Acta Pharmaceutica Sinica B 2026;16(4):2299−2316

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

TOOLS

GEMap: A comprehensive gene essentiality map for drug discovery

Saisai Tian[a,*,†], Yixue Liang[b,†], Jinbo Zhang[a,c,†], Weishan Liang[b], Jinyuan Lu[d], Jiajia Ren[e], Bojian Lin[f], Ruijun Wu[d], Chengyang Guo[a], Weidong Zhang[a,b,d,g,*]

[a]*Department of Phytochemistry, School of Pharmacy, Second Military Medical University, Shanghai 200433, China*
[b]*Pharmacy College, Guangdong Pharmaceutical University, Guangdong 350122, China*
[c]*Tianjin Rehabilitation and Recuperation Center, Tianjin 300110, China*
[d]*The Research Center for Traditional Chinese Medicine, Shanghai Institute of Infectious Diseases and Biosafety, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[e]*School of Pharmacy, Henan University, Kaifeng 475004, China*
[f]*Department of Otolaryngology-Head & Neck Surgery, Changhai Hospital, Second Military Medical University, Shanghai 200433, China*
[g]*State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China*

Received 26 June 2025; received in revised form 14 October 2025; accepted 16 November 2025

KEY WORDS

GEMap;
CRISPR-Cas9;
Gene essentiality;
Molecular feature;
Drug discovery

Abstract Gene essentiality (synonymous with dependency) mapping reveals therapeutic vulnerabilities for intractable oncogenes, yet integrated platforms bridging CRISPR functional genomics with drug discovery remain limited. To address this gap, we developed GEMap, A Gene Essentiality-Guided Platform for Drug Discovery, by combining genome-wide CRISPR-Cas9 essentiality profiles (1912 screens across 1135 cancer cell lines) with multi-omics annotations, drug profiles and drug targets information (20,000+ compounds). GEMap can be used to (1) explore multimodal gene data (Dependency/Expression/CNV/Mutation) across cell lines and tissues; (2) prioritize context-specific therapeutic targets by quantifying differential genetic dependencies in molecularly stratified cancers, such as *KRAS* mutant-cancers; and (3) discover tailored treatments and drug candidates targeting particular gene using large-scale drug perturbation data and drug physical targets. To the best of our knowledge, GEMap represents the first

*Corresponding authors.
E-mail addresses: saisai_tian@foxmail.com (Saisai Tian), wdzhangy@hotmail.com (Weidong Zhang).
†These authors contributed equally to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2026.02.004

platform bridging CRISPR-derived genetic dependencies with pharmacological responses and biological networks and establishes an open-access paradigm for accelerating precision oncology against undruggable targets. GEMap is available at https://web.biotcm.net/GEMap/.

1. Introduction

Precision medicine, also referred to as "personalized medicine", represents a cutting-edge strategy for customizing disease prevention and therapeutic interventions by considering the unique genetic makeup, environmental factors, and lifestyle choices of each individual[1,2]. Advances in precision medicine facilitate appropriate treatments by enabling therapies tailored to individual patients. For instance, the use of trastuzumab to target HER2 in breast cancer[3], cetuximab to inhibit colorectal cancer with EGFR mutation[4], and vemurafenib to address mutant BRAF in melanoma[5] has significantly improved patient outcomes since these treatments were introduced into clinical practice. Nevertheless, significant efforts are still required to fully unlock the potential of precision medicine in cancer treatment. In recent years, the pace of identifying promising new therapeutic targets has slowed, and many therapies fail in clinical trials due to insufficient efficacy[6,7]. This underscores the pressing need to discover novel and viable targets that can be effectively modulated for therapeutic success. An ideal drug target should be selectively essential for disease processes, exhibit tissue-specificity, be amenable to drug development (often referred to as "druggable" targets), and, crucially, be associated with a biomarker to enable patient selection and the monitoring of treatment effectiveness[8].

High-throughput CRISPR-Cas9 essentiality screening has become an indispensable tool in precision medicine, providing robust capabilities for therapeutic target identification[9,10]. The Cancer Dependency Map (DepMap) initiative—employing genome-wide CRISPR essentiality screens across hundreds of cancer cell lines[11,12]—systematically quantifies gene dependency (defined as the degree of essentiality for cellular proliferation and survival) to identify cancer-specific therapeutic targets[13,14]. While this rich dataset serves as an invaluable resource for biologists and drug developers, its complexity often poses significant challenges for wet-lab research communities with little computational background. Several platforms aggregating CRISPR screening results have emerged, notably including the DepMap portal, GenomeCRISPR, PICKLES v3, iCSDB, shinyDepMap and OpenTargets. The DepMap portal provides extensive annotations for cell lines, encompassing baseline transcript levels, copy number variations, and mutational profiles for genes. However, its analytical capabilities remain relatively limited, offering constrained user interactivity and lacking native functionalities for integrating drug sensitivity data. GenomeCRISPR and PICKLES v3 primarily facilitate the exploration of fitness profiles for genes of interest[15,16]. Critically, their data content is outdated, and they provide only rudimentary results. Moreover, the absence of integration between gene dependency data and pharmacological information severely curtails their utility in translational research. Similarly, iCSDB currently lacks the capacity to integrate CRISPR screening data with drug sensitivity profiles[17]. This precludes its use for elucidating drug mechanisms of action or identifying novel drug candidates through repurposing strategies. ShinyDepMap focuses on drug response prediction and the identification of sensitive cell lines, but lacks multi-omics integration and does not provide a systematic pipeline for target prioritization[18]. OpenTargets excels in target−disease association analysis by integrating genetic and clinical evidence, yet incorporates CRISPR data only superficially and fails to support in-depth dependency or co-essentiality analyses[19]. In brief, while these platforms provide valuable insights, they exhibit significant limitations in comprehensive cancer research and the interactivity may not be user-friendly for non-bioinformaticians to manipulate and interpret. Most critically, none of these databases enable the prioritization of potential therapeutic targets or drug candidates tailored to cancer patients stratified by specific molecular features (such as *KRAS* mutation) based on gene essentiality data.

There is therefore an urgent need to develop a new platform for non-programmers to identify new promising targets and screen therapeutic agents. We present GEMap, a comprehensive Gene Essentiality Map platform that enables rapid identification of context-specific therapeutic targets and drug candidates for cancer subtypes characterized by genomic biomarkers (*e.g.*, *KRAS* mutations or *CCNE1* amplifications). GEMap integrates and preprocesses molecular data from DepMap_23Q2, Sanger, Avana, and GeCKO, supplemented by manually curated literature-derived datasets from PubMed. This includes 1912 screens across 1135 cell lines spanning 27 tissue types, which represents the most comprehensive CRISPR essentiality screening resource currently available. Additionally, more than 20,000 drug entries have been integrated from the Library of Integrated Network-based Cellular Signatures (LINCS, https://clue.io/data/CMap2020#LINCS2020)[20] and The Drug Gene Interaction Database (DGIdb)[21], enabling users to easily identify drug candidates targeting genes of interest. GEMap mainly includes four modules: Essentiality Analysis, Essentiality2Target, Essentiality2Drug and DIY Tools. Overall, this platform represents a significant advancement for the broader scientific community, particularly benefiting researchers with limited computational expertise. By seamlessly integrating sophisticated analytical functionalities with user-friendly visualization features within a unified workflow, it effectively bridges the gap between target discovery and clinical translation in the field of precision oncology.

2. Materials and methods

2.1. CRISPR-Cas9 gene essentiality data

Gene essentiality profiles were integrated from four primary sources: DepMap_23Q2[22] (17,931 genes × 1094 cell lines); Sanger CRISPR (17,799 genes × 316 cell lines); Avana (17,670 genes × 341 cell lines) and GeCKO v2 (18,478 genes × 43 cell

lines), collectively profiling 1135 unique cancer cell lines across 27 tissue types. To expand tissue coverage, we systematically curated literature-derived CRISPR screens (19,812 genes × 69 cell lines) from published studies, requiring documented gene dependency scores identified *via* PubMed searches using key terms, such as "CRISPR", "CRISPRi", "gene knockout", and "knock-out". The details of data processing workflow are shown in the Supporting Information. Briefly, all datasets were cleaned and reformatted into matrices with cell lines as rows and HGNC-approved gene symbols as columns. Identifier standardization was performed using Model.csv from DepMap_23Q2, with StrippedCellLineName for cell lines, HGNC symbols for genes, and OncotreeLineage for tissues. To improve consistency, a few tissue labels were changed (*e.g.*, "Ovary/Fallopian Tube" → "Ovary"), while others were retained as defined. The final datasets provided standardized matrices with genes as rows and cell lines as columns.

To maintain consistency across the integrated dataset, rigorous data cleaning and normalization procedures were applied. Raw data were processed into a unified format, and entries ambiguously labeled as "other" in the tissue or cell line classification were excluded. Meanwhile, organizational naming conventions were harmonized to prevent redundancy and ensure accurate cross-referencing.

2.2. Multi-modal data of cancer cell lines

Large-scale multi-modal cell line profiles were collected from the Cancer Cell Line Encyclopedia (CCLE, https://sites.broadinstitute.org/ccle)[23]. The dataset encompassed genomic data (somatic mutations), copy number alteration profiles, and genome-wide transcriptomic data (gene expression). Binary mutation matrices were generated for the Broad CCLE mutation data, *i.e.*, genes were defined as mutated if they matched the variant effects of "FRAME_SHIFT_INS", "IN_FRAME_DEL", "MISSENSE", "NONSENSE", "NONSTOP", "START_CODON_INS". The copy number was denoted as gain if $\log_2$(CN ratio +1) was greater than 1.6 and as loss if less than 0.6[24]. In addition, we performed stringent cell line matching and retaininged only cell lines present in both CCLE and DepMap_23Q2.

2.3. Drug perturbation profiles and drug−target interactions curation

2.3.1. Drug perturbational profiles (DPP)

The drug perturbational profiles data were obtained from the newly released CMAP LINCS2020 datasets (LINCS, https://clue.io/data/CMap2020#LINCS2020). Replicate-collapsed differential expression signatures (Level5 dataset) of the 978 landmark genes and 9105 BING (best-inferred) genes were used in our analysis pipeline. For accessing LINCS2020 signatures, we used cmapPy Python library. In this research, we focus on gene expression signatures induced by drugs at the standardized condition of 10 μmol/L concentration and 24 h exposure. The chosen time point and concentration represent the most frequent treatment condition present in the dataset. Then, the signatures corresponding to the same conditions (treatment, plate and cell line in case of compounds) were averaged using the MODZ method[20]. Finally, tissue-specific drug response signatures were obtained by averaging all of the profiles for each molecule by ignoring cell lines in the same tissue (Supporting Information Fig. S1A). Considering the potential of drug repurposing to efficiently discover new applications for already FDA-approved drugs in a time- and cost-effective manner, our main focus was on FDA-approved drugs sourced from the Drug Repurposing Hub (version: 3/24/2020), which is a carefully curated and annotated collection encompassing FDA-approved drugs, clinical trial drugs, and pre-clinical tool compounds.

2.3.2. Drug−target interactions

Drug-target interaction data were mainly sourced from DGIdb (https://dgidb.org), a publicly available database that compiles genes, gene products, drugs, and their interactions. Interaction data of DGIdb is derived from multiple sources, such as ChEMBL, DrugBank, DTC, OncoKB, PharmGKB and TTD. In this research, we mainly extracted and filtered drug-gene interactions by 10 interaction types, categorizing them as activating (agonist, activator, positive modulator, vaccine) or inhibitory (inhibitor, blocker, antibody, negative modulator, cleavage, inverse agonist).

2.4. Dependency-based target prioritization

The prioritization system consists of two core computational modules:

2.4.1. Essentiality2Target identification

Inspired by previous research[25], Essentiality2Target quantifies gene dependencies by jointly considering effect size (the difference between stratified groups) and statistical significance ($-\log_{10}P$ value) to establish the dependency-based target prioritization system. Unlike effect−size-only methods, Essentiality2Target introduces a transparent and interpretable scoring framework that integrates both statistical rigor and biological relevance. The incorporation of P value weighting represents a methodological refinement, yielding improved sensitivity and robustness across heterogeneous cancer contexts. Based on the dependency results, Essentiality2Target aims to identify potential therapeutic targets for cancers classified by their molecular features. This scoring system will assign each gene a score, and the top-rank genes will be regarded as potential therapeutic targets based on two dimensions: Difference and P value.

Here we assume that users are interested in gene *G1* and want to find new candidate therapeutic targets under *G1* mutations:

Step 1: The cell lines are categorized into two groups: the *G1* mutant group (X cell lines) and the wild-type group (Y cell lines). For each group, we require at least three cell lines in two groups (MUT vs WT).

Step 2: Compute the difference in mean dependency scores and the corresponding P value for each gene between the *G1* mutant and wild-type groups. To assess statistical significance, we apply the Wilcoxon test to compare the two independent samples (MUT *vs.* WT) and derive a P value. For each gene, we calculates: (1) the difference in mean dependency scores between *G1* mutant (MUT) and wild-type (WT) groups, and (2) the associated P value derived from Wilcoxon test comparing these independent sample groups (Eq. (1)).

$$Difference = \frac{\sum_{i=1}^{X} dependency_{MUT}}{X} - \frac{\sum_{i=1}^{Y} dependency_{WT}}{Y} \quad (1)$$

MUT: *G1* gene mutation group.

WT: *G1* wild type group.

Step 3: A negative difference value indicates that the gene in the mutant group exhibits stronger dependency compared to the WT group, suggesting that the gene is more essential in the presence of the mutation. The *P* value is the statistical significance between MUT and WT using the Wilcoxon test for two independent samples. To integrate both the effect size and the statistical robustness into a single metric, we defined the potential therapeutic target score using Eq. (2):

$$Score = -log_{10}(P\,value) * Difference \tag{2}$$

This formulation ensures that genes with both large dependency differences are prioritized. Importantly, because lower (more negative) difference values correspond to stronger dependencies in the mutant group, negative scores indicate higher targetability potential. Finally, each gene was assigned a targetability score, with lower (more negative) values indicating higher targetability potential.

For CNV (gain/loss), the procedure is similar to that described above. For clarity, we also provide a schematic diagram illustrating the methodological details of the algorithm (Fig. S1B). Furthermore, we performed a systematic comparison between our prioritization method and existing scoring schemes or machine learning-based prioritization models, including: XGBoost[26], Boruta[27], LASSO[28], SVM[29] and LightGBM[30]. The detailed results of comparison are provided in the Supporting Information. As shown in Supporting Information Fig. S2A−S2F, our approach achieves competitive performance compared with machine learning-based models, while offering greater interpretability and computational efficiency, which are important for large-scale CRISPR screening datasets.

2.4.2. *Tissue-specific Essentiality2Target*

An ideal therapeutic target should be selectively essential for disease processes, exhibit tissue-specificity, and be amenable to drug development (often referred to as "druggable" targets)[31]. Building on the therapeutic target identified in Section 2.4.1—where we assume the user has nominated *T1* as a candidate therapeutic target for *G1*-mutant cancers—we next assess which *G1*-altered malignancies exhibit heightened sensitivity to *T1* inhibition, thereby probing its tissue-specific therapeutic vulnerability. We evaluate the tissue-specific impact of target by analyzing dependency differences across distinct biological contexts. For each tissue, we required at least three cell lines in each group (MUT *vs*. WT) and tissues were ranked along two dimensions: difference and sensitivity, as outlined in the algorithm flow (Fig. S1C).

Here, we assume that users studying *G1-T1* seek to confirm tissues demonstrating optimal sensitivity and specificity for target *T1* under *G1* mutation conditions:

Step 1: The cell lines were categorized into two groups: the *G1* mutant group (X cell lines) and the wild-type group (Y cell lines).

Step 2: We computed two metrics: Sensitivity (defined as the mean dependency score of *T1* in the *G1* mutant group, Eq. (3)) and Difference (the delta in mean dependency scores of *T1* between the *G1* mutant and wild-type groups, Eq. (4)).

$$Sensitivity = \frac{\sum_{i=1}^{X} dependency_{MUT}}{X} \tag{3}$$

$$Difference = \frac{\sum_{i=1}^{X} dependency_{MUT}}{X} - \frac{\sum_{i=1}^{Y} dependency_{WT}}{Y} \tag{4}$$

The difference in sensitivity between the mutant and wild-type groups quantifies the tissue-specific effect of the mutation. More negative values indicate higher tissue sensitivity. For CNV (gain/loss), the procedure is similar to that described above.

2.5. *Architecture*

The GEMap platform is developed using the Shiny in R 4.4.1, featuring a modular architecture designed for enhanced functionality and scalability. The frontend, constructed with the shinydashboard package, delivers a user-friendly interface. A responsive design ensures seamless cross-platform compatibility. Backend data management employs a dual-storage optimization strategy: the qs package serializes non-tabular data, while the fst package enables high-performance storage and retrieval of tabular data, ensuring low-latency access under diverse query loads. The platform leverages Shiny's modules to decouple functional webpages into independent units, improving code maintainability and supporting future development. For data visualization, static outputs are generated using ggplot2, while interactive displays are powered by DT, plotly, and visNetwork.

2.6. *Cellular experiments*

2.6.1. *Cell culture*

The human breast cancer cell line MDA-MB-453 was maintained in L-15 medium containing 10% FBS and incubated at 37 °C in a CO_2-free environment. The human clear cell ovarian carcinoma cell line ES2 was cultured in McCoy's 5A medium with 10% FBS. The human ovarian adenocarcinoma cell line Caov-3 was maintained in high-glucose DMEM supplemented with 10% FBS. The two representative *KRAS*-mutant myeloid tumor cell lines—NB4 and NOMO-1, were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, and 100 μg/mL streptomycin. The *CCNE1*-amplified uterine cancer cell line MFE-280 was cultured in MEM (with non-essential amino acids, NEAA) containing 10% FBS and KLE was grown in DMEM/F12 medium supplemented with 10% FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin. All cell lines except MDA-MB-453 were incubated at 37 °C in a humidified atmosphere with 5% CO_2. Cells were passaged when they reached the logarithmic growth phase. The NOMO-1, NB4, MDA-MB-453, KLE, and ES2 cell lines were purchased from Hongshun Biotechnology Co., Ltd. (Shanghai, China). The Caov-3 and MFE-280 cell lines were obtained from Zhongqiao XinZhou Biotechnology Co., Ltd. (Shanghai, China). All compounds were purchased from TargetMol Biotechnology Co., Ltd. (Shanghai).

2.6.2. *Cell transfection for gene silencing*

The CDK4-targeting siRNA and negative control (NC) siRNA were purchased from GenePharma (Shanghai, China). All sequence information is detailed in Supporting Information Table S1. For transfection experiments, MDA-MB-453, ES2, Caov-3, and MFE-280 cells were seeded in 6-well plates and cultured until they reach approximately 50% confluence before transfection. Transient transfection was performed using

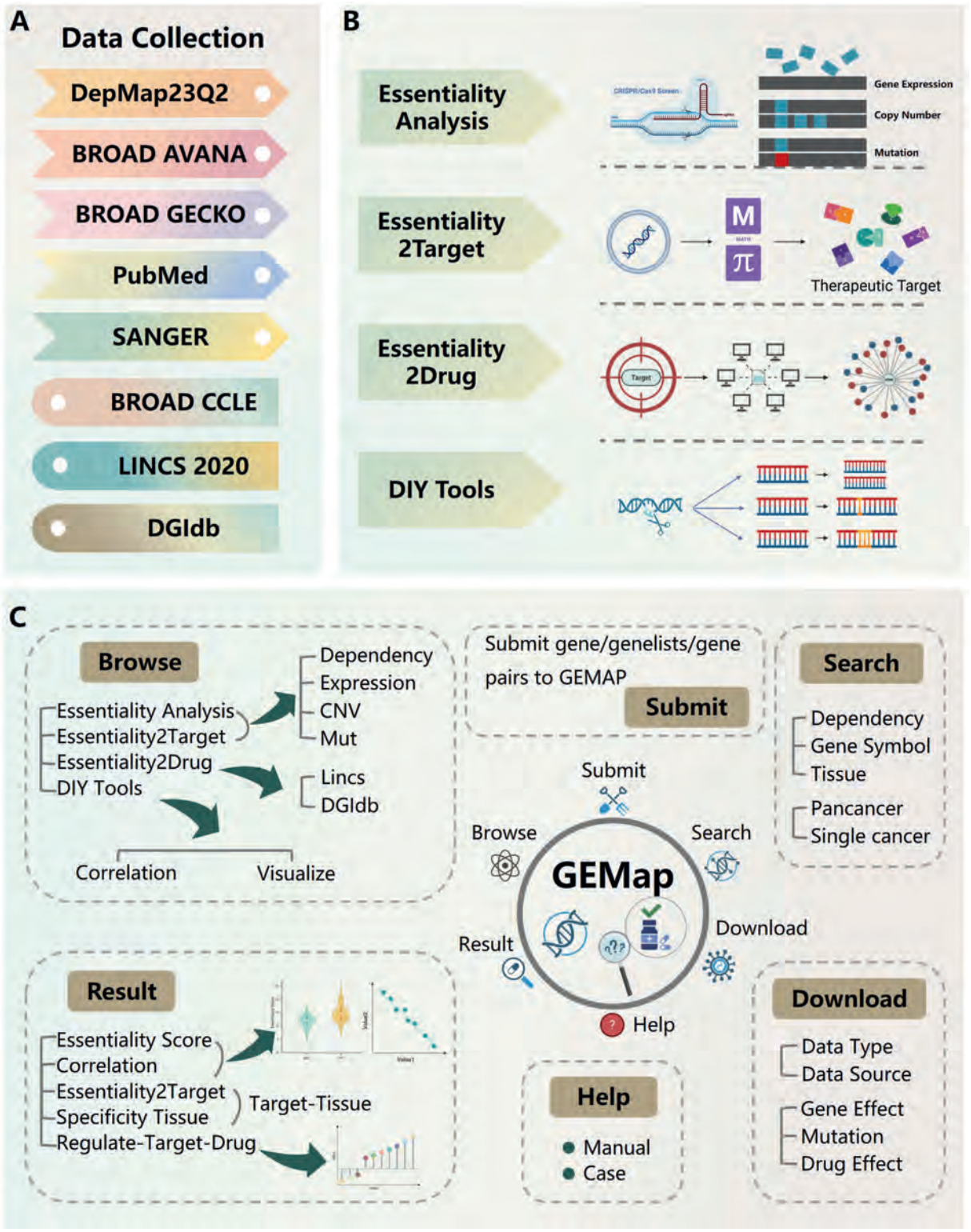

Figure 1 Overview of GEMap. (A) Data sources integrated into GEMap. (B) The functional modules of GEMap are illustrated. The Essentiality Analysis module identifies cell lines that exhibit sensitivity to gene knockout. Essentiality2Target and Essentiality2Drug are primarily designed to identify candidate therapeutic targets and drugs. The DIY Tools module presents the essentiality scores of the query gene relative to the comparison gene across several different scales. (C) The web interface that allows users to browse, submit, search, and download.

Table 1 Comparison of GEMap with six major CRISPR screening databases.

Category	Feature	Detail	GEMap	DepMap	GenomeCRISPR	PICKLES v3	iCSDB	shinyDepMap	OpenTargets
Data statistic	Screens		1912	1064	/	/	1375	/	/
	Cell lines		1135	1064	48	1162	976	423	/
	Tissues		27	27	/	31	28	28	28
	Genes		20,000+	20,000+	/	18,959	/	15,847	63,121
	Drugs		21,334	7487	/	/	265	/	18,041
	Disease		/	/	/	/	/	/	38,959
Functional features	Dependency analysis	DepMap screen	√	√	×	√	√	√	√
		RNAi screen	×	√	×	×	×	√	×
		TKOv3 screen	×	×	×	√	×	×	×
		GeCKO screen	√	×	×	×	√	×	×
		Avana screen	√	×	×	√	×	×	×
		Sanger screen	√	×	×	√	√	×	×
		PubMed screen	√	×	√	×	√	×	×
	Gene analysis	Expression	√	√	×	×	×	×	√
		Mutation	√	√	×	×	×	×	×
		Copy number	√	√	√	×	×	×	×
		Perturbagens	√	×	×	×	√	×	×
	Correlation analysis	Dep[a]-Dep[a]	√	√	×	√	×	√	×
		Dep[a]-Exp[a]	√	√	×	√	×	×	×
		Dep[a]-Mut[a]	√	×	×	√	×	×	×
		Dep[a]-CN[a]	√	√	×	×	×	×	×
	Molecular stratification	Gene-Mut[a]	√	×	×	×	×	×	×
		Gene-Amp[a]	√	×	×	×	×	×	×
		Gene-Loss[a]	√	×	×	×	×	×	×
	Therapeutic target prioritization		√	×	×	×	×	√	×
	Tissue-specific analysis		√	×	×	×	×	×	×
	Drug candidate identification		√	×	×	×	×	×	×
	Target−disease		×	×	×	×	×	×	√

[a]Note:Dep*-Dependency, Exp*-Expression, Mut*-Mutation, CN*-Copy Number, Amp*-Copy Number Amplification, Loss*-Copy Number Loss.

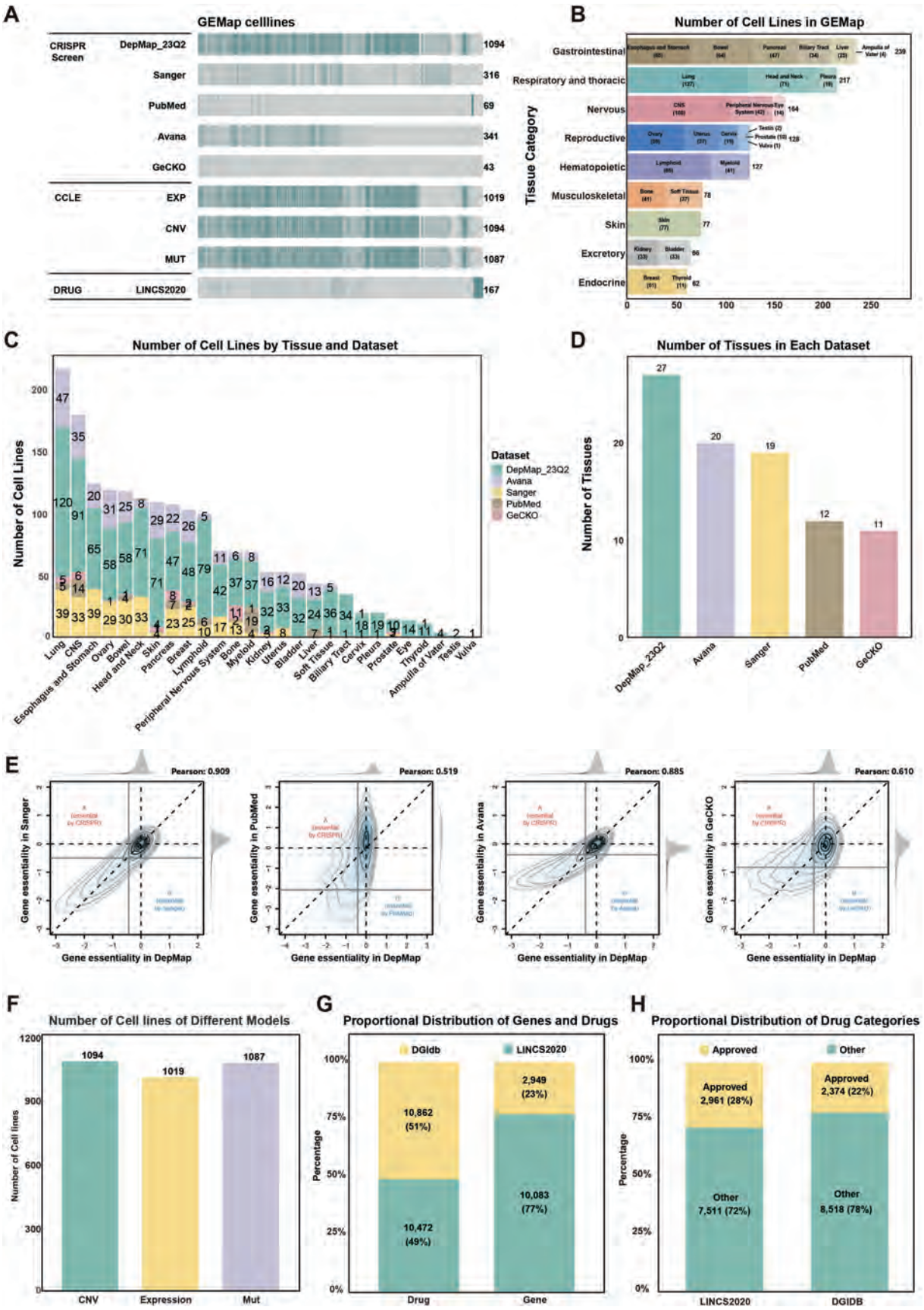

Figure 2 Statistics of the GEMap. (A) Statistical analysis of cell line distribution in GEMap. (B) Tissue type distribution analysis in GEMap. (C) Cell line information in different tissues across datasets. (D) Number of tissues in each dataset. (E) Correlation analysis between DepMap_23Q2 and (Sanger, PubMed, Avana, GeCKO) datasets. (F) CCLE dataset cell line composition analysis. (G) Gene and drug information statistics in DGIdb and LINCS. (H) FDA-approved drugs information statistics in DGIdb and LINCS.

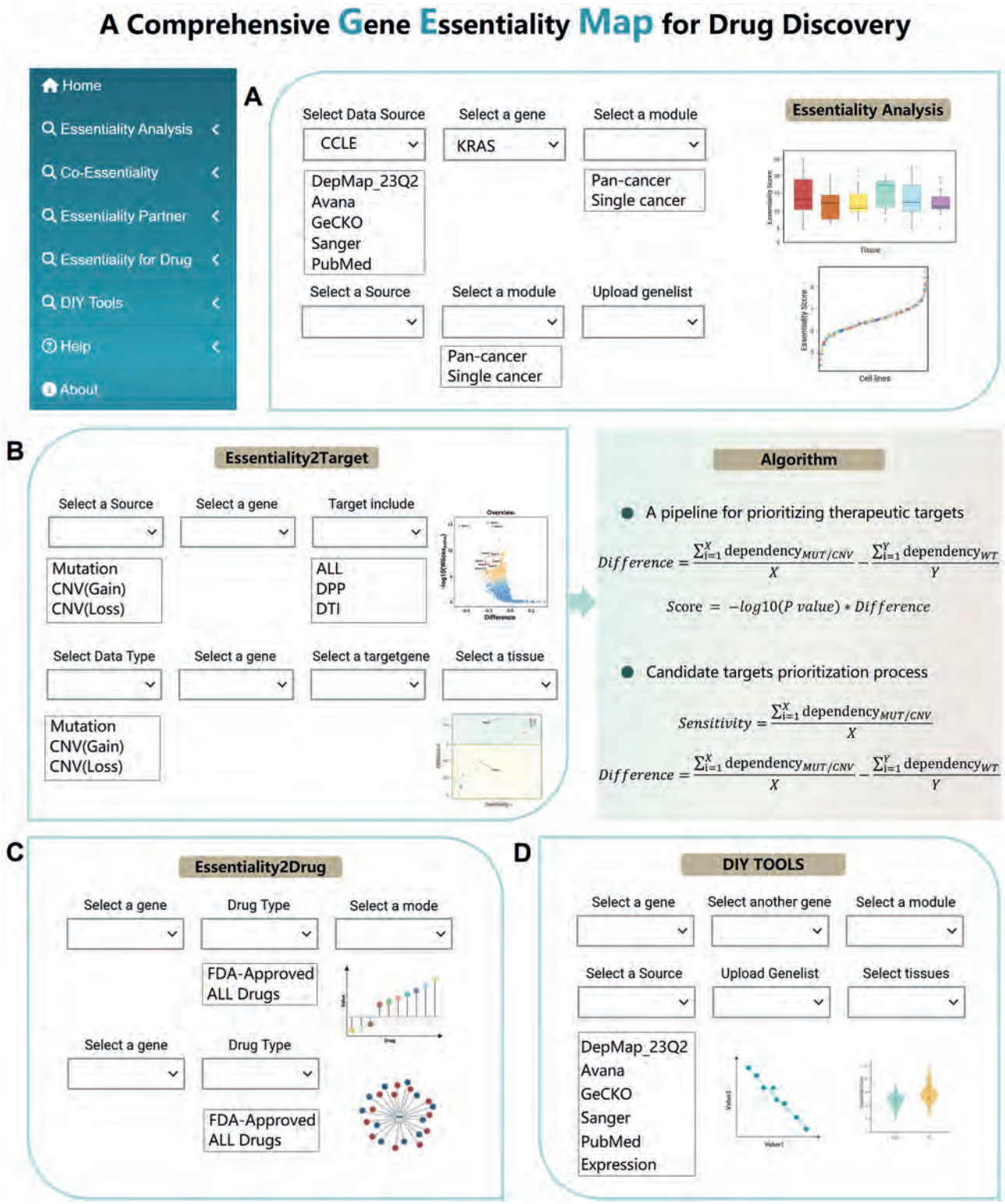

Figure 3 Schematic workflow of GEMap. (A) The interface of 'Essentiality Analysis' page. (B) The interface of 'Essentiality2Target' page. (C) The interface of 'Essentiality2Drug' page. (D) The interface of 'DIY Tools' page.

Lipofectamine 3000 reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. Post-transfection cells were maintained under standard culture conditions for 48−72 h before CCK-8 assay.

2.6.3. Cell viability assay

Cell viability was assessed using the Cell Counting Kit-8 (CCK-8; Sijia Biotechnology, China) according to the manufacturer's instructions. Briefly, NOMO-1 and NB4 cells were seeded into 96-

well plates and allowed to adhere for 2 h. Cells were then treated with the indicated compounds for 24 h. Subsequently, 20 μL of CCK-8 solution was added to each well and incubated for an additional 2 h at 37 °C. For KLE and MFE-280 cells, they were seeded and incubated for 24 h prior to compound treatment, which was applied for 48 h. After treatment, 10 μL of CCK-8 solution was added per well and incubated for 2 h. The optical density (OD) at 450 nm was measured using a BMG LABTECH POLARstar Omega microplate reader (BMG LABTECH, Ortenberg, Germany). Cell viability was calculated using Eq. (5):

$$\text{Cell viability}\ (\%) = (\text{OD value of the treatment group} / \text{OD value of the control group}) \times 100 \quad (5)$$

3. Results

3.1. Overview

Despite advances in precision oncology, targeting oncogenic drivers like *KRAS* mutations or *CCNE1* amplifications remains challenging. We present GEMap (https://web.biotcm.net/GEMap), a translational platform integrating genome-wide CRISPR essentiality profiles (1135 cell lines across 27 tissues) with drug perturbation networks (20,000+ compounds) and multi-omics annotations. The database consolidates CRISPR screening data from DepMap_23Q2, Sanger, Avana, GeCKO and supplements these with manually curated literature-derived datasets from PubMed (Fig. 1A). From the PubMed mining, we compiled 19,812 genes across 69 cell lines representing 12 distinct tissue types (Supporting Information Table S2). GEMap's core innovation lies in its cross-modal integration of molecular profiles (mutations, expression, CNV) with drug response signatures and target interactions, enabling mechanistic drug exploration through two specialized modules Essentiality2Target and Essentiality2Drug (Fig. 1B). This dependency−pharmacology bridge accelerates discovery for intractable targets and provides an intuitive user interface that enables users to browse, query, submit, download, and visualize detailed gene of interest (Fig. 1C).

Compared to existing CRISPR essentiality screening platforms that suffer from limited analytical capabilities, inability to prioritize molecularly stratified therapies, and clinical translation barriers, GEMap overcomes these limitations as the first platform that uniquely integrates CRISPR functional genomics and drug discovery, representing the most comprehensive CRISPR essentiality screening resource currently available. By quantifying differential genetic dependencies in molecularly stratified cancers, GEMap enables precise prioritization of context-specific therapeutic targets, while its integration of large-scale drug perturbation data and validated drug-target interactions facilitates identification of tailored treatments and drug candidates for specific gene (Table 1).

3.2. Data statistics

The current version of GEMap comprehensively documents 1912 CRISPR screening data across 1135 cell lines spanning 27 distinct tissue types from multiple CRISPR sources, CCLE and LINCS2020, with DepMap_23Q2 representing the largest contributor of cell lines and GeCKO containing the fewest (Fig. 2A). Then, all tissues in GEMap were classified into nine major organ systems. Notably, the gastrointestinal system showed the highest representation in both tissue diversity and cell line abundance, while skin system contained the fewest tissue types and the endocrine system had the lowest cell line coverage (Fig. 2B). For each CRISPR screening data, DepMap_23Q2 showed the broadest tissue coverage (27 types, 30.3%), while GeCKO had the fewest (12.4%) (Fig. 2D, Supporting Information Fig. S3A). Eight tissues were shared across all datasets (Fig. S3B). DepMap_23Q2 contained 58.7% of cell lines (*vs*. GeCKO's 2.3%) (Fig. S3C). Cell lines derived from Eye, Ampulla of Vater, Testis, and Vulva tissues were exclusively present in DepMap_23Q2, with Lung being most represented and Vulva least (Fig. 2C), comprising 522 unique cell lines (Fig. S3D).

As we know, integrated multi-source data analysis could provide robust cross-validation, significantly improving result reliability. Our comparative analysis of dependency scores revealed high Pearson correlation coefficients between DepMap_23Q2 and four major datasets (0.909 with Sanger, 0.519 with PubMed, 0.885 with Avana, and 0.610 with GeCKO; Fig. 2E), demonstrating robust consistency across these CRISPR screening platforms. Regarding the relatively lower correlation with PubMed compared to other datasets, this is likely due to the smaller sample size and tissue heterogeneity of PubMed data. Unlike DepMap_23Q2, Sanger, Avana, and GeCKO, which are large-scale CRISPR screens generated under standardized experimental protocols, the PubMed dataset was collected from diverse published studies using variable platforms, assay conditions, and analytical pipelines. These differences inevitably reduce the consistency with DepMap_23Q2. Nevertheless, the positive correlation still indicates a consistent signal across independent sources, supporting the robustness of our integrated analysis. Given that the DepMap_23Q2 dataset contains the largest collection of cell lines, it served as the primary data foundation for all subsequent analyses. To strengthen functional genomic analysis in GEMap, we integrate comprehensive cell line annotations for CRISPR essentiality profiles, featuring 1094 CNV profiles, 1019 expression datasets, and 1087 mutation data (Fig. 2F).

For drug discovery, we uniquely combine the DPP (LINCS2020; 10,472 compounds targeting 10,083 genes) with DTI information (DGIdb; 10,862 compounds targeting 2949 genes), enabling dependency-based drug prioritization (Fig. 2G). In DPP, the prostate tissue contained the highest number of drugs (9545 drugs), while the head and neck tissue had the fewest (108 drugs) (Supporting Information Fig. S4B). Lung, breast, and lymphoid tissues included a relatively high number of cell lines (Fig. S4A). To enable systematic drug repurposing, FDA-approved drugs were annotated in both the DPP and DTI datasets. The DPP dataset included 2961 clinically approved compounds, representing 28.3% of the total. By comparison, the DTI dataset contained 2374 clinically relevant compounds, representing 22% of the total (Fig. 2H). Within the DTI dataset, FDA-approved drugs were further categorized as either activators or inhibitors (Fig. S4C). In DTI, 36.7% of activator drugs and 40.9% of inhibitor drugs were FDA-approved (Fig. S4D). Further analysis of activators and inhibitors revealed that "agonist" and "inhibitor" were the most common drug types (Fig. S4E and S4F).

3.3. The web interface and analytical tools of GEMap

GEMap provides a user-friendly web interface that serves as an intuitive platform for browsing, searching, visualizing, and retrieving human CRISPR-Cas9 screening data (Fig. 3). The

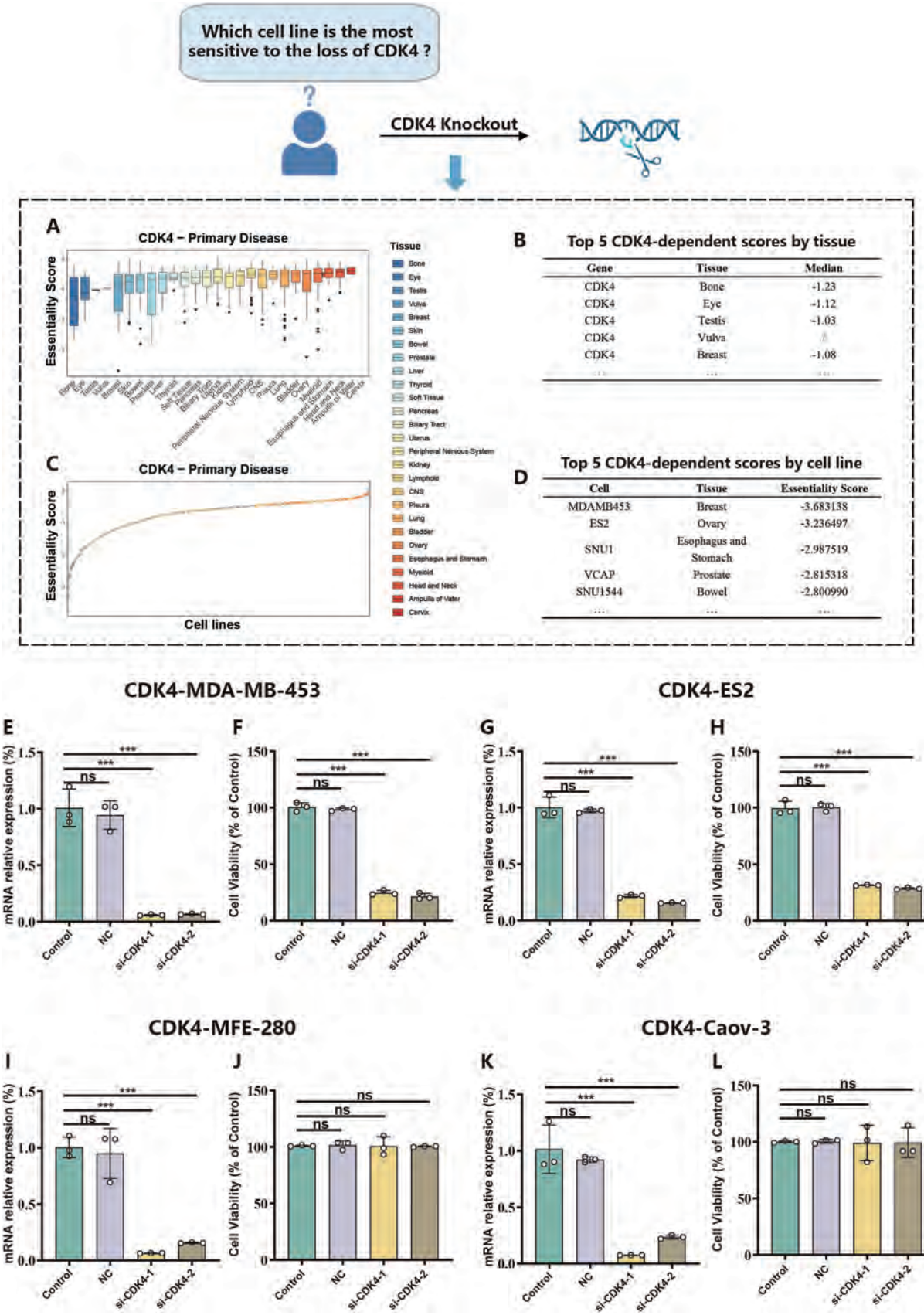

Figure 4 Identification of *CDK4* knockout-sensitive cell lines. (A) Tissues sensitive to *CDK4* knockout ranked by median score. Negative score means heightened genetic essentiality. (B) Top five tissues ranked by *CDK4* knockout sensitivity. (C) Sensitivity ranking of cell lines following *CDK4* knockout. (D) Top five cell lines ranked by *CDK4* knockout sensitivity. (E−H) Validation of *CDK4* dependency in MDA-MB-453 (breast

platform mainly includes four modules: Essentiality Analysis, Essentiality2Target, Essentiality2Drug and DIY Tools.

3.3.1. Essentiality Analysis

The Essentiality Analysis module facilitates exploration of multimodal gene data across diverse cell lines and tissue types, including Dependency, Expression, CNV and Mutation (Fig. 3A). In the Dependency, Expression, CNV section, users can input a gene symbol (*e.g.*, *KRAS*) in the "Select a gene" field and click "Go!" to generate boxplots ranked by median values. Hovering over the boxplots reveals interactive details, including max, upper fence, Q3, median, Q1, lower fence, and min values. The Mutation section presents mutation analysis through a bubble plot, where mutation frequency and distribution are visualized *via* color gradients and bubble sizes. Interactive tooltips display Gene, Tissue, and Proportion when hovering over the plot. Additionally, a result table is provided, listing the gene symbol, tissue name, number of mutant cell lines, total cell lines per tissue, and mutation rate. All visualizations and data tables are available for download.

3.3.2. Essentiality2Target

Browsing multimodal data for a gene of interest across cell lines and tissues with the Essentiality Analysis tool is simple. However, if users want to find potential therapeutic targets for patients with cancer who are classified by their molecular features, for example, finding potential targets for *KRAS*-mutant cancer, few tools provide functionality. Essentiality2Target module allows users to identify therapeutic targets for cancer patients classified by molecular features. This module includes two sections: Pancancer section and Tissue section (Fig. 3B). In the Pancancer section, users can define the analytical scope of genes, classified into DPP and DTI. The Tissue section further allows users to filter specific tissue types within the DPP.

The Pancancer section of the Essentiality2Target module enables users to predict and prioritize therapeutic targets based on the mutation or copy number variation (CNV) of a gene of interest. For example, if a user is studying *KRAS*-mutated cancers, they can select "Mutation" in the "Select a source" field, input the gene symbol (*e.g.*, *KRAS*) in the "Select a gene" field, and click "Go!" to generate results. The module provides a dotplot and a boxplot to offer a visual representation of genes most affected by the mutation by comparing mutant (MUT) versus wild-type (WT) samples. Additionally, a results table is displayed, containing Gene, Score, and *P* value information for further analysis.

If users aim to identify which cancer type with *KRAS* mutation exhibits greater sensitivity to this therapeutic target, users can navigate to the Tissue section and select "Mutation" in the "Select Data Type" field. By entering the gene symbol (*e.g.*, *KRAS*) in the "Select a Mutation" field and specifying a predicted target gene (*e.g.*, *RAF1*) in the "Select a targetGene" field, clicking "Go!" will generate a graph. Points located closer to the bottom-left corner indicate higher sensitivity in that tissue. A corresponding results table, listing Tissue, Sensitivity, and Difference, helps assess the therapeutic potential across different cancer types. We have performed a systematic comparison between our prioritization method and widely used approaches, including five machine learning−based ranking models (XGBoost[26], Boruta[27], LASSO[28], SVM[29] and LightGBM[30]). To benchmark performance, we compiled gold-standard gene sets from literatures, including 849 *KRAS*-mutant-associated genes and 34 *CCNE1*-amplification-associated genes (Supporting Information Tables S3 and S4). Using these as references, we compared Essentiality2Target against multiple widely used prioritization models. In both comparative scenarios, we adopted evaluation metrics from prior studies[32], including Top-K accuracy and Recall@K. The detailed comparison is provided in the Supporting Information. These metrics provide a more comprehensive assessment of ranking performance. As shown in Supporting Information Fig. S2, our approach achieves competitive performance compared with machine learning−based models. Meanwhile, we additionally performed enrichment validation. Notably, a significant enrichment of the gold-standard gene sets from literatures was observed at the top of the prediction results (*KRAS: $P < 0.05$; CCNE1: $P < 0.001$*; Fig. S2E and S2F), further validating the analysis.

3.3.3. Essentiality2Drug

The Essentiality2Drug module facilitates drug discovery by allowing users to identify potential therapeutic compounds through two key drug-target relationships: Drug Perturbational Profiles (DPP) and Drug Target Interactions (DTI) (Fig. 3C). DPP characterizes genome-wide transcriptional regulatory relationships between drugs and genes, identifying whether a drug upregulates or downregulates specific genes and quantifying the magnitude of these effects. Notably, DPP reflects transcriptional regulation rather than direct physical binding between a drug and its target. DTI refers to the direct binding of a pharmaceutical compound to the active site of a specific biological target, thereby modulating its functional activity, such as inhibition or activation.

In the DPP section, users start by inputting a potential therapeutic target (*e.g.*, *RAF1* for *KRAS*-mutated cancers) in the "Select a gene" field. They can then enter the most sensitive specific tissue (*e.g.*, Myeloid) identified in the Tissue section and click "Go!". The system generates a bar and rod graph displaying the top 15 upregulated and downregulated drug candidates for the specified tissue. Additionally, a results table is provided that lists the drug name, tissue, and *Z*-score ($Z > 0$ means up-regulated). This output enables users to evaluate the most promising drug candidates based on gene perturbation effects. Similarly, in the DTI section, users can input a potential therapeutic target in the "Select a gene" field to generate a drug-target bipartite network. A results table is also provided, listing the gene, drug, and directionality. Approved drugs can be screened in the Essentiality2Drug module by selecting "Approved Drugs" in the drug type filter.

3.3.4. DIY tools

DIY Tools can be categorized into Correlation and Visualize Your Way (Fig. 3D). The Correlation precomputes and visualizes the correlation between Dependency and Dependency (Co-Dependency), Expression (Dependency/Expression), CNV (Dependency/CNV), Mutation (Dependency/Mutation) of two genes. It has been reported that co-functional genes with correlated dependency profile tend to comprise a functional unit (*e.g.*, a pathway or a complex)[33,34]. These tools enable the exploration of

cancer) and ES2 (ovarian cancer) cells using CCK-8 assays after *CDK4* knockdown with two independent siRNAs. (I−L) CCK-8 assays in MFE-280 (endometrial cancer) and Caov-3 (ovarian cancer) cells, which are less dependent on *CDK4*, serving as negative controls. Data are presented as mean ± SD ($n = 3$). Statistical significance was determined by one-way ANOVA: ns, not significant; ***$P < 0.001$.

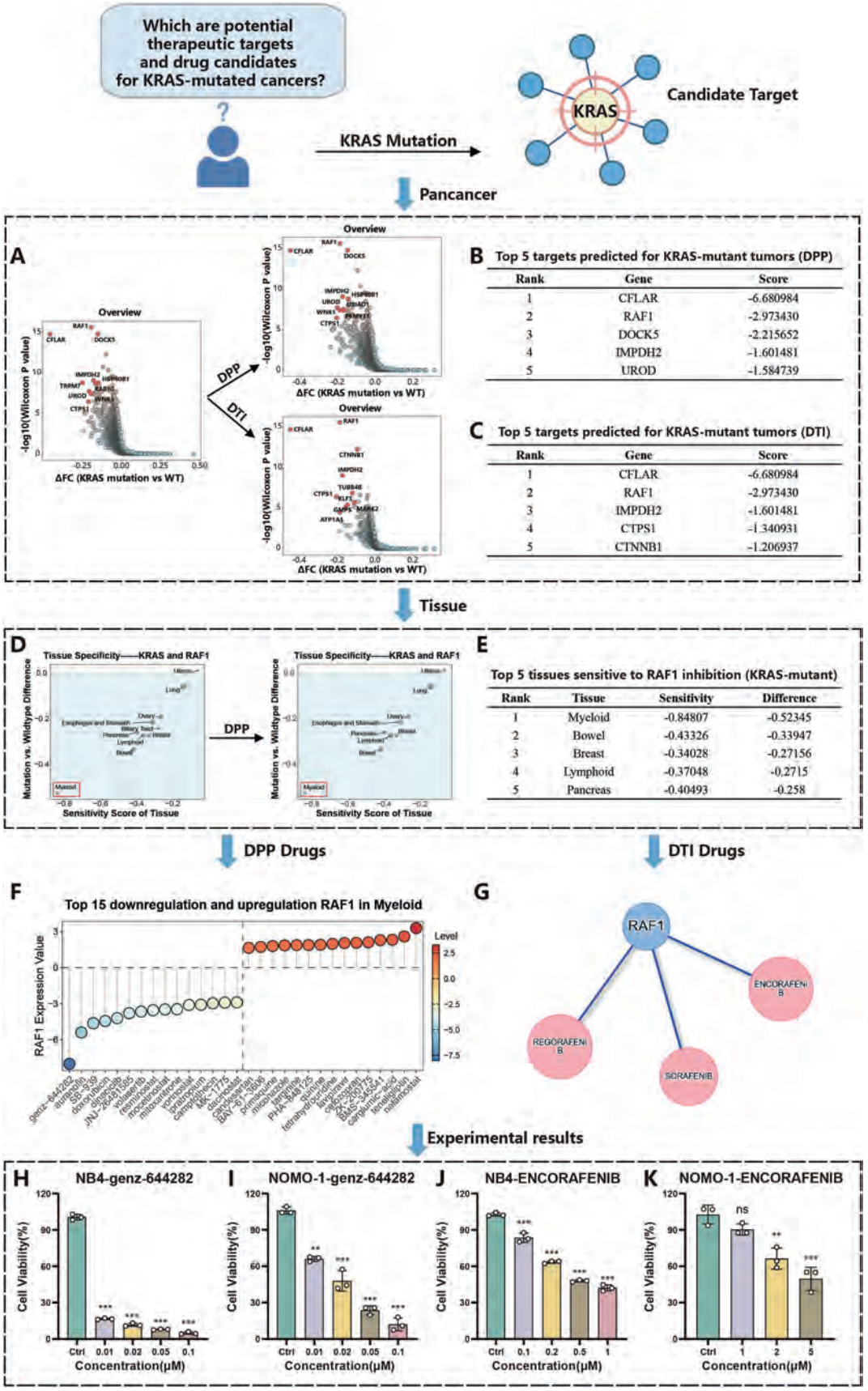

Figure 5 Discovery of therapeutic targets and drug candidates for *KRAS*-mutant cancers. (A) Potential therapeutic targets identified in ALL, DPP, and DTI. (B, C) Top 5 targets predicted for *KRAS*-mutant tumors in DPP and DTI. (D) *RAF1*-sensitive tissues stratified by *KRAS*-mutant

relationships between gene dependency and molecular features (such as expression, CNV, and mutation) to identify co-functional genes. In the first three sections, users can input two gene symbols in the "Select a gene" fields and click "Go!" to visualize the correlation across tissues for Dependence, Expression, and Copy Number. The Dependence/Mutation section allows users to enter two gene symbols and click "Go!" to display gene dependence differences between MUT and WT groups. Tissues with smaller sample sizes, such as vulva and testis, are ranked lower based on correlation strength. In the Visualize Your Way section, users can upload a gene list of interest and select target tissues to analyze gene dependence and expression patterns.

3.4. Case study 1: Which cell line is the most sensitive to the loss of CDK4?

CDK4, an essential gene in cancer development, demonstrates cell line-specific essentiality, meaning randomly selected models may not exhibit survival dependency on this gene[35]. In the Essentiality Analysis module, the gene *CDK4* was analyzed for dependency across tissues and cell lines. When sorted by median dependency scores, tissues including Bone, Eye, Testis, Vulva, and Breast displayed the most pronounced negative scores, reflecting heightened genetic essentiality (Fig. 4A and B). Analysis of all cell lines identified MDA-MB-453, ES2, SNU1, VCAP, and SNU1544 as exhibiting the strongest negative dependency scores (Fig. 4C and D). Of particular significance, the triple-negative breast cancer (TNBC) cell line MDA-MB-453 showed the most substantial *CDK4* dependence, achieving the highest negative score among all breast tissue-derived models. These results collectively indicate that MDA-MB-453 represents the most *CDK4*-knockdown-sensitive breast cancer cell line model. We also examined which cell line is the most sensitive to CDK4 loss in the Avana, PubMed, Sanger, and GeCKO datasets. Consistently, MDA-MB-453 was identified as the most sensitive cell line in both the Avana and Sanger datasets (Supporting Information Figs. S5 and S6). To experimentally validate the accuracy of computational predictions, two sensitive cell lines (top two cell lines, MDA-MB-453, ES2) were selected to test the survival dependency of *CDK4 in vitro*. Meanwhile, we also selected Caov-3, and MFE-280 cell lines, which were the bottom two non-sensitive cell lines as negative controls. We then performed qPCR and CCK-8 assays following *CDK4* knockdown in selected cell lines. Results of gene silencing demonstrated that siRNA transfection markedly reduced *CDK4* mRNA expression in MDA-MB-453, ES2, Caov-3, and MFE-280 cells (Fig. 4E, G, I, and K). Consistent with our predictions, MDA-MB-453 (breast cancer) and ES2 (ovary cancer) cells, which were identified as *CDK4*-dependent, exhibited a significant decrease in cell viability upon *CDK4* silencing (Fig. 4F and H). In contrast, MFE-280 and Caov-3 cells, predicted to be less dependent on *CDK4*, showed no significant change in viability (Fig. 4J and L). Together, these results demonstrate that our platform accurately predicts *CDK4* dependency across different cell lines.

3.5. Case study 2: Can we find potential therapeutic targets and drug candidates for KRAS-mutant cancers?

KRAS is a notorious cancer driver gene that is frequently mutated in many cancers, but *KRAS* protein has a very smooth surface, which makes it difficult to inhibit by a drug[36]. In such a case, we can think of a "target hop" strategy that targets other more druggable proteins instead of trying to inhibit *KRAS* directly[31]. In line with this hypothesis, we screened candidate targets according to the dependency-based target prioritization pipeline.

In the Pancancer section of Essentiality2Target module, the *KRAS* gene was selected for screening. Across distinct curated datasets, *RAF1* and *CFLAR* consistently emerged as top candidate targets—ranking among the top three in ALL (alongside *DOCK5*), DPP (also with *DOCK5*), and DPI (paired with *IMPDH2*) (Fig. 5A–C). Due to their recurrent prominence, *CFLAR* and *RAF1* were prioritized as key therapeutic targets for further investigation. Previous studies have identified *CFLAR* as a potential therapeutic target in *KRAS*-mutant pancreatic ductal adenocarcinoma through genome sequencing and high-throughput RNAi screening[37]. Notably, *CFLAR* silencing significantly impaired cell viability, further validating its role as a promising therapeutic target. Additional evidence has also shown that *KRAS* mutations (including G12C and G12D) lead to constitutive activation of the *KRAS* protein, resulting in persistent activation of downstream *RAF1* and subsequent tumor initiation and progression[38]. Targeting *RAF1* inhibition therefore represents a potential therapeutic strategy for *KRAS*-mutant tumors. If users aim to identify which cancer type with *KRAS* mutation exhibits greater sensitivity to this therapeutic target, users can navigate to the Tissue section. We selected *KRAS* and *RAF1* as representative examples. Their tissue specificity analysis revealed the highest sensitivity in Myeloid cancer (Fig. 5D and E), suggesting strong tissue-specific relevance.

Consequently, we focused on Myeloid cancer and proceeded to the Essentiality2Drug module to explore candidate drugs targeting *RAF1*. Our DPP analysis identified the top 15 candidate compounds for *RAF1* downregulation (Supporting Information Table S5). Screening of FDA-approved drugs similarly revealed top 15 candidates for *RAF1* downregulation (Fig. 5F, Supporting Information Table S6). Literature mining confirmed the efficacy of auranofin, doxorubicin, dinaciclib, and mocetinostat against *KRAS*-mutant tumors[39,40]. For example, auranofin suppresses *KRAS*-mutant lung adenocarcinoma growth primarily through *ECT2* targeting[41], while doxorubicin acts by disrupting cancer cell DNA repair mechanisms[42]. To further confirm the reliability of our predictions, we performed *in vitro* CCK-8 assays on two representative *KRAS*-mutant myeloid cancer cell lines, NB4 and NOMO-1, using the drugs predicted by DPP. As shown in the Supporting Information Fig. S7, nearly all the compounds showed obvious anti-tumor effects, especially for the genz-644282, which significantly inhibited the proliferation of NB4 cells and NOMO-1 cells at concentrations of 0.01 μmol/L and 0.1 μmol/L, respectively (Fig. 5H and I). Results for the other

cancers. (E) Top 5 tissues sensitive to *RAF1* inhibition. (F) The top 15 compounds that upregulate and downregulate *RAF1* in Myeloid cancer from DPP analysis. (G) Promising drug candidates from DTI analysis. (H–K) CCK-8 assays measuring cell viability of NB4 and NOMO-1 cells treated with increasing concentrations of (H, I) genz-644282 or (J, K) ENCORAFENIB for 24 h. Data are presented as mean ± SD ($n = 3$). Statistical significance was determined by one-way ANOVA: ns, not significant; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

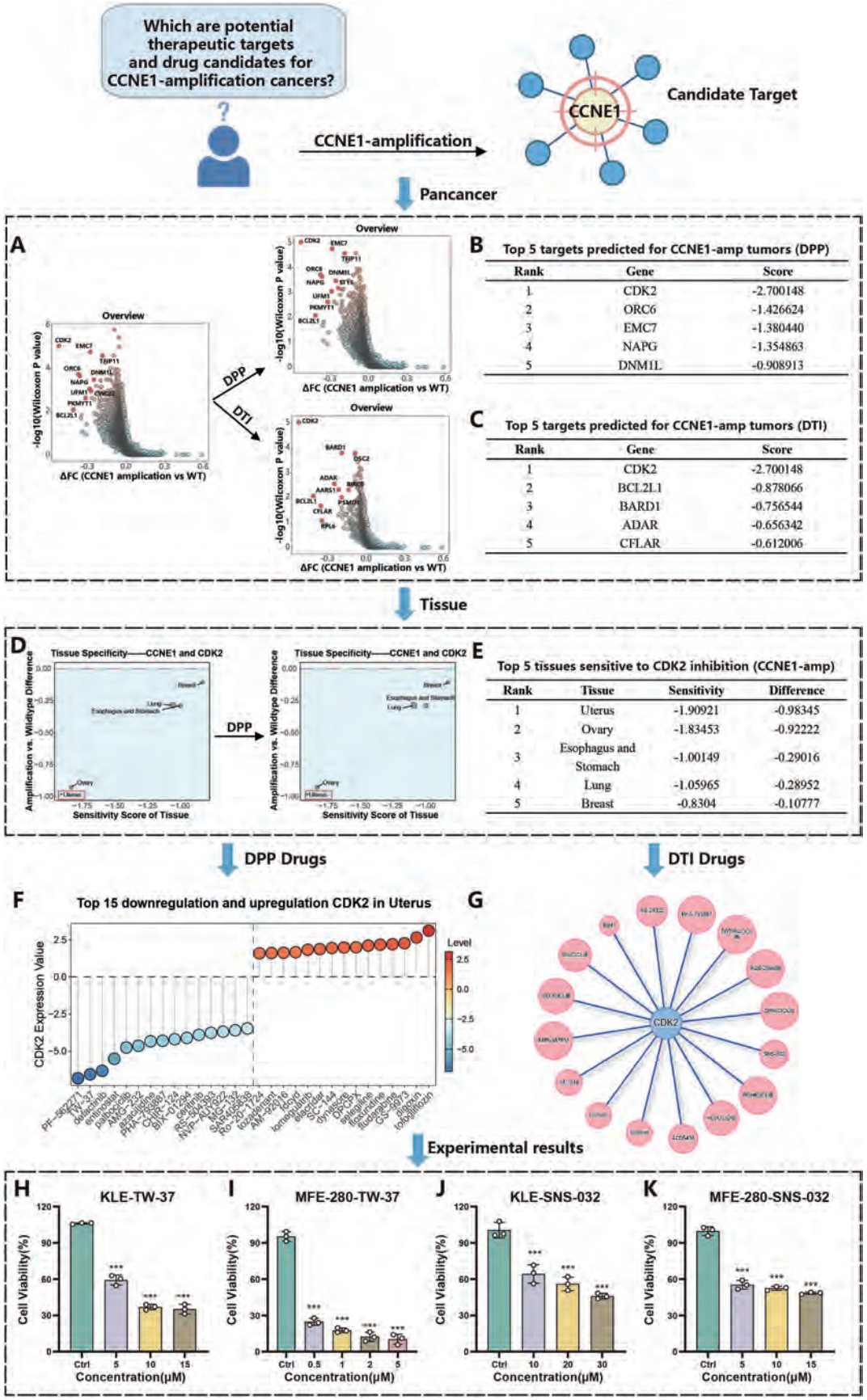

Figure 6 Discovery of therapeutic targets and drug candidates for *CCNE1*-amplified cancers. (A) Potential therapeutic targets identified in ALL, DPP, and DTI. (B, C) Top 5 targets predicted for *CCNE1*-amplified tumors in DPP and DTI. (D) *CDK2*-sensitive tissues stratified by

compounds are provided in the Supporting Information Figs. S7 and S8, confirming the accuracy of our platform's predictions.

Similarly, in the DTI analysis, we identified potential *RAF1*-targeting inhibitors (Supporting Information Table S7). For users interested in FDA-approved compounds, the tool can extract and present this information (Fig. 5G, Supporting Information Table S8). Published studies demonstrate the efficacy of LXH254, LY-3009120, BELVARAFENIB, and SORAFENIB against *KRAS*-mutant tumors[43]. LXH254 inhibits *KRAS*-mutant tumors through dual targeting of *BRAF/CRAF* dimers and *BRAF* monomers[44]. BELVARAFENIB exerts its anti-tumor effects on *KRAS*-mutant cells by blocking the *RAF-MEK-ERK* signaling pathway[44]. SORAFENIB suppresses *KRAS*-mutant tumor progression through *RAF1* inhibition[45]. To confirm prediction results in DTI, we experimentally evaluated all FDA-approved DTI-predicted compounds using CCK-8 assays. Among them, ENCORAFENIB exhibited potent growth-inhibitory effects on NB4 and NOMO-1 cells at 0.5 and 5 μmol/L, respectively (Fig. 5J and K). Similar inhibitory effects were also observed for other predicted drugs (Figs. S7 and S8), highlighting the strong predictive power of our platform.

In summary, the above findings demonstrate that *RAF1* can be regarded as a potential therapeutic target for *KRAS*-mutant cancers, especially for Myeloid cancer.

3.6. Case study 3: Can we find the potential therapeutic targets and drug candidates for CCNE1-amplified cancers?

Amplification of the *CCNE1* locus on chromosome 19q12 is prevalent in multiple tumor types, such as high-grade serous ovarian, uterine, and gastro-oesophageal cancers[25]. Elevated cyclin E correlates with genome instability, whole-genome doubling, and therapy resistance[25]. *CCNE1*-amplified tumors lack effective treatments, creating an urgent need for targeted therapies. While cyclin E is undruggable, the GEMap enables systematic discovery of therapeutic targets for *CCNE1*-amplified tumors, offering a promising strategy for these cancers.

Following the same analytical approach applied to *KRAS*, *CDK2* consistently ranked as a top candidate target—appearing in the top three across all datasets: ALL (alongside *ORC6* and *EMC7*), DPP (also with *ORC6* and *EMC7*), and DTI (with *BCL2L1* and *BARD1*) (Fig. 6A–C). Given its consistent high-ranking performance, *CDK2* was identified as the most promising therapeutic target and selected for further investigation. Accumulated evidence has demonstrated that *CCNE1* amplification results in cyclin E1 overexpression[46]. Cyclin E1 binds to *CDK2*, forming the *CCNE1–CDK2* complex, which enhances *CDK2* kinase activity[47]. Consequently, *CDK2* inhibitors exhibit strong anti-tumor effects in *CCNE1*-amplified tumors, suggesting that *CDK2* is a promising therapeutic target for this tumor subtype. To further investigate tissue sensitivity to *CDK2* inhibition, our tissue-specific analysis identified Uterus cancer as exhibiting the highest therapeutic vulnerability (Fig. 6D and E), indicating pronounced tissue-selective dependence. Based on these findings, we utilized the Essentiality2Drug module to systematically identify potential *CDK2*-targeting therapeutics. In the DPP analysis, we identified the top 15 candidate compounds for *CDK2* downregulation (Supporting Information Table S9). Screening of FDA-approved drugs likewise yielded 15 high-priority candidates for *CDK2* suppression (Fig. 6F, Supporting Information Table S10). Noticeably, many compounds exhibit anti-tumor activity, such as PF-562271[48], ceritinib[49], NVP-AUY922[50]. To verify the robustness of our computational predictions, we performed CCK-8 assays using these top 15 drugs prioritized by DPP on the two representative *CCNE1*-amplified cell lines in Uterus cancer, KLE and MFE-280 cells. As shown in Fig. S9–S12, nearly all the compounds showed obvious anti-tumor effects, especially for the TW-37, which markedly inhibited the proliferation of KLE and MFE-280 cells at concentrations of 5 and 0.5 μmol/L, respectively (Fig. 6H and I).

Similarly, the DTI analysis identified potential *CDK2*-targeting inhibitors (Fig. 6G, Supporting Information Table S11). Findings from previous research indicate that dinaciclib exerts anti-tumor effects in *CCNE1*-amplified tumors through *CDK2* inhibition. Dinaciclib, a cyclin-dependent kinase (*CDK*) inhibitor, suppresses *CDK2* activity, disrupting the *CCNE1-CDK2* complex and inducing cell cycle arrest, thereby suppressing tumor cell proliferation[51]. Other *CDK* inhibitors—including SELICICLIB, MILCICLIB, ZOTIRACICLIB, ALVOCIDIB, AT-7519, and R547—may also suppress *CCNE1*-amplified tumors. The predictive validity of our DTI results was also further experimentally confirmed through *in vitro* assays and all FDA-approved compounds showed obvious anti-tumor effects. For example, SNS-032 significantly reduced the proliferative activity of KLE and MFE-280 cells at 10 and 5 μmol/L, respectively (Fig. 6J and K). Comparable inhibitory activity was observed for additional predicted drugs (Figs. S10 and S12), reinforcing the reliability of our prediction framework.

Taken together, our findings indicate that *CDK2* represents a promising therapeutic target for *CCNE1*-amplified cancers, particularly for uterine cancer.

4. Discussion

Currently, human CRISPR screen data is increasing very rapidly because of its importance in disease prevention, treatment, and drug discovery. To enhance the accessibility of this drug discovery resource for the scientific community, we developed GEMap, an intuitive web-based browser designed to empower researchers. GEMap integrates as many as 1912 screens across 1135 cell lines in 27 tissue types by manual curation on massive CRISPR screen-related resource. Meanwhile, more than 20,000 drug entries have been integrated, enabling users to easily identify drug candidates targeting genes of interest. To our knowledge, this is the first platform to integrate CRISPR screening data, drug-induced transcriptomic changes, and drug target information. It provides a systematic framework for identifying novel drug candidates. As demonstrated in the case study, this tool allows users to swiftly assess the essentiality and selectivity of specific genes across diverse cell lines, while also identifying co-functional genes that exhibit similar essentiality profiles. GEMap streamlines the exploration of gene function, facilitating the discovery of potential

CCNE1-amplified cancers. (E) Top 5 tissues sensitive to *CDK2* inhibition. (F) The top 15 compounds that upregulate and downregulate *CDK2* in Uterus from DPP analysis. (G) Promising drug candidates from DTI analysis. (H–K) CCK-8 assay measuring cell viability of KLE and MFE-280 cells treated with increasing concentrations of (H, I) TW-37 or (J, K) SNS-032 for 48 h. Data are presented as mean ± SD ($n = 3$). Statistical significance was determined by one-way ANOVA; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

therapeutic targets and advancing research in functional genomics. To further assess how the large-scale integration enhances GEMap's robustness and analytical accuracy, we compared gene dependency estimates across multiple CRISPR datasets. These cross-dataset concordance results demonstrate that large-scale data integration effectively enhances the reliability of gene essentiality maps. Moreover, the incorporation of extensive drug perturbation and target data (>20,000 compounds) facilitates the discovery of novel and reliable drug−gene associations. Collectively, these integrations improve both the robustness and precision of the GEMap analytical framework.

In contrast to existing public databases that curate CRISPR screen-related data, GEMap not only integrates the largest number of heterogeneous datasets (including CRISPR screens, PubMed-derived studies, and drug perturbation data), but also uniquely supports modules such as tissue-specific analysis, molecular stratification analysis, and drug candidate identification that are absent in most other tools. As shown in Table 1, we quantified feature coverage: GEMap supports ∼86% of all listed features, while DepMap covers ∼36%, GenomeCRISPR ∼10%, PICKLES ∼32%, iCSDB ∼23%, shinyDepMap ∼18%, and OpenTargets ∼14%. This quantitative analysis highlights that GEMap not only provides broader data integration (CRISPR, multi-omics, drug, PubMed-derived datasets) but also integrates unique modules absent in other platforms, such as therapeutic target prioritization, and drug candidate discovery.

More importantly, the GEMap introduces two specialized modules—Essentiality2Target and Essentiality2Drug—leveraging multimodal data to overcome fundamental limitations in precision oncology. These modules not only enable the rapid identification of therapeutic targets and the customization of treatments for cancer patients categorized by their molecular profiles but also facilitate the discovery of therapeutic drugs linked to candidate treatment targets. One notable application of this tool for drug discovery is 'target hopping', a strategic approach that involves shifting focus between drug targets while maintaining selectivity[52]. This method aims to identify druggable alternatives for traditionally undruggable genes by leveraging their similar dependency profiles. A well-known example is *KRAS*, a key oncogene involved in many cancers and historically deemed undruggable. Using the Essentiality2Target module, we identified *RAF1* and *CFLAR* as top candidate targets. These genes represent promising therapeutic targets for *KRAS*-driven cancers. Together, these features enhance the robustness of gene essentiality maps, improve the accuracy of therapeutic target prioritization, and enable the discovery of novel or more reliable drug−gene relationships.

Nonetheless, our work has limitations: Firstly, GEMap data characterizes the genetic requirements for cells grown in culture, which differs from the *in vivo* tumor environment in critical ways. Secondly, GEMap includes only cultured cancer lines and no wild-type cell lines or tissues; therefore, a highly selective gene in our analysis is not guaranteed to be less toxic to normal tissues when inhibited. Thirdly, it cannot automatically update CRISPR-related data, we will employ natural language processing technology to improve manual update efficiency and use high-performance computing to reduce the processing time for similarity computing in the future. While GEMap is designed to serve the cancer research community, we welcome user feedback and are committed to its continuous improvement. We plan to update GEMap annually to incorporate the latest DepMap, DGIdb, and related datasets, so that users can benefit from the most up-to-date information. New datasets, analytical methods, and visualization features will be incorporated into GEMap as they become available in the future.

In conclusion, the GEMap's mission is to provide researchers with an intuitive and user-friendly platform to uncover gene vulnerabilities and identify potential therapeutic targets. All results are presented through dynamic visualizations and downloadable tables, ensuring reproducibility and further validation in downstream research. We believe that the GEMap would be a useful resource for research communities in drug discovery as well as in basic molecular and cellular biology.

5. Conclusions

Despite advances in precision oncology, oncogenic drivers like *KRAS* mutations and *CCNE1* amplifications remain clinically intractable. In this study, we developed a computational pharmacology platform, GEMap (https://web.biotcm.net/GEMap), that directly addresses the therapeutic intractability of oncogenic drivers (*e.g.*, KRAS mutations, *CCNE1* amplifications) by integrating genome-wide CRISPR essentiality profiles with drug perturbation signatures and protein interaction networks. This unified platform enables multi-modal interrogation of gene dependencies, expression dynamics, copy number variations, and mutational landscapes across diverse cancer models. Through quantitative scoring of genetic vulnerabilities—exemplified by identifying *RAF1* as a druggable surrogate for *KRAS*-mutant tumors—GEMap empowers context-specific target prioritization for molecularly stratified cancers. Furthermore, its systematic matching of drug candidates to targets accelerates bench-to-bedside translation. To our knowledge, GEMap represents the first integrated platform combining genome-wide CRISPR screening data with drug perturbation transcriptomic profiles and target interaction networks, establishing a novel systematic framework for therapeutic candidate discovery. Therefore, this work not only showcases the feasibility and potential of GEMap as a valuable tool in drug discovery but also highlights its versatile and widely applicable framework within the realm of drug development.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (82574713, 82430119, 82273810), National Key Research and Development Program of China (2022YFC3502000), National Science and Technology Major Project of the Ministry of Science and Technology of China (2025ZD1800300), the Computational Biology Program (25JS2830400) of Science and Technology Commission of Shanghai Municipality (STCSM), Shanghai Municipal Science and Technology Major Project (ZD2021CY001), the Chenguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (23CGA45, Saisai Tian), Natural Science Foundation of Shanghai (25ZR1402574), the ability establishment of sustainable use for valuable Chinese medicine resources (2060302), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2023-I2M-3-009).

Author contributions

Weidong Zhang and Saisai Tian: conceptualization, original draft, methodology, review and editing, funding acquisition and supervision. Saisai Tian, Yixue Liang, and Jinbo Zhang analyzed the

data, designed the website, generated the figures and wrote the paper. Weishan Liang, Jinyuan Lu, Jiajia Ren, Bojian Lin, Ruijun Wu, and Chengyang Guo participated in part of the experiments. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that they have no competing interests.

Data availability

Data supporting the conclusions of this study are available upon request from the corresponding author. In addition, we have made all resources publicly available: the complete core code ensuring reproducibility is accessible on GitHub (https://github.com/YixueLiang/GEMap_ms), and the website implementation scripts are hosted on Gitee (https://gitee.com/auptz/gemap).

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2026.02.004.

References

1. Puccetti M, Pariano M, Schoubben A, Giovagnoli S, Ricci M. Biologics, theranostics, and personalized medicine in drug delivery systems. *Pharmacol Res* 2024;**201**:107086.
2. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized health care. *Clin Transl Sci* 2021;**14**:86−93.
3. Incorvati JA, Shah S, Mu Y, Lu J. Targeted therapy for HER2 positive breast cancer. *J Hematol Oncol* 2013;**6**:38.
4. Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au HJ, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med* 2007;**357**:2040−8.
5. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* 2011;**364**:2507−16.
6. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov* 2013;**12**:569.
7. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016;**47**:20−33.
8. Gashaw I, Ellinghaus P, Sommer A, Asadullah K. What makes a good drug target?. *Drug Discov Today* 2011;**16**:1037−43.
9. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 2019;**568**:511−6.
10. Chan EM, Shibue T, McFarland JM, Gaeta B, Ghandi M, Dumont N, et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* 2019;**568**:551−6.
11. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem* 2016;**85**:227−64.
12. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell* 2017;**170**:564−76.
13. Sandoval GJ, Pulice JL, Pakula H, Schenone M, Takeda DY, Pop M, et al. Binding of TMPRSS2-ERG to BAF chromatin remodeling complexes mediates prostate oncogenesis. *Mol Cell* 2018;**71**:554−66.
14. Wang X, Wang S, Troisi EC, Howard TP, Haswell JR, Wolf BK, et al. BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat Commun* 2019;**10**:1881.
15. Rauscher B, Heigwer F, Breinig M, Winter J, Boutros M. GenomeCRISPR — a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Res* 2017;**45**:D679−86.
16. Novak LC, Chou J, Colic M, Bristow CA, Hart T. PICKLES v3: the updated database of pooled *in vitro* CRISPR knockout library essentiality screens. *Nucleic Acids Res* 2023;**51**:D1117−21.
17. Choi A, Jang I, Han H, Kim MS, Choi J, Lee J, et al. iCSDB: an integrated database of CRISPR screens. *Nucleic Acids Res* 2021;**49**:D956−61.
18. Shimada K, Bachman JA, Muhlich JL, Mitchison TJ. Shinydepmap, a tool to identify targetable cancer genes and their functional connections from Cancer Dependency Map data. *eLife* 2021;**10**:e57116.
19. Buniello A, Suveges D, Cruz-Castillo C, Llinares MB, Cornu H, Lopez I, et al. Open targets platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res* 2025;**53**:D1467−75.
20. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437−52.
21. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res* 2023;**52**:D1227−35.
22. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. *Cell* 2017;**170**:564−76.
23. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald 3rd ER, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature* 2019;**569**:503−8.
24. Jaiswal A, Gautam P, Pietilä EA, Timonen S, Nordström N, Akimov Y, et al. Multi-modal meta-analysis of cancer cell line omics profiles identifies ECHDC1 as a novel breast tumor suppressor. *Mol Syst Biol* 2021;**17**:e9526.
25. Gallo D, Young JTF, Fourtounis J, Martino G, Álvarez-Quilón A, Bernier C, et al. CCNE1 amplification is synthetic lethal with PKMYT1 kinase inhibition. *Nature* 2022;**604**:749−56.
26. Inoue T, Ichikawa D, Ueno T, Cheong M, Inoue T, Whetstone WD, et al. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. *Neurotrauma Rep* 2020;**1**:8−16.
27. Huang Y, Huang W, Ma X, Zhao G, Kang J, Li H, et al. Nomogram for predicting mild cognitive impairment in Chinese elder CSVD patients based on Boruta algorithm. *Front Aging Neurosci* 2025;**17**:1431421.
28. Mei Z, Shi Z. On LASSO for high dimensional predictive regression. *J Econom* 2024;**242**:105809.
29. Guido R, Ferrisi S, Lofaro D, Conforti D. An overview on the advancements of support vector machine models in healthcare applications: a review. *Information* 2024;**15**:235.
30. Zhang J, Mucs D, Norinder U, Svensson F. LightGBM: an effective and scalable algorithm for prediction of chemical Toxicity-Application to the Tox21 and mutagenicity data sets. *J Chem Inf Model* 2019;**59**:4150−8.
31. Shimada K, Bachman JA, Muhlich JL, Mitchison TJ. shinyDepMap, a tool to identify targetable cancer genes and their functional connections from cancer dependency map data. *eLife* 2021;**10**:e57116.
32. Ni S, Kong X, Zhang Y, Chen Z, Wang Z, Fu Z, et al. Identifying compound-protein interactions with knowledge graph embedding of perturbation transcriptomics. *Cell Genom* 2024;**4**:100655.
33. Hakobyan A, Meyenberg M, Vardazaryan N, Hancock J, Vulliard L, Loizou JI, et al. Pan-cancer analysis of the interplay between mutational signatures and cellular signaling. *iScience* 2024;**27**:109873.
34. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet* 2019;**20**:175.
35. Palmer CL, Boras B, Pascual B, Li N, Li D, Garza S, et al. CDK4 selective inhibition improves preclinical anti-tumor efficacy and safety. *Cancer Cell* 2025;**43**:464−81.
36. Sulahian R, Kwon JJ, Walsh KH, Pailler E, Bosse TL, Thaker M, et al. Synthetic lethal interaction of SHOC2 depletion with MEK inhibition in RAS-driven cancers. *Cell Rep* 2019;**29**:118−34.

37. Moser R, Annis J, Nikolova O, Whatcott C, Gurley K, Mendez E, et al. Pharmacologic targeting of TFIIH suppresses KRAS-mutant pancreatic ductal adenocarcinoma and synergizes with TRAIL. *Cancer Res* 2022;**82**:3375−93.
38. García-Alonso S, Mesa P, Ovejero LP, Aizpurua G, Lechuga CG, Zarzuela E, et al. Structure of the RAF1-HSP90-CDC37 complex reveals the basis of RAF1 regulation. *Mol Cell* 2022;**82**:3438−52.
39. Costa-Cabral S, Brough R, Konde A, Aarts M, Campbell J, Marinari E, et al. CDK1 is a synthetic lethal target for KRAS mutant tumours. *PLoS One* 2016;**11**:e0149099.
40. Peng DH, Kundu ST, Fradette JJ, Diao L, Tong P, Byers LA, et al. ZEB1 suppression sensitizes KRAS mutant cancers to MEK inhibition by an IL17RD-dependent mechanism. *Sci Transl Med* 2019;**11**: eaaq1238.
41. Justilien V, Ali SA, Jamieson L, Yin N, Cox AD, Der CJ, et al. Ect2-dependent rRNA synthesis is required for KRAS-TRP53-driven lung adenocarcinoma. *Cancer Cell* 2017;**31**:256−69.
42. Kim HR, Park SJ, Cho YS, Moyo MK, Choi JU, Lee NK, et al. Stimulating macropinocytosis of peptide-drug conjugates through DNA-dependent protein kinase inhibition for treating KRAS-mutant cancer. *J Control Release* 2024;**372**:176−93.
43. Ramurthy S, Taft BR, Aversa RJ, Barsanti PA, Burger MT, Lou Y, et al. Design and discovery of *N*-(3-(2-(2-hydroxyethoxy)-6-morpholinopyridin-4-yl)-4-methylphenyl)-2-(trifluoromethyl)isonicotinamide, a selective, efficacious, and well-tolerated RAF inhibitor targeting RAS mutant cancers: the path to the clinic. *J Med Chem* 2020;**63**:2013−27.
44. Moore AR, Rosenberg SC, McCormick F, Malek S. RAS-targeted therapies: is the undruggable drugged?. *Nat Rev Drug Discov* 2020;**19**: 533−52.
45. Saturno G, Lopes F, Niculescu-Duvaz I, Niculescu-Duvaz D, Zambon A, Davies L, et al. The paradox-breaking panRAF plus SRC family kinase inhibitor, CCT3833, is effective in mutant KRAS-driven cancers. *Ann Oncol* 2021;**32**:269−78.
46. Kwiatkowski N, Liang T, Sha Z, Collier PN, Yang A, Sathappa M, et al. CDK2 heterobifunctional degraders co-degrade CDK2 and cyclin E resulting in efficacy in CCNE1-amplified and overexpressed cancers. *Cell Chem Biol* 2025;**32**:556−69.
47. Dietrich C, Trub A, Ahn A, Taylor M, Ambani K, Chan KT, et al. INX-315, a selective CDK2 inhibitor, induces cell cycle arrest and senescence in solid tumors. *Cancer Discov* 2024;**14**: 446−67.
48. Lu D, Wang Y, Zhang T, Wang F, Li K, Zhou S, et al. Metabolic radiolabeling and *in vivo* PET imaging of cytotoxic T lymphocytes to guide combination adoptive cell transfer cancer therapy. *J Nanobiotechnol* 2021;**19**:175.
49. Shaw AT, Kim DW, Mehra R, Tan DS, Felip E, Chow LQ, et al. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med* 2014;**370**:1189−97.
50. Wang Y, Chu T. Progress in detection of EGFR exon 20 insertion mutation and targeted therapies in non-small cell lung cancer. *Chin J Lung Cancer* 2023;**26**:135−47.
51. Etemadmoghadam D, Au-Yeung G, Wall M, Mitchell C, Kansara M, Loehrer E, et al. Resistance to CDK2 inhibitors is associated with selection of polyploid cells in CCNE1-amplified ovarian cancer. *Clin Cancer Res* 2013;**19**:5960−71.
52. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 2013;**9**:232−40.