*Briefings in Bioinformatics*, 2025, **26(3)**, bbaf266
https://doi.org/10.1093/bib/bbaf266
**Case Study**

# Systematic identification of cancer-type-specific drugs based on essential genes and validations in lung adenocarcinoma

Xiang Lian[1,2,‡], Xia Kuang[1,‡], Dong-Dong Zhang[1], Qian Xu[1], Anqiang Ye[1], Cheng-Yu Wang[1], Hong-Tu Cui[1], Hai-Xia Guo[3], Ji-Yun Zhang[1], Yuan Liu[4], Ge-Fei Hao[5,*], Zhenshun Cheng[4,6,7,*], Feng-Biao Guo[1,2,*]

[1]Department of Respiratory and Critical Care Medicine, Zhongnan Hospital of Wuhan University, School of Pharmaceutical Sciences, Wuhan University, 185 Donghu Road, Wuchang District, Wuhan 430071, China
[2]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, Wuhan University, 185 Donghu Road, Wuchang District, Wuhan 430071, China
[3]School of Life Science and Technology, University of Electronic Science and Technology of China, 2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu 611731, China
[4]Department of Respiratory and Critical Care Medicine, Zhongnan Hospital of Wuhan University, Wuhan University, 169 Donghu Road, Wuchang District, Wuhan 430071, China
[5]State Key Laboratory of Green Pesticide, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Center for R&D of Fine Chemicals, Guizhou University, 2708 South Section of Huaxi Avenue, Huaxi District, Guiyang 550025, China
[6]Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences, 169 Donghu Road, Wuchang District, Wuhan 430071, China
[7]Hubei Engineering Center for Infectious Disease Prevention, Control and Treatment, 169 Donghu Road, Wuchang District, Wuhan 430071, China

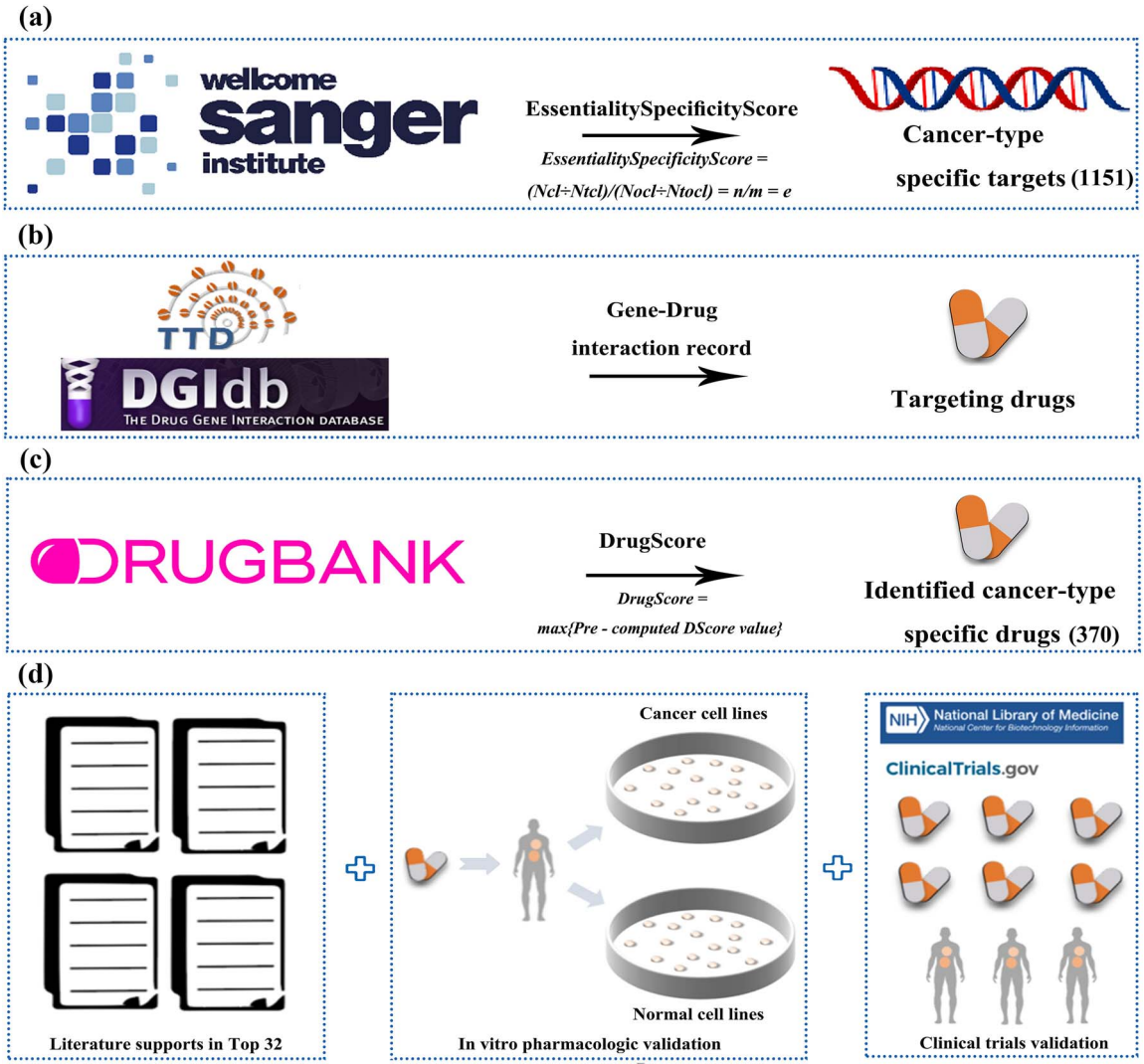
[*]Corresponding authors. Feng-Biao Guo, Department of Respiratory and Critical Care Medicine, Zhongnan Hospital of Wuhan University, School of Pharmaceutical Sciences, Wuhan University, 185 Donghu Road, Wuchang District, Wuhan 430071, China. Tel: 0086-13880956017; E-mail: fbguoy@whu.edu.cn; Zhenshun Cheng, Department of Respiratory and Critical Care Medicine, Zhongnan Hospital of Wuhan University, Wuhan University, 169 Donghu Road, Wuchang District, Wuhan 430071, China. E-mail: zhenshun_cheng@126.com; Ge-Fei Hao, State Key Laboratory of Green Pesticide, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Center for R&D of Fine Chemicals, Guizhou University, 2708 South Section of Huaxi Avenue, Huaxi District, Guiyang 550025, China. E-mail: gefei_hao@foxmail.com
[‡]Xiang Lian and Xia Kuang contributed equally to this work.

## Abstract

Depicting a global landscape of essential gene-targeting drugs would provide more opportunities for cancer therapy. However, a systematic investigation on drugs targeting essential genes still has not been reported. We suppose that drugs targeting cancer-type-specific essential genes would generally have less toxicity than those targeting pan-cancer essential genes. A scoring function-based strategy was developed to identify cancer-type-specific targets and drugs. The *EssentialitySpecificityScore* ranked the essential genes in 19 cancer types, and 1151 top genes were identified as cancer-type-specific targets. Combining target-drug interaction databases with research/marketing status, 370 cancer-type-specific drugs were identified, bound to 100 out of all identified targets. Profiles of applied cancer types of identified targets and drugs illustrate the scoring strategy's effectiveness: most drugs apply to cancer types $<10$. Seven drugs with no previous anticancer evidence were validated in 11 lung adenocarcinoma cell lines, and lower inhibition rates (from 9.4% to 44.0%) were observed in 10 normal cell lines. This difference is statistically significant (Student's *t*-test, $P \leq .0001$), confirming the rationality of our supposition. Our built EGKG (Essential Gene Knowledge Graph) forms a computational basis to uncover essential gene targets and drugs for specific cancer types. It is available at http://gepa.org.cn/egkg/. Also, our experimental result suggests that combining drugs with orthogonal essentiality may be an alternative way to improve anticancer effects while maintaining biocompatibility. The code and data are available at https://github.com/KKINGA1/EGKG_data_process.

**Received:** November 28, 2024. **Revised:** May 3, 2025. **Accepted:** May 15, 2025

**Graphical Abstract**




**Keywords:** cancer-type-specific drugs; essential genes; *EssentialitySpecificityScore*; lung adenocarcinoma; cell inhibition rate

## Introduction

The 90% failure rate in the development of cancer drugs is due to a lack of efficacy, and effective anticancer drug entities need to be developed [1, 2]. One of the most effective means of cancer treatment is targeted therapy [3, 4], which uses a gene with a definite effect as a drug target [5]. Functional genomics approaches can guide the mining of novel targets [6] from tumor genomic data to develop personalized anticancer drugs [7, 8]. However, the abundance of genomic information [7], limited reliable target identification [9], lack of clinical efficacy [1], high toxicity, and side effects of anticancer drugs [10] are still obstacles to developing personalized cancer drugs [11]. Identifying essential genes provides the basis for targeted cancer therapy [12–16]. Essential genes support vital cell functions [17] and can serve as excellent target genes [13, 18, 19]. Prioritizing essential genes as targets can provide more reliable choices, increase the feasibility of personalized cancer therapy [20], and facilitate the development of novel cancer drugs. Due to the high mortality rate and the lack of effective drugs [21], the reuse of existing drugs has become a hot topic of research in recent years [22], which is expected to overcome the above limitations that hinder the development of cancer therapy.

At present, many studies have focused on drug target screening [23] and vaccine design [13, 24] around essential genes. Bacterial essential genes have been well studied and widely used to identify antimicrobial target genes [25]. A variety of essential gene recognition models have also emerged for human cell lines. In 2019, the Sanger Center identified essential genes in 325 cell lines using CRISPR high-throughput knockout technology, enabling the exploitation of these genes for drug screenings [26]. In 2021, Liu's group from Tongji University analyzed the synthetic lethality of cancer cell lines and established a map of cancer-type therapeutic drugs based on the synthetic lethal gene pairs [27]. However, until now, there is still a lack of systematic studies for cancer therapeutics targeting essential genes [28]. The drug targets need to be highly specific to kill tumor cells without affecting normal cells

[29]. Cancer-type specific targets are usually identified by their abnormal expression level caused by coding gene mutation or epigenetic modification of the promoter region [30–34]. Behan and Colleagues connected cancer-type-specific targets with essential genes [26]. Cancer-type-specific essential genes, the functions of which are only necessary for certain cancer types or most cell lines of those cancer types and are non-essential in most cell lines of the other cancer types, constitute ideal drug targets [35, 36]. Now, it is urgent to depict a global landscape of drugs targeting cancer-type-specific essential genes. Precise antineoplastic drugs kill tumors with little or no toxicity to normal cells [28]. It is reasonable that drugs targeting essential genes specific to a certain cancer type would meet the above requirements in principle.

Accordingly, we proposed a prioritization strategy of antitumor drugs based on essential genes. This systematic strategy contains three steps and involves three scores expressed in two scoring functions. A unique feature of our cancer drug-repurposing method lies in that cancer-type-specific essential genes are identified according to a defined essentiality specificity score, and cancer-type-specific drugs are identified by their database-recorded binding with identified target genes. However, most cancer drug-repurposing methods are based on interaction networks of multiple elements (drugs, syndromes, targets, diseases). After identifying cancer-type-specific drugs, an online service (http://gepa.org.cn/egkg/) was graphically established to display the drug prioritization map. Then, we indeed observed the lower toxicity of the top-prioritizing drugs to normal cells than cancer cells for lung cancer *in vitro*. Therefore, our work obtained the complete list of drugs targeting cancer-type-specific essential genes and validated the relatively higher safety of such types of drugs at the cell level. In two works [37, 38] developing accurate computational frameworks to predict and mitigate risks of drug combinations, the authors emphasized that drug–drug interactions (DDIs) can substantially influence therapeutic outcomes by potentially altering drug efficacy and inducing severe adverse effects. They inspired us to combine drugs against the same cancer types based on the gene essentiality relationship of their targets. As a result, we experimentally observed that combining drugs with orthogonal essentiality has a stronger anticancer effect than using them alone, whereas biocompatibility is superior. Our computational and experimental results collectively offer valuable insights and implications for cancer-targeted therapy.

## Materials and methods

### Drug reorientation scoring model

In this study, we developed an effective drug repurposing scoring scheme by integrating two core scoring functions, *EssentialitySpecificityScore* (Fig. 1a) and Drug Status Score (*DrugScore*) (Fig. 1c).

*EssentialitySpecificityScore* is calculated as follows:

$$\text{EssentialitySpecificityScore} = \frac{\text{Ncl} \div \text{Ntcl}}{\text{Nocl} \div \text{Ntocl}} = \frac{\text{n}}{\text{m}} = \text{e} \quad (1)$$

where Ncl: The number of cell lines for which a gene is essential in a cancer type; Ntcl: Total number of cell lines of this cancer type; Nocl: The number of cell lines for which a gene is essential in other cancer types (in the case of lung adenocarcinoma, this number denotes the cell lines for which a gene is essential in cancer types other than lung adenocarcinoma); Ntocl: Total number of cell lines for other cancer types.

Ideally, the higher the *e* score, the more significant the difference between the essentiality of this gene in a given cancer type and that in normal cells. To guarantee the statistical significance of the *n* value, we only retain these cancer types with five or more cell lines. That is to say, we identified cancer-type-specific targets for 19 cancer types.

Secondly, referring to the strategy applied in SLKG, to integrate drug research and marketing (R&M) status and target gene type, drug scoring is defined as [27]:

$$DrugScore = maxPre - computedDScorevalue \quad (2)$$

$$Pre - computedDScorevalue = [DrugStatus + TargetGeneType] \quad (3)$$

The detailed scores corresponding to specific status and drug target type are listed in Table 1. *DrugScore* is defined as the maximum of the set of values of *Pre-computed DScore* values to avoid multiple sources leading to multiple drug R&M statuses. A drug may act on multiple target genes, and this may cause one drug to have multiple *Target-Gene-Type* [22]. In addition, we also unified the drug names in the TTD [35] database and DGI [39] database according to Drugbank's records [40]. Matching the index ID of all three databases (pubchem_cid/pubchem_sid/chembl_id) at the same time makes it easy to query for drug-specific information and delete drugs that do not have these three IDs (more specific information cannot be obtained).

To determine the threshold scores for the best-repurposed drug candidates, we further examined the probability and cumulative distribution curves of the *EssentialitySpecificityScore* and *DrugScore*. In the probability distribution curves (data not shown), the *EssentialitySpecificityScore* is enriched in the regions with low scores. Therefore, the thresholds for the two scoring functions are as follows:

$$DrugScore \geq 0.8 \quad (4)$$

$$e > 1.3 \; and \; n > 0.5 \quad (5)$$

In total, we calculated the three scores for 10376 drugs derived from public databases and identified 370 drugs that are potentially effective in at least one cancer type. We have used three scores affiliated with two functions to identify potential cancer-type-specific drugs. After finishing the identification process, we will rank the identified drugs in ascending order of the *e* score.

The ranked drugs and targets are graphically exhibited on the EGKG site. Method details for building EGKG, entity and relationship establishment of EGKG, EGKG web server development, *in vitro* experimental section, molecular docking, and simulation sections were described in the Supplementary Material.

## Results

### Using a three-step strategy to prioritize cancer-type-specific genes and identify potential drugs

We proposed the first systematic investigation of repurposed drugs targeting genes, specifically essential in 19 different cancer types. In our study, we have found that distinguishing at the subtype level, such as lung adenocarcinoma within non-small cell lung carcinoma, provides a more precise understanding of gene essentialities. This approach allows us to identify specific genetic vulnerabilities that may be unique to a particular subtype. Consequently, the effectiveness of drugs targeting these vulnerabilities is more likely to be generalizable within that subtype. The general framework of this study is outlined in Fig. 1. First, we

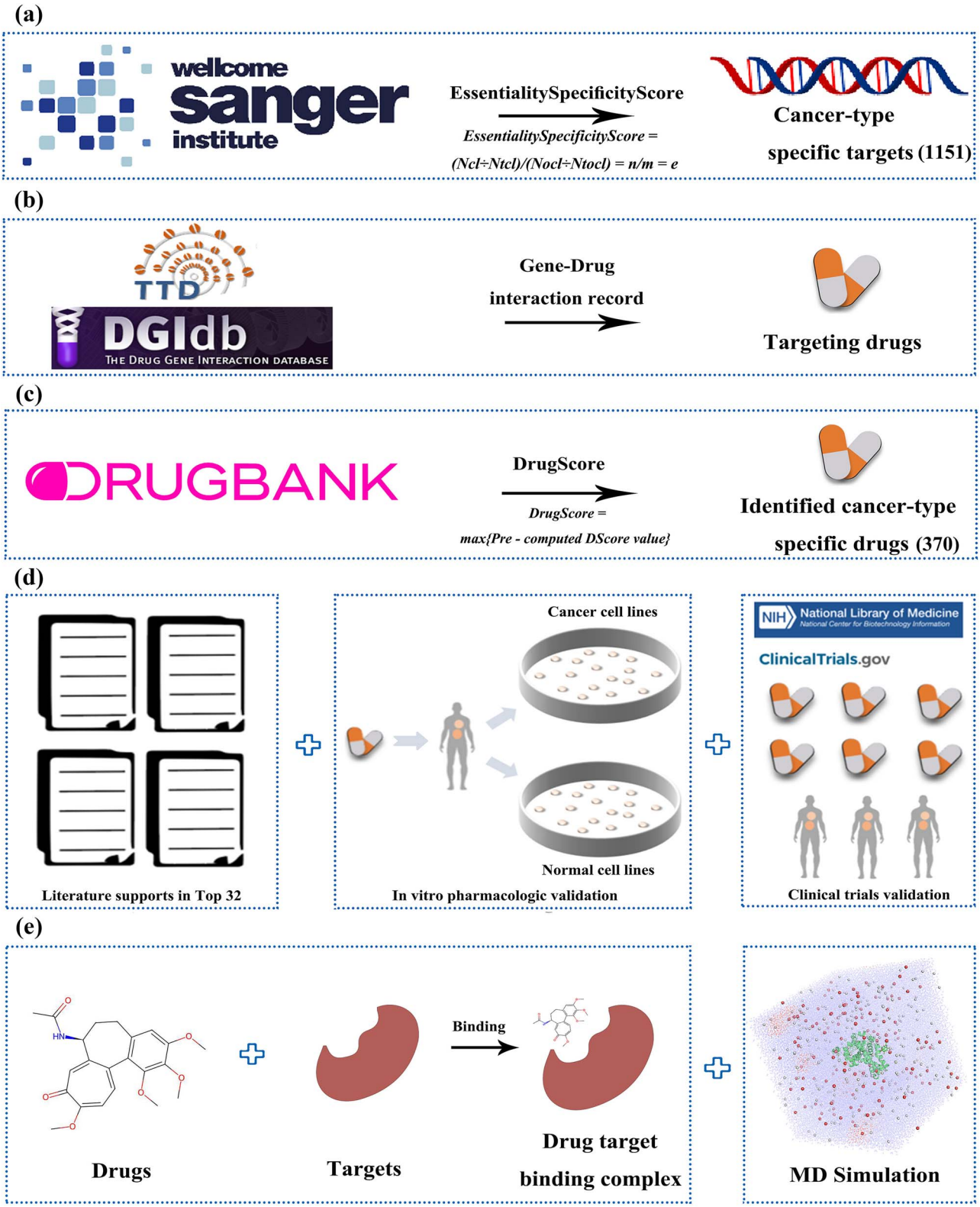



Figure 1. The general framework of the EGKG (Essential Gene Knowledge Graph) and the subsequent analysis. This study consists of five steps. (a–c) Retrieving essential genes from the Sanger Center and selecting cancer-type specific targets; matching top targets with interacted drugs according to database records; selecting the drugs targeting top genes and with acceptable market/research status. The above steps could identify drugs specific to all 19 cancer types. (d) Validating repurposed drugs with literature support and *in vitro* pharmacological evidence for lung adenocarcinoma. (e) Validating the binding of colchicine and rosiglitazone with SCD, progesterone with CDK4 by molecular docking and dynamic simulation. Both genes encoding the two target proteins have a higher *EssentialitySpecificityScore*.

Table 1. Scoring criteria for *DrugScore*

| Development status of the drug/compound | Type of target gene [27] | *DrugScore* value |
|---|---|---|
| Approved | Direct target | 1 |
| | Others | 0.8 |
| Clinical trials | Direct target | 0.6 |
| | Others | 0.4 |
| Experimental | Direct target | 0.2 |
| | Others | 0.1 |
| Others | Direct target | 0.05 |
| | Others | 0.01 |

obtained information on whether the 7470 genes were essential in each of the 325 cell lines, reported by the Sanger Center [26], and excluded 13 cancer types with cell line numbers less than five in that gene dependence project. Then, we counted the *EssentialitySpecificityScore* (*e* score) of these genes in each cancer type. With this score, we constructed relationships between genes and cancer types. We also considered the ratio of cell lines for which a gene is essential in a cancer type and defined it as *n* in the formula of the scoring function for the *e* score. With the two scores, we will retain only those genes larger than certain thresholds as cancer-type-specific targets.

The methodology for obtaining gene-drug records refers to the PanDrugs (http://www.Pandrugs.org) [7]. Gene-drug records indicate whether a gene is a potential target of a drug. The specific source of gene-drug records and the number of records in each source are shown in Supplementary Table S1. For all the cancer-type-specific gene targets, we match them with drugs based on gene-drug records.

Meanwhile, referring to SLKG [27], *DrugScore* was calculated as described in the methods to integrate the drug R&M status and target gene type, and the scoring criteria for *DrugScore* are shown in Table 1. When matching drugs to specific cancers through gene-drug records, drugs with *DrugScore* less than 0.8 were filtered out.

Hence, we identified drugs for each cancer type based on the relationship between genes and cancer types, gene-drug records, as well as the R&M status of drugs. All the identified drugs have $n > 0.5$ to guarantee the wide-spectrum in a given cancer type; $e > 1.3$ means the high specificity and less toxicity associated with inhibiting the target gene; $DrugScore \geq 0.8$ reflects the clinical safety. Among the 19 cancer types, the number of identified drugs and their corresponding targets are shown in Supplementary Table S2. All information was uploaded onto the EGKG (Essential Gene Knowledge Graph) (http://gepa.org.cn/egkg/index.php), which presents a potential option for the precision treatment of tumors based on the specificity of essential genes. For example, there are a total of 39 lung adenocarcinoma-specific targets; among them, seven have matching drugs. However, only five have matching drugs with *DrugScore* larger than or equal to 0.8, and the number of so-called lung adenocarcinoma-specific drugs is 32.

Among our chosen 1151 cancer-type-specific targets, most have no matched drugs. We also listed these 1051 targets without bound drugs in 19 cancer types. For example, gene YPEL5 has an *n* value of 0.8 and an *e* score of 1.656 for lung adenocarcinoma. If, in the future, novel drugs could be developed targeting it, they would be wide-spectrum in lung adenocarcinoma and may cause less toxicity because of the high specificity of essentiality for this gene in the cancer type. Novel drugs targeting gene SEC63 (*n*: 0.650, *e*: 2.401) are also expected to treat lung adenocarcinoma with high safety.

The workflow of screening cancer type-specific targets and identifying drug candidates by gene specificity scores, gene-drug interaction information, and drug scores is shown in Fig. 2a and b, respectively. Consequently, different numbers of their applied cancer types exist among the finally identified targets and drugs. For example, 368 identified targets apply to only one cancer type, and all identified targets (both with or without matching drugs, Fig. 2c and d) are applied to less than 10 cancer types, indicating these targets' specificity and agents specifically targeting them would cause lower toxicity. As shown in Fig. 2e, there are 69 identified drugs applied to only one cancer type, and the cumulative number of identified drugs applicable to nine or fewer cancer types is 335, which accounts for 90.5% of all identified drugs. Although some drugs may bind multiple essential genes, we consider their collective effects in our screening procedure. Hence, very rare, wide-spectrum anticancer drugs appear in our final list. These results are consistent with our expectation of cancer-type-specific drug screening.

## Identified drugs were validated with *in vitro* pharmacological evidence

We recognize the significance of selecting a representative cancer type for validation purposes. We chose lung adenocarcinoma for two compelling reasons: (i) Prevalence and relevance: lung cancer, particularly lung adenocarcinoma, is one of the most prevalent and lethal cancers worldwide, highlighting its importance in therapeutic development. (ii) Data availability: The extensive cell line data available for lung adenocarcinoma in the Sanger dataset, which has the highest number of cell lines with gene essentialities information, provides a solid foundation for our validation process. For lung adenocarcinoma, we selected the top six genes (CDK4, WNK1, HSP90B1, OGDH, VHL, SCD) with the highest *e* value to match with interacting drugs (among the first six genes, there were no targeting drugs for HSP90B1 and OGDH). After removing the existing cancer applications (Supplementary Table S3) (TTD) as well as the drugs that were not available, we finally selected seven drugs out of the top 32 drugs with the highest *e* value for cell experiments and listed their information in Table 2. In addition, CDK4 (Cyclin-dependent kinase 4) and SCD (stearoyl-CoA desaturase) are complementarily essential in 11 lung adenocarcinoma cell lines (Supplementary Table S4), which theoretically have the potential to be combined. In other words, the two genes have orthogonal essentiality profiles across lung cancer cell lines. So, their targeting drugs, progesterone and rosiglitazone, were combined to treat lung adenocarcinoma cell lines.

Top-identified drugs have often been validated through clinical trial results on the ClinicalTrials.gov website. Of the top 32 reusable lung adenocarcinoma drugs, 23 are in clinical trials, with 21 having cancer clinical trials. These results suggest that a significant proportion of the top-ranked reusable drugs are registered in clinical trials, further demonstrating the reliability of our predictive results (Supplementary Table S5).

To validate the virtual screening results of EGKG for lung adenocarcinoma, the top seven drugs were selected to be experimentally validated *in vitro*. We also verified the effect of these seven drugs on 10 normal cell lines from different parts of the human body. The average survival rate (Fig. 3a) and standard deviation of 11 lung adenocarcinoma cell lines and 10 normal

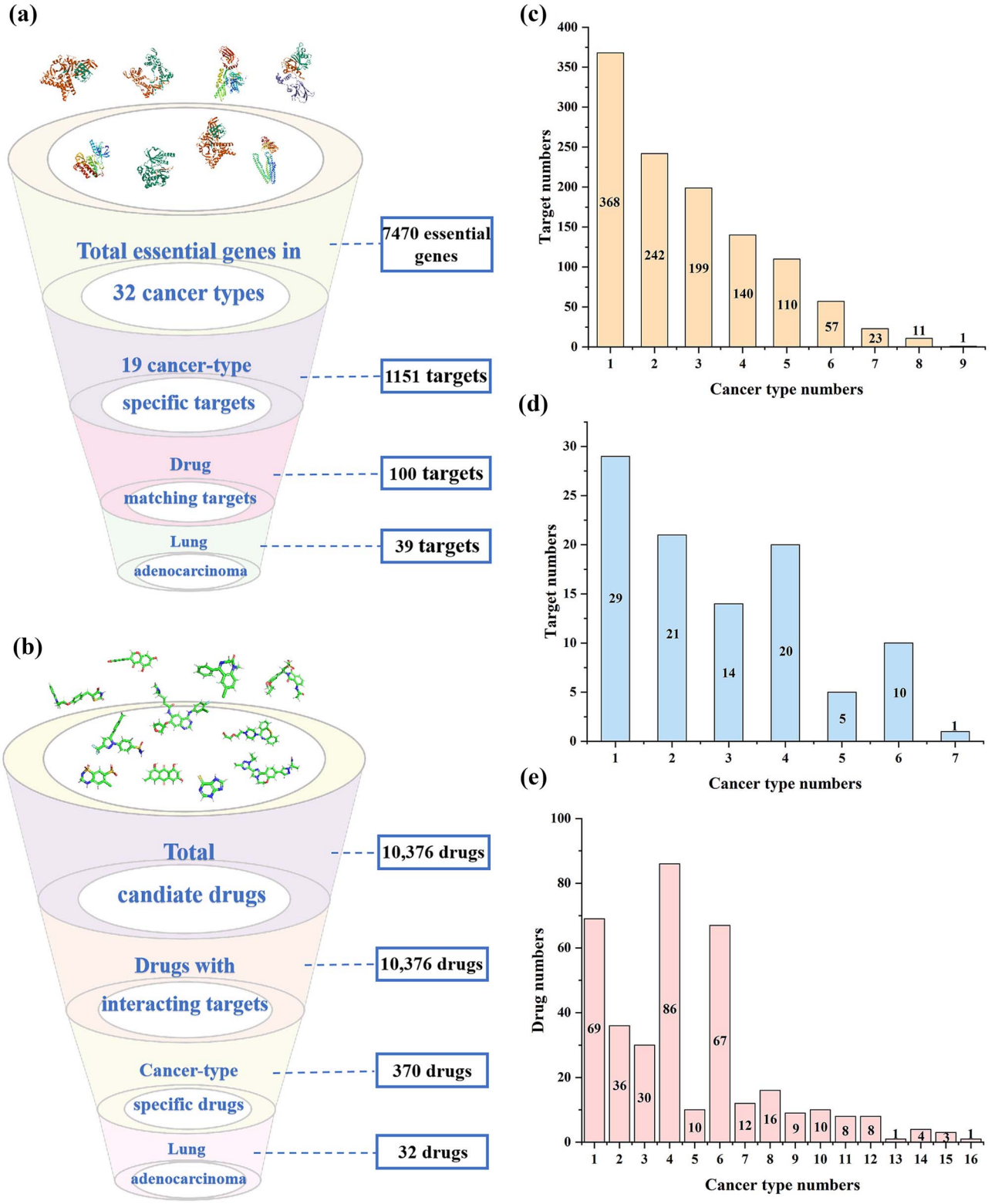



Figure 2. The number of targets and drugs at various stages and applied in a given number of cancer types. Workflows for screening cancer type-specific targets (a) and identifying drug candidates (b) by gene-specific scores, drug-target interaction information, and drug scores. (c) The number of target genes changes with the number of applied cancer types. (d) The number of target genes with matched drugs changes with the number of their applied cancer types. (e) The number of drugs changes with the number of their applied cancer types.

Table 2. The sources of gene-drug records, *e* score, *DrugScore* of seven chosen drugs to validate, and their affinities with interacting targets in lung adenocarcinoma

| Drugs | Target | Source | Source provider | *n* score | *e* score | *DrugScore* | AA (kcal/mol) |
|---|---|---|---|---|---|---|---|
| Apremilast | CDK4 | TTD | TTD | 0.55 | 2.13 | 1 | −8.2 |
| Colchicine | SCD | NCI | DGIdb | 0.8 | 1.32 | 1 | −6.2 |
| Dexamethasone | CDK4 | NCI, CIViC | DGIdb | 0.55 | 2.13 | 1 | −9.2 |
| Hydrochlorothiazide | WNK1 | PharmGKB | DGIdb | 0.5 | 1.56 | 1 | −6.1 |
| Ibuprofen | VHL | NCI | DGIdb | 0.65 | 1.33 | 1 | −5.1 |
| Progesterone | CDK4 | NCI | DGIdb | 0.55 | 2.13 | 1 | −9.9 |
| Rosiglitazone | SCD | NCI | DGIdb | 0.8 | 1.32 | 1 | −8.9 |

Abbreviations: AA, average affinity (kcal/mol); ABS, absolute value.

cell lines, and the *P*-value of the Student's t-test for the two types of cells are shown in Supplementary Table S6. The survival rate of each lung adenocarcinoma cell line and normal cell line is also shown (Fig. 3c–f, Supplementary Figs S1 and S2). It can be observed that the cell survival rate of lung cancer cell lines is significantly reduced compared with normal cell lines. More specifically, colchicine has significant anticancer efficacy with no significant toxic side effects on normal cells (Fig. 3c and d). In addition, the *P*-value of the Student's *t*-test of colchicine survival rate was 0.00066 (Supplementary Table S6), indicating that its effect on cancer cell lines and normal cell lines was significantly different. We also used IC50 values as parameters for drug efficacy, and the results of IC50 values for these drugs are shown in Supplementary Table S7. Colchicine can be seen to have lower IC50 values for 11 lung cancer cell lines compared to the other six drugs (Fig. 3b). From the above results, it can be seen that colchicine has an excellent killing effect on a variety of lung cancer cell lines and has good biocompatibility with normal cell lines.

Because their targeting genes, CDK4 and SCD, have orthogonal essentialities across cell lines of lung adenocarcinoma, it is important to test the combination effect of progesterone and rosiglitazone. In the combination drug experiment, the average inhibition rate of 11 lung adenocarcinoma cell lines could reach 68.77% at a concentration of 30 $\mu g{\cdot}ml^{-1}$. When these two drugs were used alone, the average inhibition rates of progesterone and rosiglitazone against lung cancer cell lines were 53.64% and 47.85%, respectively (Supplementary Table S6). It can be seen that the combination of drugs has a more significant anticancer effect than the use of drugs alone (Fig. 3e; Supplementary Figs S1c and S2c). Subsequently, we further verified their combined effect on the viability of a variety of normal cell lines (Fig. 3f), and the average cell survival rate was 94.31% (Supplementary Table S6). In addition, the *P*-value of the Student's t-test of the survival rate of combined treatment with progesterone and rosiglitazone was 1.26E-11 (Supplementary Table S6). The above results indicate that the drug combination has a stronger anticancer effect than the use alone, and the biocompatibility is superior.

## Molecular docking and molecular dynamics simulation of SCD with colchicine and rosiglitazone, CDK4 with progesterone

We obtained preliminary affinity calculations by molecular docking (Table 2). Based on gene-drug interactions deposited in the public database (https://dgidb.org/), we picked out SCD for colchicine and rosiglitazone, and CDK4 for progesterone. SCD and CDK4, which are widely essential in lung cancer cell lines, were nine and seven in 11 lung cancer cell lines, respectively (Supplementary Table S4). The results of molecular docking showed that the binding energy of colchicine to SCD protein was −6.2 kcal/mol, the binding energy of progesterone to CDK4 protein was −9.9 kcal/mol, and the binding energy of rosiglitazone to SCD protein was −8.9 kcal/mol, indicating that the three drugs could bind well to their corresponding targets. In addition, molecular docking models of progesterone with CDK4 (PDB ID: 7SJ3) and colchicine and rosiglitazone with SCD (PDB ID: 4ZYO), respectively, showed that colchicine, rosiglitazone, and progesterone docked to the binding site of the SCD original ligand and the binding site of the CDK4 original ligand, respectively (Fig. 4). Molecular dynamics simulation results (Supplementary Figs S3, S4, S5, and S6) support the stable binding of these drugs with their essential receptors.

## Most of the 370 identified drugs are repurposed as anticancer agents and have a restricted number of bound targets

We have identified 370 cancer-type-specific essential genes bound by drugs, which are distributed among 19 cancer types. Some of these drugs' originally approved indications may be cancer therapy, while others may not. We extracted the information on treating indications from Drugbank and showed their distribution in Fig. 5a. As we can see, only 24.3% of drugs are developed for tumor therapy, and over 70% are identified as antitumor because they may bind targets that are essential in specific cancer cells (types). These results could preliminarily demonstrate that our identifying strategy has the capacity for drug repurposing.

Ideal drugs are expected to have specific and precise targeting, while off-targeting would cause toxicity to normal cells or the human body [41]. This rule holds for our identified anticancer drugs. Since we assess the specificity of one drug by each of all potentially bound targets, if this drug could bind simultaneously with many essential genes, its specificity would be compromised. In Fig. 5b, we show the distribution of the number of recorded targets that are essential in any cancer cells for all 370 drugs. It is acceptable that about half of the 370 drugs have essential targets in one to three, indicating relatively high specificity of these drugs as antitumor agents.

We think that drug toxicity to normal cells is determined by the aggregate essentiality of all simultaneously bound targets in all cancer cells. In the inhibitory experiments, we observed a correlation (Pearson's correlation analysis, $R = -0.224$, $P = .062$, 77 inhibitory rates with 7 aggregative $m$ values of all validated drugs) between the aggregative $m$ values of potentially bound targets and the viability of the normal cells. We think that the observed correlation is not highly significant just because the tested drugs are as low as seven. Based on this result, if a drug

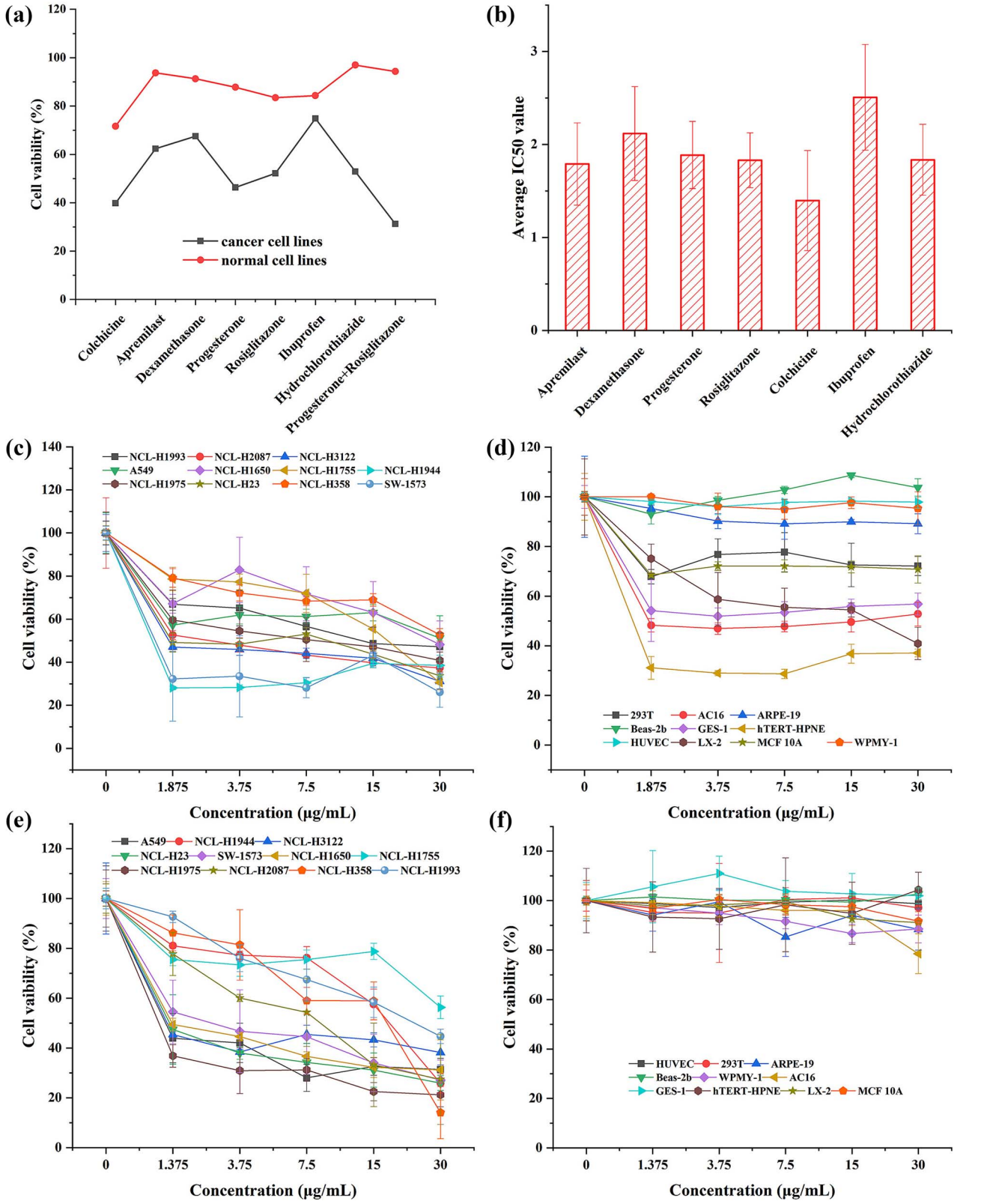



Figure 3. *In vitro* validation of seven identified drugs in lung adenocarcinoma. (a) Average cell survival rate after drug treatment (30 $\mu g \cdot ml^{-1}$) with 11 lung cancer cell lines and 10 normal cell lines. (b) Mean $\log_{10}$(IC50[$\mu$M]) values of seven drugs against 11 lung cancer cell lines. (c and d) Cell viability of colchicine in 11 lung cancer cell lines and 10 normal cell lines, respectively. (e and f) Cell viability of progesterone in combination with rosiglitazone in 11 lung cancer cell lines and 10 normal cell lines, respectively.

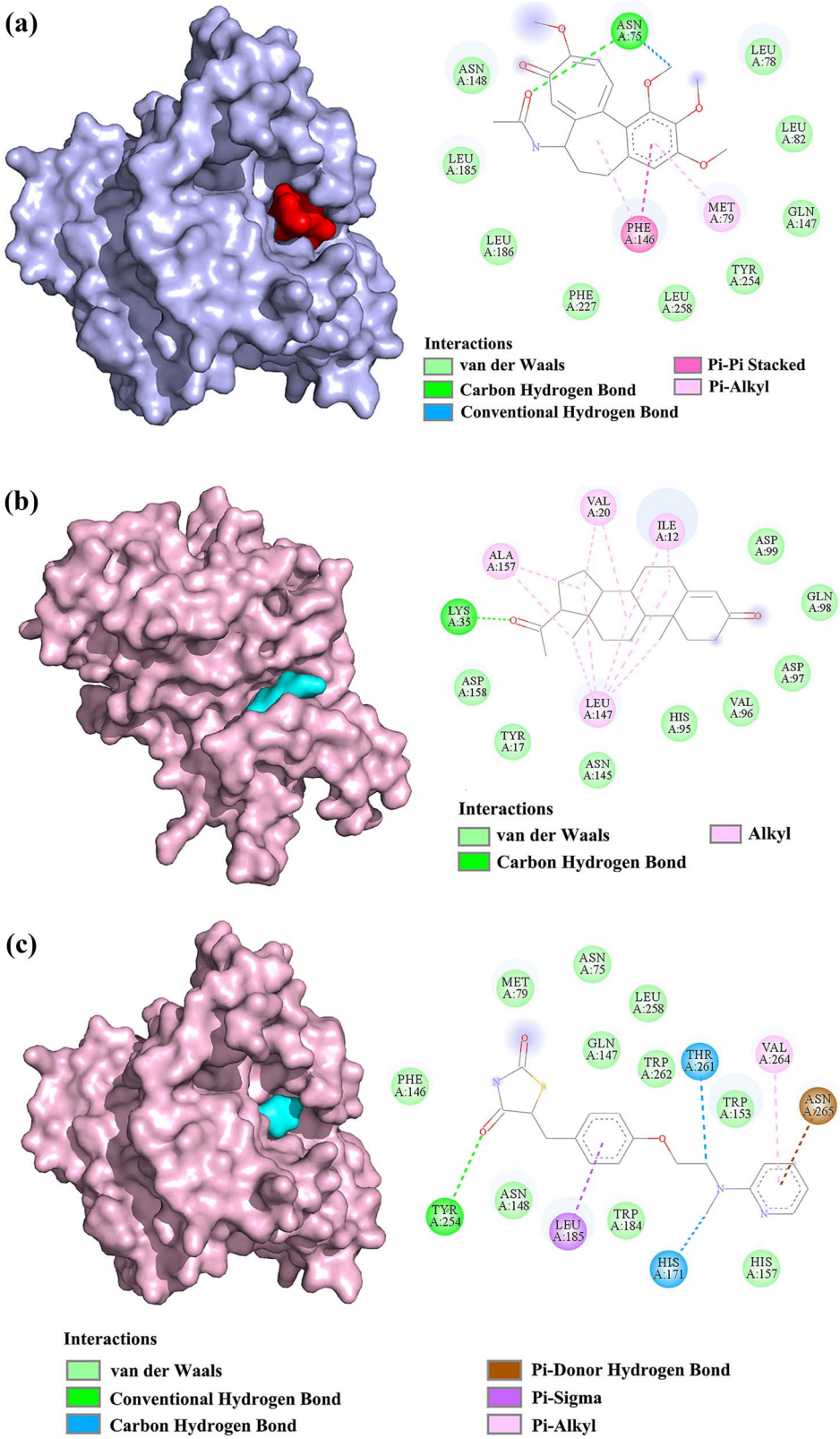



Figure 4. The binding mechanisms of the three drugs to proteins, respectively. (a) SCD and colchicine. (b) CDK4 and progesterone. (c) SCD and rosiglitazone.

has $m$ values less than 0.5, it could not hold high inhibitory efficiency on normal cells. Therefore, if we figure out the aggregate essentiality of all the potential targets for one drug, we can roughly estimate its maximum toxicity to normal cells. Hence, we sum the essential number of all potential targets recorded in the public databases for all drugs. For example, prednisone has 10 binding targets, and the $m$ value of the aggregate 10 targets is 0.41 (132/325). Consequently, we got Fig. 5c and luckily found that about half of the 370 drugs have a maximum essentiality ($m$ value) less than 0.5. A drug could not simultaneously

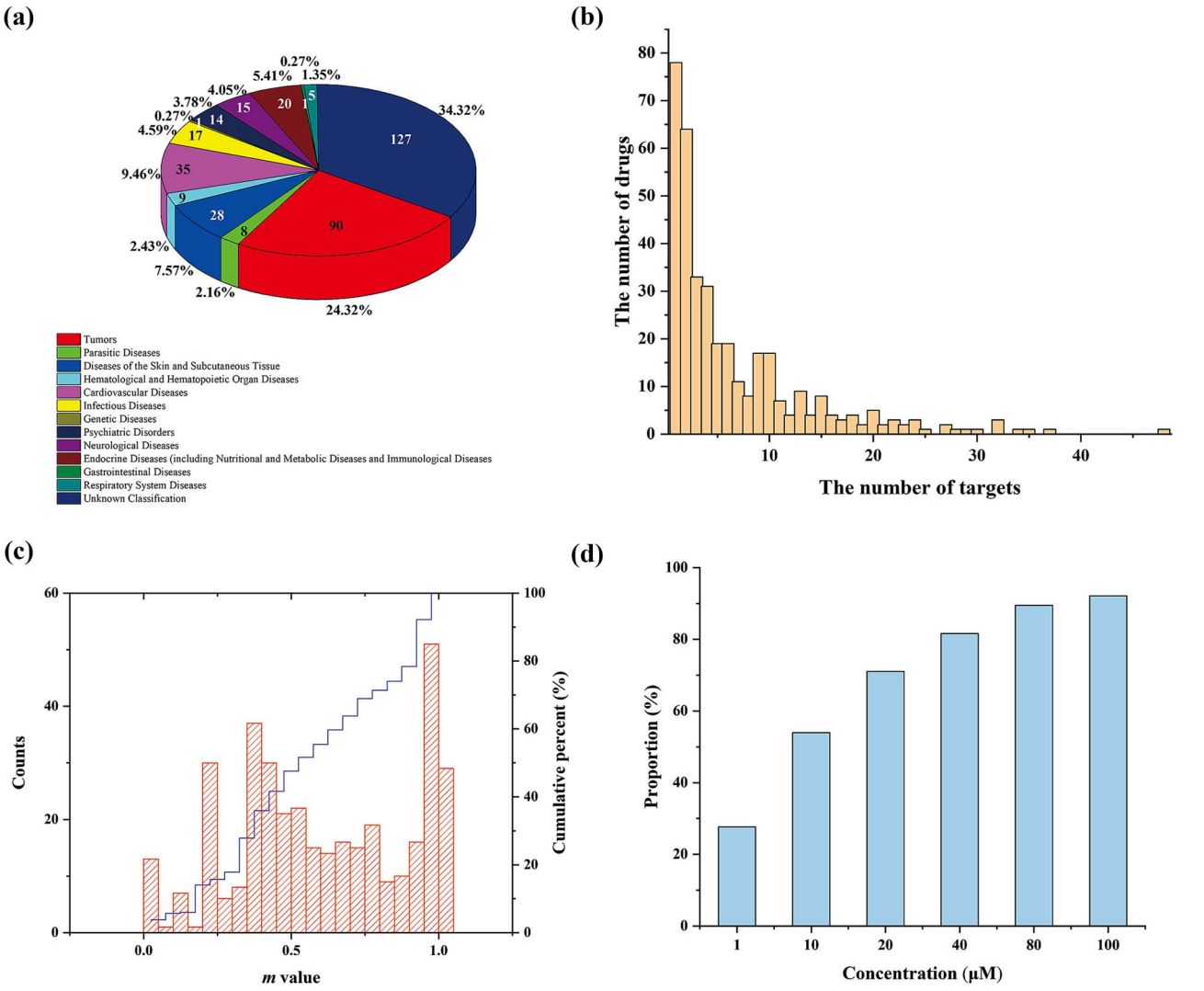



Figure 5. Distribution characteristics of identified drugs and targets. (a) Originally approved indications of the 370 identified drugs according to Drugbank. (b) The distribution of the number of targets potentially bound to one of the 370 drugs and also essential in any of the 325 cancer cells. (c) The aggregative essentiality (*m* value) of all potential targets for each of the 370 drugs (parallel targets). (d) The proportion of effective drugs across various cancer types based on IC50 concentration thresholds, as verified with 76 drugs from the DepMap and ChEMBL databases.

bind all potential targets, and hence, many more drugs will be safe because their factual aggregative essentiality will be less than 0.5.

## Validation of identified drugs' efficacy across cancer types

To further substantiate our findings, we have compiled IC50 value information for all drugs from DepMap (https://depmap.org/portal/) and ChEMBL (https://www.ebi.ac.uk/chembl/) and found that 76 of our identified 370 drugs have effective IC50 values within consistent cancer types. The graph (Fig. 5d) illustrates the proportion of effective drugs (average IC50 < 1–100 $\mu$M) targeting their respective cancer cell lines for 76 drugs. Notably, 41 drugs showed an average IC50 value <10 $\mu$M across the cancer types they target, indicating that our predicted drugs are effective against more than half of the predicted cancer types. This finding underscores the robustness of our computational predictions and supports the validity of our identified drug candidates.

## Validation of combined drug efficacy and synergy

To address the intriguing combination strategy for drugs targeting orthogonal essentialities and to strengthen our conclusions with robust experimental support, we have expanded our experimental validation process. Our approach involved identifying drug combinations that target genes with orthogonal essentiality, meaning they are crucial for cancer cell survival. We enhanced our criteria to focus on where the combined target essentiality score was increased to 1, aiming to identify cancer-type-specific drug combinations that are more likely to be effective.

We utilized the CombDrug (https://drugcomb.org/) database to extract IC50 values for drug combinations. We matched our identified drug combinations with the available IC50 data and found that 126 combinations have IC50 values of consistent types. After setting a threshold where both single-drug IC50 values after drug combinations were less than 10 $\mu$M, we found that 93 out of 126 matched items were retained, indicating an effectiveness rate of 73.8%. This indicates that our approach for identifying synergistic drug combinations is feasible and promising.

## Discussion

Many studies have shown that essential genes can be used for antimicrobial drug targets [25], but no method has been proposed to screen anticancer drugs based on a single essential gene. To this end, we have developed EGKG, which would help the precision treatment of cancers based on essential genes from the perspective of drug repurposing and provides a new idea for developing anticancer drugs targeting cancer-type-specific essential genes. We validated the effectiveness of all seven drugs identified by this strategy with lung adenocarcinoma cell inhibiting experiments. More importantly, it was found that there are significantly different ($P \leq .0001$) inhibiting rates between cancer cells and normal cells for the seven drugs. This suggests that those identified drugs with higher $e$ values could have higher inhibiting rates on associated cancer cells than on normal cells and could be regarded as having high inhibiting specificity or treatment safety. Furthermore, we observed a weakly negative correlation ($P = .062$) for the aggregative targets between the aggregative $m$ values of potentially bound targets and the viability of normal cells. Hence, the $m$ value of one essential gene could partly reflect the safety of the drug when binding to the target gene, and the $e$ score helped to identify those drugs having lower inhibiting rates on associated cancer cells than normal cells. These results validated our supposition that cancer-type-specific essential genes could be regarded as ideal targets, whose bound drugs have less toxicity than those bound to less specific essential genes. Precision medicine needs more antitumor drugs with higher safety and lower body toxicity [42]. Although it has been implied that drugs targeting specific essential genes could meet this requirement, here for the first time, we performed a systematic investigation of ranking drugs targeting cancer-type-specific essential genes and identified hundreds of such drugs applied in 19 cancer types.

Combinational therapy by multiple drugs is an emerging intervention, and several drugs may play roles through different mechanisms [43]. Here, we validated a novel combining strategy by targeting simultaneously genes with orthogonal essentialities. Consequently, treating the lung adenocarcinoma cell lines and normal cells simultaneously with both progesterone and rosiglitazone, the inhibiting rate for the former increased, whereas that for the latter decreased. Hence, this way of targeting genes with orthogonal essentialities could improve chemotherapy and reduce damage to normal cells.

These findings still need to be further updated and improved: (i) When using our identified drugs as clinical treatment for a specific cancer type, *in vitro* and/or *in vivo* experiments are expected to validate the chosen drugs. (ii) Find more examples where multiple targets with orthogonal essentialities can be used in combination with drugs and conduct experiments to verify their efficacy. (iii) Among the 19 cancer types, there are still 1051 targets without bound drugs. These cancer-type-specific targets are listed on the EGKG site, and this information, particularly on those targets with top rank, deserves special attention from researchers in the fields of targeted therapy and vaccine design. (iv) The EGKG table details the number of cell lines for each cancer type in our study. Gene essentiality is assessed based on its presence across these lines. Fewer cell lines might cause assessment inaccuracies. It is planned to improve our analysis with more extensive cell line data in the future for better identification of gene essentialities and drug targets. (v) We have evaluated the potential integration of additional datasets like TWOSIDES and PubChem. TWOSIDES focuses on DDIs rather than drug–target interactions, and PubChem's drug-target data largely overlap with our existing sources. In future work, we plan to leverage PubChem's molecule-protein interaction information to include more bioactive chemical molecules and further validate our findings.

**Key Points**

- A scoring-based strategy was developed to identify cancer-type-specific essential genes as preferred targets and also identify their bound drugs, which are supposed to have lower toxicity than those targeting pan-cancer essential genes.
- Seven drugs were validated in 11 lung adenocarcinoma cell lines, and lower inhibition rates (from 9.4% to 44.0%) were observed in 10 normal cell lines. This difference is statistically significant (Student's $t$-test, $P \leq .0001$), confirming the rationality of our supposition. Molecular docking results confirmed that these drugs targeted cancer-type-specific essential genes.
- Among the 370 anticancer drugs identified in 19 cancer types, 69 are applied to only one cancer type, and the cumulative number of identified drugs applicable to nine or fewer cancer types is 335, which accounts for 90.5% of all identified drugs. Hence, most of our identified drugs are not pan-cancer agents, and this feature would improve their safety.
- We have built the EGKG (Essential Gene Knowledge Graph) resource to reserve all our identified cancer-type-specific targets and drugs. Researchers could easily access these data through the functions of item browse and keyword search. EGKG forms a computational basis to uncover essential gene targets and drugs applicable to specific cancer types.
- We proposed and validated that combining drugs having targets with orthogonal essentiality could enhance anticancer effects while maintaining biocompatibility, inspiring an alternative approach to cancer synergetic therapy.

## Acknowledgements

We thank Miss Candy Sizhe Wu at Cornell University for polishing the language of our manuscript.

## Author contributions

F.G. and G.H. conceived the concept. X.L. performed the experiments under the supervision of Y.L. and Z.C. X.K. collected data, and X.L. carried out *in vitro* cell experiments. D.Z., Q.X., A.Y., C.W., H.C., and H.G. helped in data analysis. X.L. wrote the paper. J.Z. maintained the server.

## Supplementary data

Supplementary data is available at Briefings in Bioinformatics online.

Conflict of interest: None declared.

## Funding

This work was supported by the National Natural Science Foundation of China (grant no. 32370696), the Climbing Project for

Medical Talent of Zhongnan Hospital of Wuhan University (grant no. PDJH202406 and PDJH202205), and Noncommunicable Chronic Diseases-National Science and Technology Major Project (grant no. 2023ZD0506200).

## Data availability

The code used to process the data in EGKG of this work has been uploaded to FigShare (https://figshare.com/articles/software/EGKG_data_process/26771053?file=48632788).

## References


1. Hay M, Thomas DW, Craighead JL. *et al.* Clinical development success rates for investigational drugs. *Nat Biotechnol* 2014;**32**:40–51. https://doi.org/10.1038/nbt.2786
2. Kong J, Lee H, Kim D. *et al.* Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nat Commun* 2020;**11**:5485. https://doi.org/10.1038/s41467-020-19313-8
3. Zhou Z, Li M. Targeted therapies for cancer. *BMC Med* 2022;**20**:90. https://doi.org/10.1186/s12916-022-02287-3
4. Dai M, Yan G, Wang N. *et al. In vivo* genome-wide CRISPR screen reveals breast cancer vulnerabilities and synergistic mTOR/Hippo targeted combination therapy. *Nat Commun* 2021;**12**:3055. https://doi.org/10.1038/s41467-021-23316-4
5. Schenone M, Dančík V, Wagner BK. *et al.* Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 2013;**9**:232–40. https://doi.org/10.1038/nchembio.1199
6. Konda P, Garinet S, Van Allen EM. *et al.* Genome-guided discovery of cancer therapeutic targets. *Cell Rep* 2023;**42**:112978. https://doi.org/10.1016/j.celrep.2023.112978
7. Piñeiro-Yáñez E, Reboiro-Jato M, Gómez-López G. *et al.* PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data. *Genome Med* 2018;**10**:41. https://doi.org/10.1186/s13073-018-0546-1
8. Kuijk E, Kranenburg O, Cuppen E. *et al.* Common anti-cancer therapies induce somatic mutations in stem cells of healthy tissue. *Nat Commun* 2022;**13**:5915. https://doi.org/10.1038/s41467-022-33663-5
9. Zehir A, Benayed R, Shah RH. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017;**23**:703–13. https://doi.org/10.1038/nm.4333
10. Mirahmadi M, Azimi-Hashemi S, Saburi E. *et al.* Potential inhibitory effect of lycopene on prostate cancer. *Biomed Pharmacother* 2020;**129**:110459. https://doi.org/10.1016/j.biopha.2020.110459
11. Coombes RC. Drug testing in the patient: toward personalized cancer treatment. *Sci Transl Med* 2015;**7**:284ps10. https://doi.org/10.1126/scitranslmed.aab1214
12. Patel SJ, Sanjana NE, Kishton RJ. *et al.* Identification of essential genes for cancer immunotherapy. *Nature* 2017;**548**:537–42. https://doi.org/10.1038/nature23477
13. Xu T, Wang S, Ma T. *et al.* The identification of essential cellular genes is critical for validating drug targets. *Drug Discov Today* 2024;**29**:104215. https://doi.org/10.1016/j.drudis.2024.104215
14. Luo H, Lin Y, Liu T. *et al.* DEG 15, an update of the database of essential genes that includes built-in analysis tools. *Nucleic Acids Res* 2021;**49**:D677–86. https://doi.org/10.1093/nar/gkaa917
15. Gurumayum S, Jiang P, Hao X. *et al.* OGEE v3: online GEne essentiality database with increased coverage of organisms and human cell lines. *Nucleic Acids Res* 2021;**49**:D998–1003. https://doi.org/10.1093/nar/gkaa884
16. Liang YT, Luo H, Lin Y. *et al.* Recent advances in the characterization of essential genes and development of a database of essential genes. *Imeta* 2024;**3**:e157. https://doi.org/10.1002/imt2.157
17. Funk L, Su KC, Ly J. *et al.* The phenotypic landscape of essential human genes. *Cell* 2022;**185**:4634–4653.e22. https://doi.org/10.1016/j.cell.2022.10.017
18. Juhas M, Eberl L, Church GM. Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol* 2012;**30**:601–7. https://doi.org/10.1016/j.tibtech.2012.08.002
19. Ji X, Rajpal DK, Freudenberg JM. The essentiality of drug targets: an analysis of current literature and genomic databases. *Drug Discov Today* 2019;**24**:544–50. https://doi.org/10.1016/j.drudis.2018.11.002
20. Tyner JW. Functional genomics for personalized cancer therapy. *Sci Transl Med* 2014;**6**:243fs26. https://doi.org/10.1126/scitranslmed.3009586
21. Nguyen TTT, Kim YT, Jeong G. *et al.* Immunopathology of and potential therapeutics for secondary hemophagocytic lymphohistiocytosis/macrophage activation syndrome: a translational perspective. *Exp Mol Med* 2024;**56**:559–69. https://doi.org/10.1038/s12276-024-01182-6
22. Gonzalez-Fierro A, Dueñas-González A. Drug repurposing for cancer therapy, easier said than done. *Semin Cancer Biol* 2021;**68**:123–31. https://doi.org/10.1016/j.semcancer.2019.12.012
23. Wu Y, Zhang C, Liu X. *et al.* ARIH1 signaling promotes anti-tumor immunity by targeting PD-L1 for proteasomal degradation. *Nat Commun* 2021;**12**:2346. https://doi.org/10.1038/s41467-021-22467-8
24. Mayer RL, Verbeke R, Asselman C. *et al.* Immunopeptidomics-based design of mRNA vaccine formulations against listeria monocytogenes. *Nat Commun* 2022;**13**:6075. https://doi.org/10.1038/s41467-022-33721-y
25. Wang R, Shu X, Zhao H. *et al.* Associate toxin-antitoxin with CRISPR-Cas to kill multidrug-resistant pathogens. *Nat Commun* 2023;**14**:2078. https://doi.org/10.1038/s41467-023-37789-y
26. Behan FM, Iorio F, Picco G. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 2019;**568**:511–6. https://doi.org/10.1038/s41586-019-1103-9
27. Zhang B, Tang C, Yao Y. *et al.* The tumor therapy landscape of synthetic lethality. *Nat Commun* 2021;**12**:1275. https://doi.org/10.1038/s41467-021-21544-2
28. Chang L, Ruiz P, Ito T. *et al.* Targeting pan-essential genes in cancer: challenges and opportunities. *Cancer Cell* 2021;**39**:466–79. https://doi.org/10.1016/j.ccell.2020.12.008
29. Setton J, Zinda M, Riaz N. *et al.* Synthetic lethality in cancer therapeutics: the next generation. *Cancer Discov* 2021;**11**:1626–35. https://doi.org/10.1158/2159-8290.CD-20-1503
30. Martínez-Jiménez F, Muiños F, Sentís I. *et al.* A compendium of mutational cancer driver genes. *Nat Rev Cancer* 2020;**20**:555–72. https://doi.org/10.1038/s41568-020-0290-x
31. Bates SE. Epigenetic therapies for cancer. *N Engl J Med* 2020;**383**:650–63. https://doi.org/10.1056/NEJMra1805035
32. Terekhanova NV, Karpova A, Liang WW. *et al.* Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* 2023;**623**:432–41. https://doi.org/10.1038/s41586-023-06682-5
33. Hahn WC, Bader JS, Braun TP. *et al.* Cancer target discovery and development network. An expanded universe of cancer targets. *Cell* 2021;**184**:1142–55. https://doi.org/10.1016/j.cell.2021.02.020
34. Dietlein F, Wang AB, Fagre C. *et al.* Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science* 2022;**376**:eabg5601. https://doi.org/10.1126/science.abg5601

35. Zhou Y, Zhang Y, Lian X. *et al.* Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res* 2022;**50**:D1398–407. https://doi.org/10.1093/nar/gkab953
36. Pacini C, Duncan E, Gonçalves E. *et al.* A comprehensive clinically informed map of dependencies in cancer cells and framework for target prioritization. *Cancer Cell* 2024;**42**:301–316.e9. https://doi.org/10.1016/j.ccell.2023.12.016
37. Masumshah R, Eslahchi C. DPSP: a multimodal deep learning framework for polypharmacy side effects prediction. *Bioinform Adv* 2023;**3**:vbad110. https://doi.org/10.1093/bioadv/vbad110
38. Masumshah R, Aghdam R, Eslahchi C. A neural network-based method for polypharmacy side effects prediction. *BMC Bioinformatics* 2021;**22**:385. https://doi.org/10.1186/s12859-021-04298-y
39. Freshour SL, Kiwala S, Cotto KC. *et al.* Integration of the drug-gene interaction database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* 2021;**49**:D1144–51. https://doi.org/10.1093/nar/gkaa1084
40. Wishart DS, Feunang YD, Guo AC. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82. https://doi.org/10.1093/nar/gkx1037
41. Fang RH, Gao W, Zhang L. Targeting drugs to tumours using cell membrane-coated nanoparticles. *Nat Rev Clin Oncol* 2023;**20**:33–48. https://doi.org/10.1038/s41571-022-00699-x
42. Galetin A, Brouwer KLR, Tweedie D. *et al.* Membrane transporters in drug development and as determinants of precision medicine. *Nat Rev Drug Discov* 2024;**23**:255–80. https://doi.org/10.1038/s41573-023-00877-1
43. Chen S, Xie J, Ye R. *et al.* Structure-aware dual-target drug design through collaborative learning of pharmacophore combination and molecular simulation. *Chem Sci* 2024;**15**:10366–80. https://doi.org/10.1039/d4sc00094c


*Briefings in Bioinformatics*, 2025, **26(3)**, bbaf266
https://doi.org/10.1093/bib/bbaf266
Case Study