from shRNA screens[18]. Furthermore, we determined Combined Essentiality Score (CES[50]) by integrating CERES and DEMETER2 with the molecular profiles of the cell lines, including the TPM gene expression values derived from RNA-seq, the gene expression values from the microarray screens, the gene-level copy numbers, and the somatic mutations. We considered the three variants of gene essentiality features (CERES, DEMETER2, and CES) in the subsequent analyses. Genes with a missing rate of more than 20% were removed, after which the R MissForest package was used to impute the rest of the genes with the default parameters.

Drug sensitivity data

We retrieved the CTRPv2 dataset from the NCI's CTD-squared data portal, containing 545 small molecules and 907 cancer cell lines[51]. Cell viability percentages were determined at multiple doses of the drugs, after which a logistic curve was fitted. The area under the curve (AUC) was considered the drug sensitivity score, where an AUC of 0 represents complete cell growth inhibition, and an AUC of 15 represents no effect.

We also extracted the GDSCv2 data consisting of 198 small molecules across a panel of 809 cancer cell lines[52]. Furthermore, we obtained the PRISM data containing the drug sensitivity scores for 1,448 compounds against 499 cell lines[10]. In GDSC and PRISM, the AUC was calculated from the fitted dose viability curve and normalized further to the [0, 1] interval.

Chemical fingerprints and gene expression signatures

The Python RDkit was used to generate Extended Connectivity FingerPrint (ECFP, 1024 features) and Molecular ACCess System fingerprints (MACCS, 256 features) for each compound based on its SMILES. In addition, we retrieved the LINCS L1000 Connectivity Map data to get the gene expression signatures[11]. The consensus drug signatures with 978 features were used following the procedures described in the previous work[53].

Drug target ground truth data

Binary drug targets were extracted from the corresponding drug screening studies, where either original drug library annotation or Drug Repurposing Hub[54] (DRH) was used as the source. Furthermore, DrugTargetProfiler[55] was employed to extract and integrate quantitative drug target data from DrugTargetCommons[56], where measurements across multiple binding affinity assays were harmonized into a consensus score between 0 (non-binding) and 1 (strong binding).

Determination of gene essentiality signatures.

We built a drug sensitivity prediction model using three variants of gene essentiality signatures (CES, CERES, and DEMETER2). For each $drug_i$we fit an L1-penalized multivariate linear regression model M_i by regressing its drug sensitivity AUC against genome-wide gene essentiality for multiple cell lines:

$$M_i: auc_i \sim \beta_i ESS \quad (1)$$

Where auc_i is a $1 \times n$ vector, and *ESS* is an $m \times n$ matrix representing the essentiality profiles of m genes for the n matching cells. The $1 \times m$ coefficient vector β_i is then extracted to represent the gene essentiality signature of the drug.

Nested cross-validation (CV) was used for evaluating the model fitness, where the hyperparameters were tuned with a ten-bootstrap inner layer and the prediction accuracy was evaluated on the outer layer of five folds with three replicates. Model fitness was estimated using the holdout sets from the CV, including accuracy metric, i.e., coefficient of determination (R^2), Spearman correlation coefficient (Spearman COR), area under the precision-recall curve (PRAUC), area under the receiver-operator curve (AUROC), as well as error metrics, i.e., mean absolute error (MAE) and root-mean-square error (RMSE). Regression metrics were calculated using the original numeric drug sensitivity scores and classification metrics were calculated by binarizing using the 90% percentile as the drug sensitivity threshold.
The overall difference among the three variants of gene essentiality signatures was evaluated by ANOVA test, and if significant, followed by post hoc Tukey's test or Dunnett's test. For each drug, the performance of the three variants of gene essentiality signatures was compared using paired t-test with Bonferroni correction.

Evaluation of gene essentiality signatures

We evaluated which drug features (i.e., gene essentiality signatures, gene expression signatures, and chemical fingerprints) predicts drug targets. We considered a drug-target prediction method that was developed in recent competitions, significantly outperforming the other methods in an independent experimental validation[14]. This method is conceptually similar to the *k* nearest neighbors learning, where the targets of a queried drug are predicted by their neighbors. For a queried $drug_i$, we determined its correlations with the training drugs in the feature space as shown below:

$$r_i = cor(x_i, X) \quad (2)$$

where x_i represents the feature vector for the queried drug, X is a $m \times n$ feature matrix for the n training drugs. The drug targets for drug i were then predicted as an average of the drug target score of the training drugs, weighted by the correlation coefficients transformed by rectified linear activation function (ReLU):

$$y_i = ReLU(r_i) \times Y \quad (3)$$

Where the $p \times n$ matrix, Y represents the target profiles for all the n training drugs in p genes, $ReLU(r_i)$ is the transformed coefficient vector and y_i is the predicted target profiles for $drug_i$. Cross validation with five folds and three replicates was employed to estimate the AUROC against the ground truth.

In addition, we evaluated the similarity of a drug pair based on multiple feature spaces. For each drug pair, we used Spearman correlation to determine the similarity in the drug sensitivity space, as well as in the predicted drug target space. We used the Tanimoto distance to determine the similarity in the ground truth target space, as well as in the

chemical fingerprint space. We considered pairs of drugs with similar drug sensitivity profiles and drug-target profiles to be mutually repurposable. For example, for drug pair A and B in the CTRP dataset, we considered them mutually repurposable if both their drug sensitivity and drug target similarity are higher than 95% of other drugs paired with drug A or drug B. The threshold was set to 90% in GDSC due to insufficient ground truth drug pairs. For each mutually repurposable drug pair, we checked whether their similarity in the predicted target space is also at the top. The performance of different drug signatures was compared using the paired Wilcox signed rank test with Holm's correction for multiple testing. In addition, only the top mutually-repurposable drug pairs in the CTRP dataset (N = 16, with drug sensitivity similarity > 99%, target similarity 100%, and identical mechanisms of action) were shown.

PPI network and pathway analysis

To study the functional relevance of the gene essentiality signatures, we mapped the genes in the PPI network constructed in the recent publication[23]. Each signature gene was annotated as the drug target, first-degree, second-degree, or other neighbors, depending on the shortest path between the gene and the putative target of the given drug. GSEA[57] (Gene Set Enrichment Analysis) was conducted for the gene essentiality signatures against the KEGG and GO pathway gene sets collected from the MSigDB database[58].

Supplementary Info

Supplementary Figure 1 Model fitness of different gene features using cross-validation.

Supplementary Figure 2 Rank of similarity for repurposable drug pairs.
ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001 by the paired Wilcox signed rank test with the gene essentiality signature as the reference. Ess-sig: gene essentiality signature, Exp-sig: gene expression signature.

Supplementary Figure 3 Accuracy of unsupervised prediction of primary targets.

Supplementary Figure 4 Percentage of drugs the target neighbors of which are prioritized by their top signature genes (top/bottom 50 genes).

Supplementary Figure 5 Distribution of model fitness for PRISM drugs.

Supplementary Table 1 Mean estimates of drug sensitivity model fitness.

Supplementary Table 2 Comparison of drug sensitivity model fitness between CES and CERES.

Supplementary Table 3 Comparison of drug sensitivity model fitness between DEMETER2 and CERES.

Supplementary Table 4 Comparing variance of RMSE for drug sensitivity model fitness.

Supplementary Table 5 Statistical significance for comparing the accuracies in supervised target prediction.

Supplementary Table 6 Statistical significance for comparing the similarity percentile of repurposable drug pairs.

Supplementary Table 7 List of KEGG pathways associated with the drugs' essentiality signatures.

Supplementary Table 8 List of Gene Ontologies (GO) terms associated with the drugs' essentiality signatures.

Supplementary Table 9. Immuno-oncology (IO) combination clinical trials involving drugs associated with DNA mismatch repair.
Supplementary Table 10 PRISM drug target annotation.
Supplementary Table 11 Gene essentiality model fitness for CTRP, GDSC, and PRISM drugs.
Supplementary Table 12 Gene essentiality signatures for the drugs screened in all the three studies.
Supplementary Table 13 Performance of supervised target prediction for drugs with varying gene essentiality model fitness.

Data availability
Supplementary tables are available online**.** The datasets are publicly available in the following repositories, CTRP [https://portals.broadinstitute.org/ctrp.v2.1/], GDSC [https://www.cancerrxgene.org/] and DepMap [https://depmap.org/portal/]. DTC [https://drugtargetcommons.fimm.fi], Drug Repurposing Hub [https://clue.io/repurposing#download-data]. Specifically, DepMap 21Q1 release was employed to derive the input datasets including**,** CERES (Achilles_gene_effect.csv), TPM RNA-seq gene expression matrix (CCLE_expression.csv), the gene level CNV file (CCLE_gene_cn.csv), and the mutations (CCLE_mutations.csv). In addition, DEMETER2 (D2_combined_gene_dep_scores.csv) was collected from DEMETER2 Data v6. The gene expression from the microarray (CCLE_Expression_Entrez_2012-09-29.gct) was collected from the CCLE release available in DepMap portal. For the CTRP dataset, the curve-fitting parameters from the post-quality control dataset (v20). For GDSC we employed the curve statistics (GDSC2_fitted_dose_response_25Feb20.xlsx) from its version2. PRISM data (secondary-screen-dose-response-curve-parameters.csv) was collected from its 19Q4 release.

Code availability.
The GitHub repository is available upon request.

Acknowledgment.
We appreciate the open science data sharing of the Broad Institute of MIT and Harvard for providing The Cancer Dependency Map (DepMap), PRISM, and the Drug Repurposing Hub. We thank CSC, Finland for providing the IT services. The illustrative cartoons in Figure 1, Figure 2A, and Figure 4D are generated using bioRender.

Author contribution.
The work was supervised by J.T. W.W. and J.T. conceived of the study. W.W. developed the models and led the computational analysis. S.Z., Z.T., Y.W., and J.A. contributed to the data collection, while J.B., S.H, and J.E. contributed to the interpretation of the findings. W.W and J.T. wrote the manuscript. All authors have read and agreed to the final version of the manuscript.

Corresponding authors.
Correspondence to Wenyu Wang (wenyu.wang@helsinki.fi) or Jing Tang (jing.tang@helsinki.fi).

Funding

This work was supported by the EU H2020 (EOSC-LIFE, No. 824087, JT), the European Research Council (DrugComb, No. 716063, JT), and the Academy of Finland (No. 317680, JT). W.W. was funded by the FIMM-EMBL international PhD program, Doctoral Program of Biomedicine at University of Helsinki, K. Albin Johanssons stiftelse and Ida Montinin Säätiö. S.H was supported by the National Natural Science Foundation of China (No. 82203120) as well as the Key Research and Development Program of Shaanxi (Program No.2022SF-092).

Competing interests.
None to declare

References

1 Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aaw8412 (2019).
2 Wong, C. H., Siah, K. W. & Lo, A. W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273-286, doi:10.1093/biostatistics/kxx069 (2019).
3 Kubota, K., Funabashi, M. & Ogura, Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. *Biochim Biophys Acta Proteins Proteom* **1867**, 22-27, doi:10.1016/j.bbapap.2018.08.002 (2019).
4 Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat Rev Drug Discov* **16**, 531-543, doi:10.1038/nrd.2017.111 (2017).
5 Terstappen, G. C., Schlüpen, C., Raggiaschi, R. & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov* **6**, 891-903, doi:10.1038/nrd2410 (2007).
6 Swinney, D. C. Phenotypic vs. target-based drug discovery for first-in-class medicines. *Clin Pharmacol Ther* **93**, 299-301, doi:10.1038/clpt.2012.236 (2013).
7 Jafari, R. *et al.* The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat Protoc* **9**, 2100-2122, doi:10.1038/nprot.2014.138 (2014).
8 Gaetani, M. *et al.* Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution. *J Proteome Res* **18**, 4027-4037, doi:10.1021/acs.jproteome.9b00500 (2019).
9 Sun, J. *et al.* Recent advances in proteome-wide label-free target deconvolution for bioactive small molecules. *Med Res Rev* **41**, 2893-2926, doi:10.1002/med.21788 (2021).
10 Corsello, S. M. *et al.* Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* **1**, 235-248, doi:10.1038/s43018-019-0018-6 (2020).
11 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e1417, doi:10.1016/j.cell.2017.10.049 (2017).
12 Bush, E. C. *et al.* PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat Commun* **8**, 105, doi:10.1038/s41467-017-00136-z (2017).
13 Chen, W. & Zhou, X. Drug Signature Detection Based on L1000 Genomic and Proteomic Big Data. *Methods Mol Biol* **1939**, 273-286, doi:10.1007/978-1-4939-9089-4_15 (2019).

14 Douglass, E. F., Jr. *et al.* A community challenge for a pancancer drug mechanism of action inference from perturbational profile data. *Cell Rep Med* **3**, 100492, doi:10.1016/j.xcrm.2021.100492 (2022).
15 Gao, S. *et al.* Modeling drug mechanism of action with large scale gene-expression profiles using GPAR, an artificial intelligence platform. *BMC Bioinformatics* **22**, 17, doi:10.1186/s12859-020-03915-6 (2021).
16 Pham, T. H., Qiu, Y., Zeng, J., Xie, L. & Zhang, P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nat Mach Intell* **3**, 247-257, doi:10.1038/s42256-020-00285-9 (2021).
17 Rancati, G., Moffat, J., Typas, A. & Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat Rev Genet* **19**, 34-49, doi:10.1038/nrg.2017.74 (2018).
18 McFarland, J. M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat Commun* **9**, 4610, doi:10.1038/s41467-018-06916-5 (2018).
19 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784, doi:10.1038/ng.3984 (2017).
20 Schaefer, C. *et al.* Target discovery screens using pooled shRNA libraries and next-generation sequencing: A model workflow and analytical algorithm. *PLoS One* **13**, e0191570, doi:10.1371/journal.pone.0191570 (2018).
21 Yan, X. *et al.* High-content imaging-based pooled CRISPR screens in mammalian cells. *J Cell Biol* **220**, doi:10.1083/jcb.202008158 (2021).
22 Tang, Y. C. & Gottlieb, A. Explainable drug sensitivity prediction through cancer pathway enrichment. *Sci Rep* **11**, 3128, doi:10.1038/s41598-021-82612-7 (2021).
23 Gonçalves, E. *et al.* Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. *Mol Syst Biol* **16**, e9405, doi:10.15252/msb.20199405 (2020).
24 Kurilov, R., Haibe-Kains, B. & Brors, B. Assessment of modelling strategies for drug response prediction in cell lines and xenografts. *Sci Rep* **10**, 2849, doi:10.1038/s41598-020-59656-2 (2020).
25 Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol* **32**, 1202-1212, doi:10.1038/nbt.2877 (2014).
26 García-Martínez, J. M. *et al.* Ku-0063794 is a specific inhibitor of the mammalian target of rapamycin (mTOR). *Biochem J* **421**, 29-42, doi:10.1042/bj20090489 (2009).
27 Engelman, J. A. *et al.* ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. *Proc Natl Acad Sci U S A* **102**, 3788-3793, doi:10.1073/pnas.0409773102 (2005).
28 Furcht, C. M., Buonato, J. M. & Lazzara, M. J. EGFR-activated Src family kinases maintain GAB1-SHP2 complexes distal from EGFR. *Sci Signal* **8**, ra46, doi:10.1126/scisignal.2005697 (2015).
29 Shi, T. *et al.* Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway. *Sci Signal* **9**, rs6, doi:10.1126/scisignal.aaf0891 (2016).
30 Gunnarsson, U. *et al.* Association between local immune cell infiltration, mismatch repair status and systemic inflammatory response in colorectal cancer. *J Transl Med* **18**, 178, doi:10.1186/s12967-020-02336-6 (2020).
31 Pećina-Šlaus, N., Kafka, A., Salamon, I. & Bukovac, A. Mismatch Repair Pathway, Genome Stability and Cancer. *Front Mol Biosci* **7**, 122, doi:10.3389/fmolb.2020.00122 (2020).
32 Yu, C. *et al.* High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat Biotechnol* **34**, 419-423, doi:10.1038/nbt.3460 (2016).
33 Kabir, A. & Muth, A. Polypharmacology: The science of multi-targeting molecules. *Pharmacol Res* **176**, 106055, doi:10.1016/j.phrs.2021.106055 (2022).

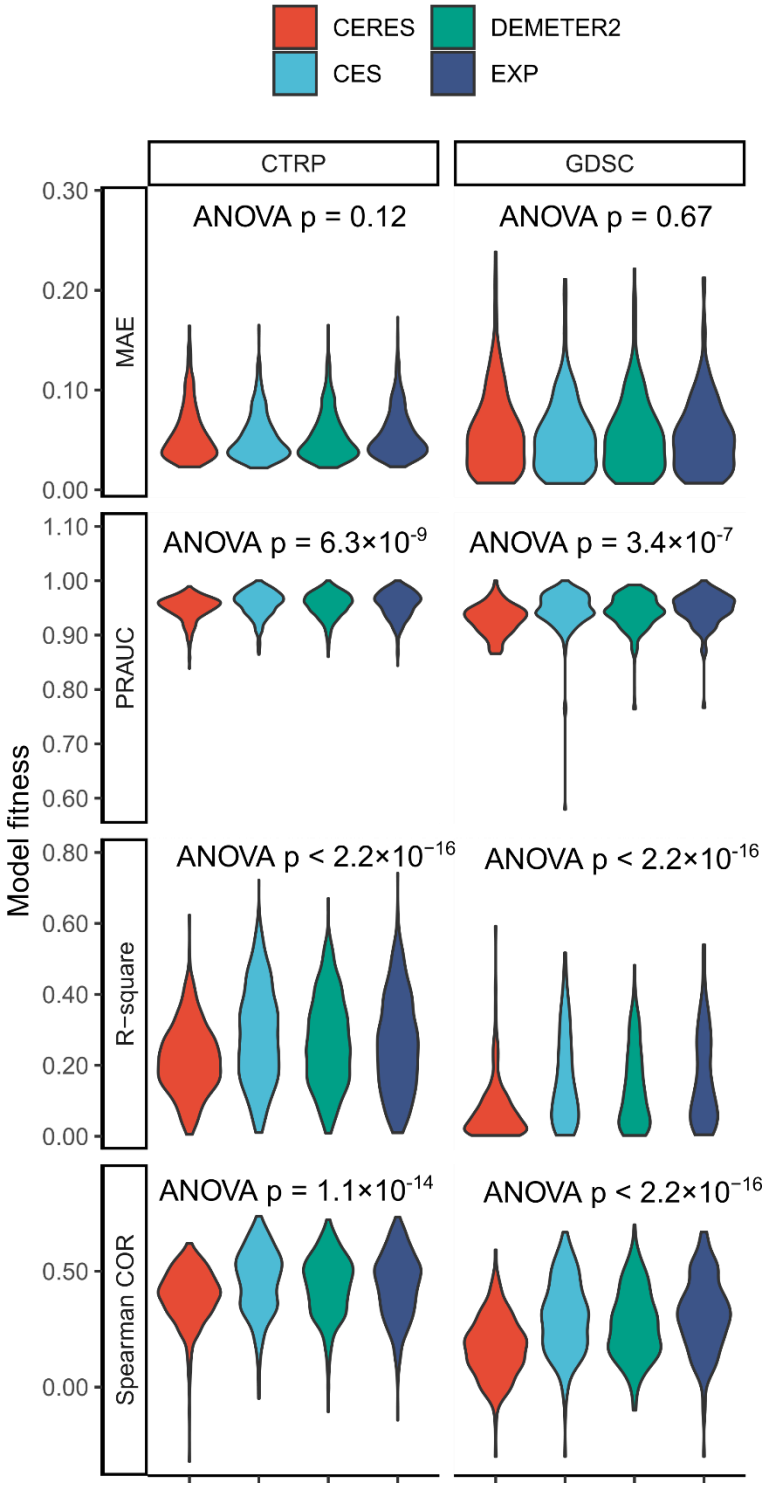
34 Marin-Acevedo, J. A., Kimbrough, E. O. & Lou, Y. Next generation of immune checkpoint inhibitors and beyond. *J Hematol Oncol* **14**, 45, doi:10.1186/s13045-021-01056-8 (2021).
35 Pistamaltzian, N. F., Georgoulias, V. & Kotsakis, A. The role of immune checkpoint inhibitors in advanced non-small cell lung cancer. *Expert Rev Respir Med* **13**, 435-447, doi:10.1080/17476348.2019.1593828 (2019).
36 Shields, B. D. *et al.* Indicators of responsiveness to immune checkpoint inhibitors. *Sci Rep* **7**, 807, doi:10.1038/s41598-017-01000-2 (2017).
37 Zheng, M. Tumor mutation burden for predicting immune checkpoint blockade response: the more, the better. *J Immunother Cancer* **10**, doi:10.1136/jitc-2021-003087 (2022).
38 Wan, F. *et al.* DeepCPI: A Deep Learning-based Framework for Large-scale in silico Drug Screening. *Genomics Proteomics Bioinformatics* **17**, 478-495, doi:10.1016/j.gpb.2019.04.003 (2019).
39 Antolin, A. A. *et al.* The kinase polypharmacology landscape of clinical PARP inhibitors. *Sci Rep* **10**, 2585, doi:10.1038/s41598-020-59074-4 (2020).
40 Gholizadeh, E. *et al.* Identification of Celecoxib-Targeted Proteins Using Label-Free Thermal Proteome Profiling on Rat Hippocampus. *Mol Pharmacol* **99**, 308-318, doi:10.1124/molpharm.120.000210 (2021).
41 Jia, J. *et al.* Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* **8**, 111-128, doi:10.1038/nrd2683 (2009).
42 Zagidullin, B. *et al.* DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res* **47**, W43-w51, doi:10.1093/nar/gkz337 (2019).
43 Zheng, S. *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* **49**, W174-w184, doi:10.1093/nar/gkab438 (2021).
44 Andrei, S. A. *et al.* Stabilization of protein-protein interactions in drug discovery. *Expert Opin Drug Discov* **12**, 925-940, doi:10.1080/17460441.2017.1346608 (2017).
45 Badr-Eldin, S. M., Aldawsari, H. M., Kotta, S., Deb, P. K. & Venugopala, K. N. Three-Dimensional In Vitro Cell Culture Models for Efficient Drug Discovery: Progress So Far and Future Prospects. *Pharmaceuticals (Basel)* **15**, doi:10.3390/ph15080926 (2022).
46 Calandrini, C. & Drost, J. Normal and tumor-derived organoids as a drug screening platform for tumor-specific drug vulnerabilities. *STAR Protoc* **3**, 101079, doi:10.1016/j.xpro.2021.101079 (2022).
47 Engel, M., Belfiore, L., Aghaei, B. & Sutija, M. Enabling high throughput drug discovery in 3D cell cultures through a novel bioprinting workflow. *SLAS Technol* **27**, 32-38, doi:10.1016/j.slast.2021.10.002 (2022).
48 Langhans, S. A. Three-Dimensional in Vitro Cell Culture Models in Drug Discovery and Drug Repositioning. *Front Pharmacol* **9**, 6, doi:10.3389/fphar.2018.00006 (2018).
49 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e516, doi:10.1016/j.cell.2017.06.010 (2017).
50 Wang, W. *et al.* Combined gene essentiality scoring improves the prediction of cancer dependency maps. *EBioMedicine* **50**, 67-80, doi:10.1016/j.ebiom.2019.10.051 (2019).
51 Rees, M. G. *et al.* Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* **12**, 109-116, doi:10.1038/nchembio.1986 (2016).
52 Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740-754, doi:10.1016/j.cell.2016.06.017 (2016).
53 Szalai, B. *et al.* Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. *Nucleic Acids Res* **47**, 10010-10026, doi:10.1093/nar/gkz805 (2019).
54 Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* **23**, 405-408, doi:10.1038/nm.4306 (2017).

55 Tanoli, Z. *et al.* Interactive visual analysis of drug-target interaction networks using Drug Target Profiler, with applications to precision medicine and drug repurposing. *Brief Bioinform*, doi:10.1093/bib/bby119 (2018).

56 Tang, J. *et al.* Drug Target Commons: A Community Effort to Build a Consensus Knowledge Base for Drug-Target Interactions. *Cell Chem Biol* **25**, 224-229.e222, doi:10.1016/j.chembiol.2017.11.009 (2018).

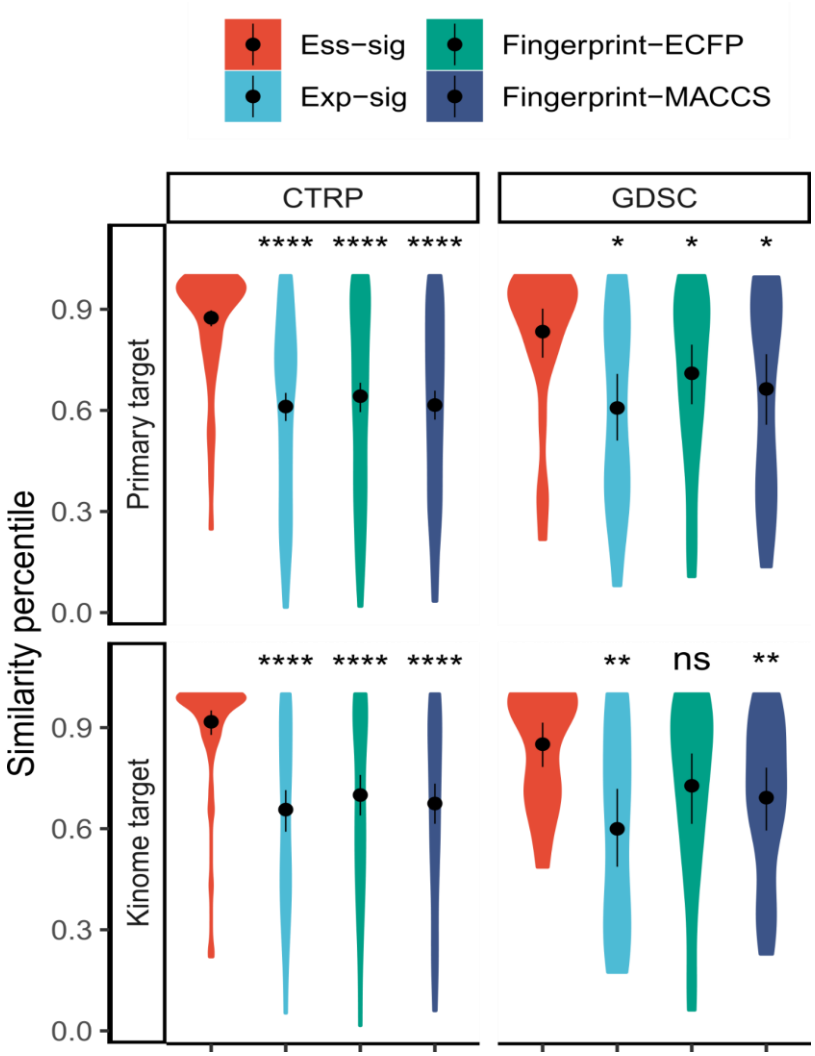
57 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

58 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

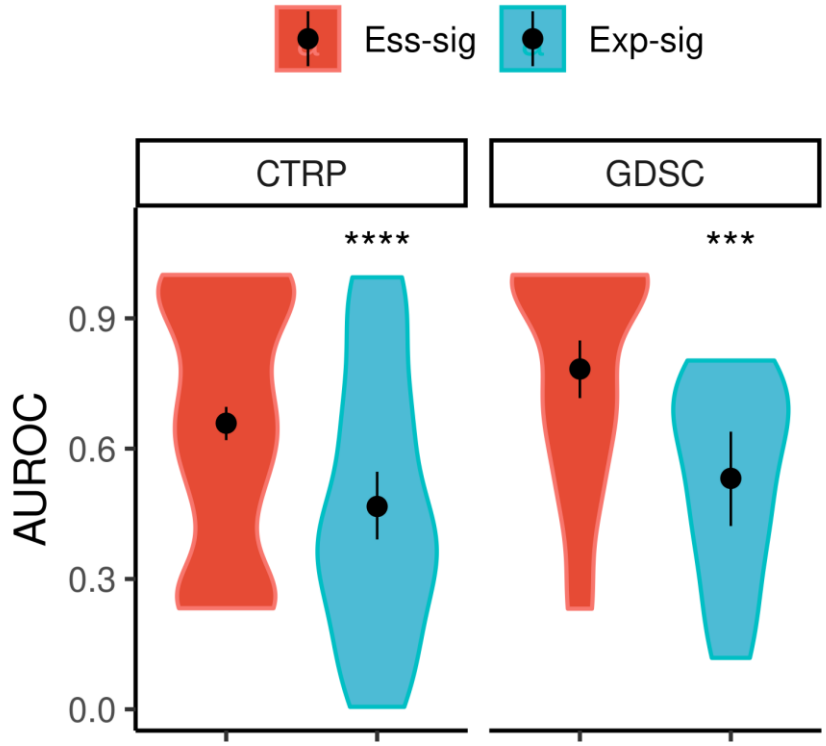


Supplementary Figure 1

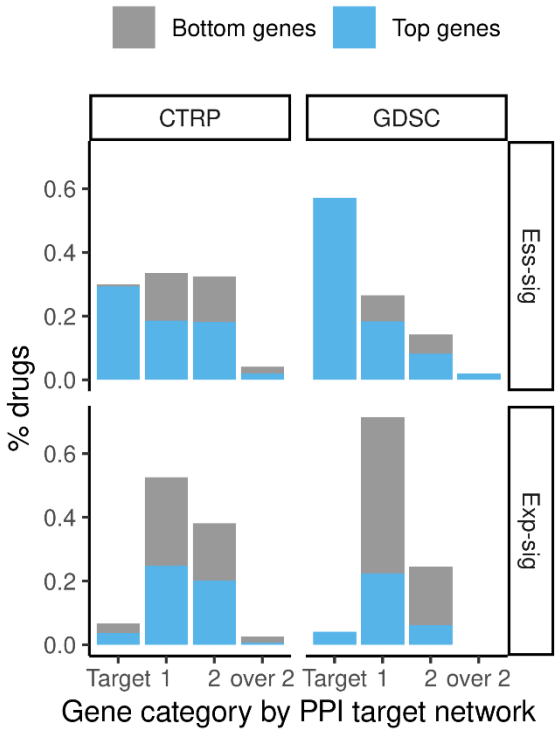
Model fitness of different gene features using cross-validation.



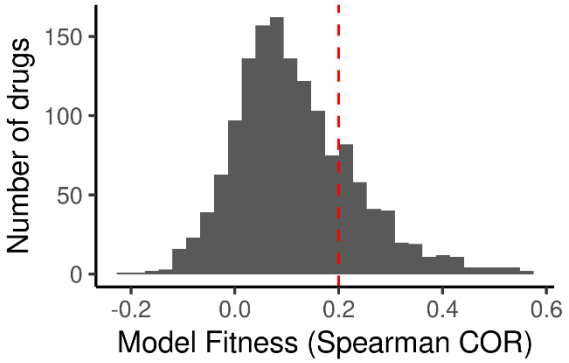
Supplementary Figure 2 Rank of similarity for repurposable drug pairs.
ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001 by the paired Wilcox signed rank test with the gene essentiality signature as the reference. Ess-sig: gene essentiality signature, Exp-sig: gene expression signature.



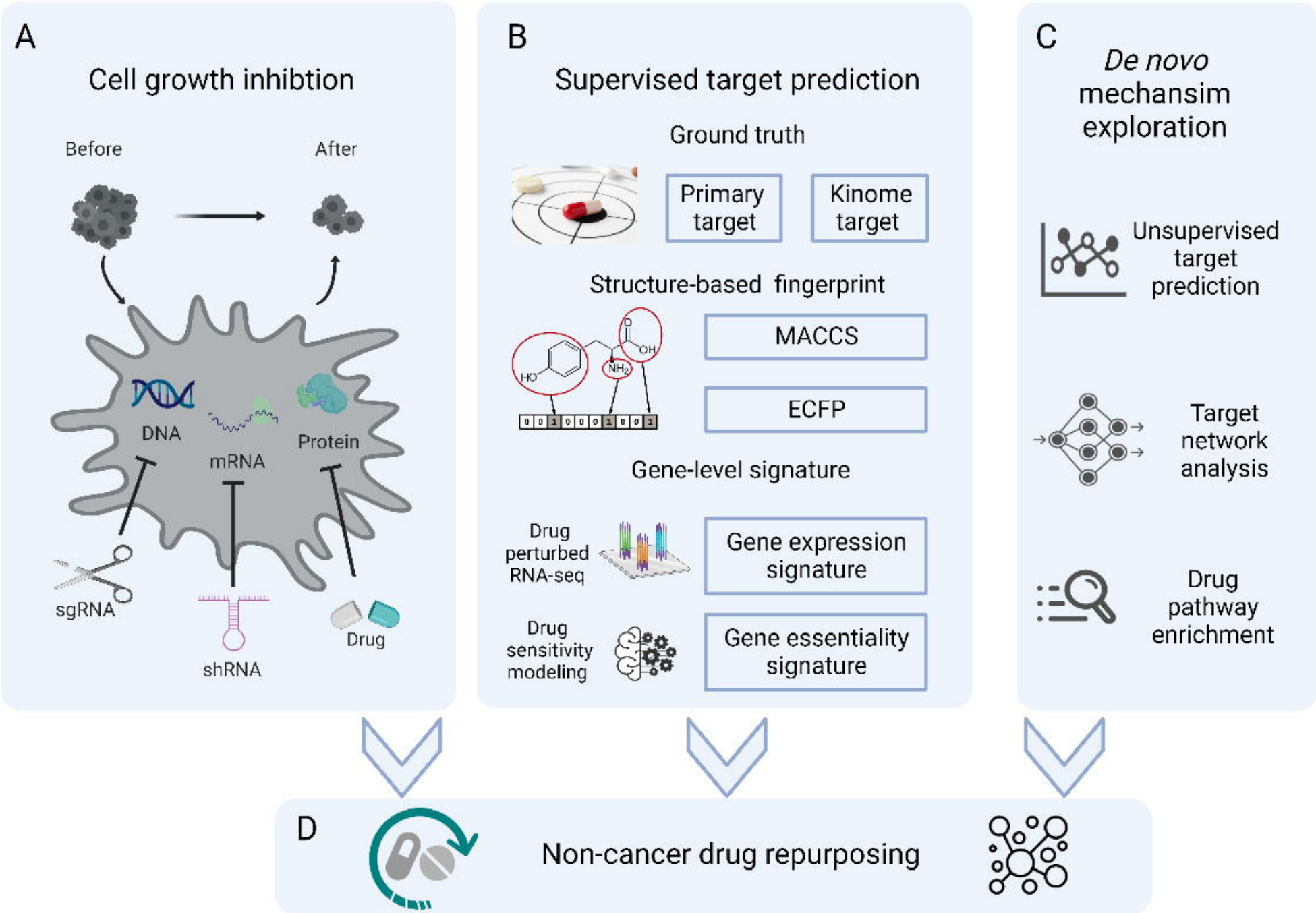
Supplementary Figure 3 Accuracy of unsupervised prediction of primary targets.



Supplementary Figure 4 Percentage of drugs the target neighbors of which are prioritized by their top signature genes (top/bottom 50 genes).



Supplementary Figure 5 Distribution of model fitness for PRISM drugs.


A
Cell growth inhibtion
Before
After
DNA
mRNA
Protein
sgRNA
shRNA
Drug
B
Supervised target prediction
Ground truth
Primary target
Kinome target
Structure-based fingerprint
MACCS
ECFP
Gene-level signature
Drug perturbed RNA-seq
Gene expression signature
Drug sensitivity modeling
Gene essentiality signature
C
De novo mechansim exploration
Unsupervised target prediction
Target network analysis
Drug pathway enrichment
D
Non-cancer drug repurposing