

SOFTWARE ARTICLE

Open Access



# EZCancerTarget: an open-access drug repurposing and data-collection tool to enhance target validation and optimize international research efforts against highly progressive cancers

David Dora<sup>1\*</sup> , Timea Dora<sup>2</sup>, Gabor Szegvari<sup>3</sup>, Csongor Gerdán<sup>4,5</sup> and Zoltan Lohinai<sup>3,4\*</sup>

\*Correspondence:  
dora.david@med.semmelweis-univ.hu; zoltan.lohinai@koranyi.hu

<sup>1</sup> Department of Anatomy, Histology, and Embryology, Semmelweis University, Tuzolto st. 58, Budapest 1094, Hungary

<sup>2</sup> Department of Management and Business Economics, Budapest University of Technology and Economics, Budapest, Hungary

<sup>3</sup> Translational Medicine Institute, Semmelweis University, Budapest, Hungary

<sup>4</sup> National Korányi Institute of Pulmonology, Piheno ut 1., 1121 Budapest, Hungary

<sup>5</sup> Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

## Abstract

The expanding body of potential therapeutic targets requires easily accessible, structured, and transparent real-time interpretation of molecular data. Open-access genomic, proteomic and drug-repurposing databases transformed the landscape of cancer research, but most of them are difficult and time-consuming for casual users. Furthermore, to conduct systematic searches and data retrieval on multiple targets, researchers need the help of an expert bioinformatician, who is not always readily available for smaller research teams. We invite research teams to join and aim to enhance the cooperative work of more experienced groups to harmonize international efforts to overcome devastating malignancies. Here, we integrate available fundamental data and present a novel, open access, data-aggregating, drug repurposing platform, deriving our searches from the entries of Clue.io. We show how we integrated our previous expertise in small-cell lung cancer (SCLC) to initiate a new platform to overcome highly progressive cancers such as triple-negative breast and pancreatic cancer with data-aggregating approaches. Through the front end, the current content of the platform can be further expanded or replaced and users can create their drug-target list to select the clinically most relevant targets for further functional validation assays or drug trials. EZCancerTarget integrates searches from publicly available databases, such as PubChem, DrugBank, PubMed, and EMA, citing up-to-date and relevant literature of every target. Moreover, information on compounds is complemented with biological background information on eligible targets using entities like UniProt, String, and GeneCards, presenting relevant pathways, molecular- and biological function and subcellular localizations of these molecules. Cancer drug discovery requires a convergence of complex, often disparate fields. We present a simple, transparent, and user-friendly drug repurposing software to facilitate the efforts of research groups in the field of cancer research.

**Keywords:** Drug-repurposing, Cancer research, Drug database, Data mining, Lung cancer



## Introduction

There is an increasing need for open-access drug repurposing databases for researchers in the translational and clinical field due to the emergence of many potential new therapeutic targets every year. This trend will likely continue in the future with exponentially increasing repositories full of data, and data mining requires particular expertise and qualifications that cannot be expected from a single group of researchers. The development of novel pharmaceuticals takes enormous effort and immense financial and human resources, a decade of research and clinical trials until approval with more advancements in prevalent cancers. Research groups usually select their target and related research directions without cross-optimization of efforts across individual groups. Industry AI-based drug developments have limitations and also require preclinical validation. Therefore, *in silico* approaches with solid human resources might assist than fully overtake all steps.

The therapeutic potential of a molecular target, the suitable inhibitor (or agonist), might already be available in another indication, or a small molecule lead compound is available that needs further modification and preclinical testing. Nowadays, a wide selection of databases is available for various goals in drug repurposing [1–3]. Omics data about molecular targets can be retrieved from *Uniprot* [4], *Genecards* [5], *TTD* [6], *STITCH* [7], *BioGRID* [8], and *STRING* [9], or corresponding pathways from *KEGG* [10], *Pathways Common* [11], or *Reactome* [12] with to a certain level of deficient or overlapping information. Information about drugs launched or in development (drug omics data) can be obtained from *Pubchem* [13], *Drug Bank* [14], *Drug Map Central* [15], or *FDA or EMA Label* repositories. Also, the *Clinicaltrials.gov*, *SIDER* [16], or *FDA Adverse Event Reporting System (FAERS)* platforms provide essential information on stages of drug testing.

Platforms for browsing and visualizing drug-target interactions and drugs in disease-context are already available for many users, such as *Drug Target Profiler* [17], *Cancer Genome Interpreter* [18], *SwissTargetPrediction* [19], *OpenTargets* [20], and *PharmGKB* [21]. These platforms include many entries and a gargantuan amount of unstructured information that is sometimes time-consuming and difficult to handle, especially for those lacking expertise in the field. The data mining techniques require specific processes usually nonexistent in less prevalent cancers and often lack financial incentive for clinical testing [22]. Cancer drug discovery requires a convergence of complex, often disparate fields. There is a great need for simple, transparent and user-friendly drug repurposing databases.

One recent endeavor with high impact is Broad Institute's *Clue.io*, which serves as a drug repurposing hub curated and annotated collection of FDA-approved drugs, clinical trial drugs, and preclinical tool compounds with a companion information resource [23]. In this paper, we present a novel, open access, data-mining, drug repurposing platform, deriving our searches from the entries of *Clue.io*. *EZCancerTarget* provides researchers and clinicians (especially in cancer research) an easy-to-use platform to retrieve all the necessary information for a freely selected array of potential therapeutic targets with a parallelly-working, data-mining application. In addition, *EZCancerTarget* provides detailed biological information on the selected set of target molecules using open-access databases, such as *Uniprot*, *GeneCards*, *Gene Ontology* and *STRING*. This way, the user

receives a concise summary on the biological relevance of every target, that is explicitly important for researchers who are not experts in molecular biology.

## Methods

### Clue.io and dataPatch R script

Target inclusion/exclusion depends on search results from a clue.io query. *EZCancer-Target* consists of 3 separate R scripts. The first script -clue.R- calls various clue.io REST API endpoints to build up a result table. If the main API call does not find any component for a target, that target will not be involved in further processing steps since no known drug repurposing approaches are available in Clue.io. R looks up the input target list in two ways. First, it tries to access a private and/or shared Google Sheet file. It requires a unique "key" (a token) given to clue.R via a simple environmental variable. If this secret key is available for the script, it authenticates by *gargle* package [24] to access Google Sheet API services. Next, it reads the sheet and takes the values from its first three columns. An ID string also identifies the Google Sheet, and it is passed via an operating system environment. If there is no API key/Google Sheet identifier, then clue.R tries to load a TSV file from the INPUT directory of the EasyCancerTarget directory. Clue.R merges the outputs of various clue.io API calls and saves the composing table into an RDS file (R-specific data format to store and load R objects). At the next stage, data patch.R reads this RDS file and restores the data frame from it.

### PubMed function

This function searches for the compound name received from clue.io and sends a search request to PubMed<sup>®</sup> service of *National Center for Biotechnology Information* (NCBI). It restricts the result set by including only clinical trials, meta-analyses, randomized controlled trials, reviews and systematic reviews. The results are ordered by the best match algorithm of PubMed<sup>®</sup>. Our function picks the top 3 of the result set and stores it in the global datatable. If there is no hit at all, dataPatch.R provides the used search URL and this search can be re-initiated and/or refined by users of *EZCancerTarget*. pubMed function uses a simple XPath query to extract PubMed identifiers embedded into the resulting HTML source code.

### XmlUniProt function

This function collects data from the *UniProt* website. *UniProt* provides APIs to access and query its data. Easy to access the human readable contents in machine readable formats (for example XML, RDE, etc.). The usage of the *UniProt website REST API* is straightforward, since the input target list also contains UniProt identifiers. xmlUniProt extracts GO (*Gene Ontology*) molecular function and cellular component terms, *STRING* and *Reactome* references from the received XML data. These specific entries are stored in simple R lists and added to the already collected data in a new column: "UniProtData".

### Presentation layer

A web browser is a "mandatory" software of each end user's computer, so HTML is a clear choice to summarize, visualize and deliver collections of texts, images and

hyperlinks. An important part of this rendering is building an HTML source file and populating it with the collected data in an user-friendly way. *EZCancerTarget* follows the popular *Model-View-Controller* design pattern even though it composes only static HTML output (View) from the data (Model) at this stage of the workflow. (*NOTE: However previous actions and functionalities of the workflow can be interpreted as the Controller part of the MVC pattern*).

Further detail on methods and functions is available at <https://cycle20.github.io/EZCancerTarget/methods.html>. All the R scripts and versions' descriptions and runtime environment are freely accessible at: <https://github.com/cycle20/EZCancerTarget>.

## Results

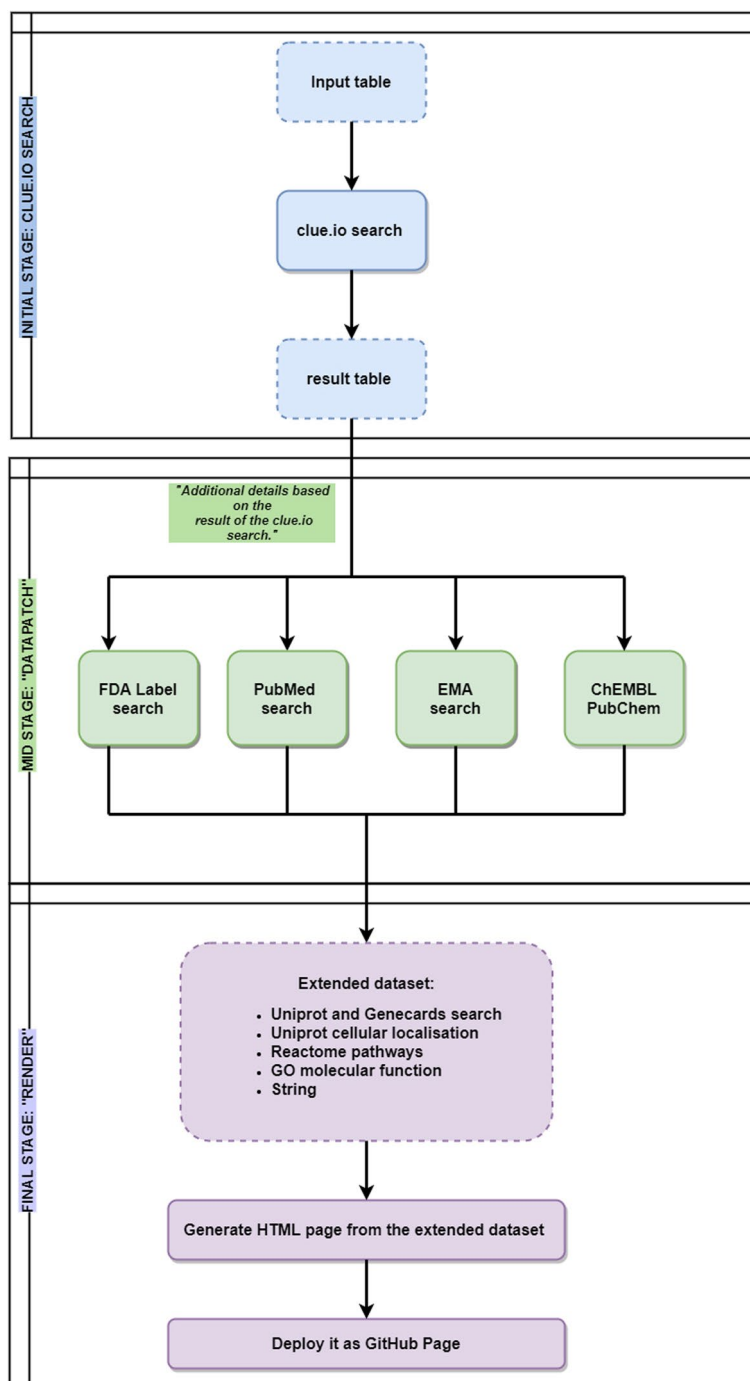
### Construction and content

Each entry obtained from the search results in the interactive online platform of *EZCancerTarget* is referenced and has at least one scientific piece of evidence. Figure 1 shows a flowchart on the processing steps and workflow of *EZCancerTarget*. First, users of *EZCancerTarget* can start their workflow by opening the project's starting page on Github (<https://cycle20.github.io/EZCancerTarget/>). Then, users can upload their target list with three pieces of information into a google spreadsheet directly (Target INPUT), which requires a quick authorization step from the project administrator (to avoid interference from simultaneous and multiple users), or can manually install it in R environment. Installation process is described in detail on the aforementioned Github page.

*EZCancerTarget* fetches its input from a simple data source. It can be a TSV file in case of running scripts on user-controlled computer. Another option is updating a shared Google Spreadsheet that is processed by workflow scripts on *GitHub*. *EZCancerTarget* load data from *clue.io*, then download data from other sources: *FDA Label* service, *PubMed*, *EMA*, *ChEMBL*, *PubChem* (Fig. 1). After data collection it generates a user-friendly report file and a summary output data file. The report HTML file can be opened by a browser. If the workflow runs on *GitHub*, it will deploy the report file as a public *GitHub* web page.

In the input table, the first piece (Fig. 2 Column A), asks for the HUGO ID. In column B, users can give a "Label" for every target for classification and clustering, useful in later work. The third piece (Fig. 2 column C) is the UniprotKB ID of the searched gene. Inputs for the HUGO and Label columns are limited to 12 characters. On the right side of the spreadsheet (columns E-K), hyperlinks provide access to the results page on Github, and in the "Results of Update Request" box, users can check the query's status. Hitting the "Start Rendering" button located on columns E–F starts the query. The area within E1-K8 are protected and automatically overwritten if edited (Fig. 2).

By clicking on the "Result page" link on the target spreadsheet, we can access the results of our query within approximately 30 min. Clicking on the hyperlink in cell H6 we can follow the progress of the query (Fig. 2, arrow). A new query overwrites the earlier one in the web application, but every previous version is saved on Github under the "Result of Update Request" link (<https://github.com/cycle20/EZCancerTarget/actions/workflows/clue.yml>). A scrollable panel displays all the targets on the left with at least one valid drug compound available. The software automatically excludes



**Fig. 1** Flowchart of functionality. Flowchart describes the main steps of *EZCancerTarget's* functionality, including data input, Clue.io target search, cross-referencing in databases (Datapatch) and molecular background information on selected targets (Render)

entries where no drug or small molecule inhibitor/agonist is available according to the Clue.io repurposing hub.

In the first entry of the results list ("summary") the evaluation report on the search is accessible. In the "overview" section, it displays the total number of found compounds

	A	B	C	D	E	F	G	H	I	J	K
1	HUGO	Label	UNIPROT_KB_ID		Baseline page:	<a href="https://cyclac20.github.io/scancer/baseline/index.target.with.data.html">https://cyclac20.github.io/scancer/baseline/index.target.with.data.html</a>					
2	ITGB6	SCLC NE-LOW	P18564		Result page:	<a href="https://cyclac20.github.io/scancer/index.target.with.data.html">https://cyclac20.github.io/scancer/index.target.with.data.html</a>					
3	RAB27B	SCLC NE-LOW	Q00194		<div style="border: 1px solid black; padding: 5px;"> <p>Start Rendering ...</p> </div> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Program messages</p> <p>Result of Update Request</p> <p>Workflow started. Please follow its progress on GitHub</p> <p><a href="https://github.com/cyclac20/scancer/actions/workflows/clue.yml">https://github.com/cyclac20/scancer/actions/workflows/clue.yml</a></p> </div>						
4	KRT5	SCLC NE-LOW	P13647								
5	YBX3	SCLC NE-LOW	P16989								
6	ANXA1	SCLC NE-LOW	P04083								
7	PGC	SCLC NE-LOW	P20142								
8	KRT17	SCLC NE-LOW	Q04695								
9	MYC	SCLC NE-LOW	P01106								
10	TP63	SCLC NE-LOW	Q9H3D4								
11	LRMP	SCLC NE-LOW	Q12912								
12	MIMP7	SCLC NE-LOW	P09237								
13	ITGB4	SCLC NE-LOW	P16144								
14	CD70	SCLC NE-LOW	P32970								
15	PRAME	SCLC NE-LOW	P78395								
16	FOXJ1	SCLC NE-LOW	Q12951								
17	VSNL1	SCLC NE-LOW	P62760								
18	CES1	SCLC NE-LOW	P23141								
19	GPR126	SCLC NE-LOW	Q86S04								
20	MAGEA4	SCLC NE-LOW	P43358								
21	CD44	SCLC NE-LOW	P16070								
22	BCAT1	SCLC NE-LOW	P54687								
23	ROS1	SCLC NE-LOW	P08922								
24	CXCR2	SCLC NE-LOW	P25025								
25	SPP1	SCLC NE-LOW	Q3LGB0								
26	IFI27	SCLC NE-LOW	P40305								
27	CYP3A5	SCLC NE-LOW	P20815								
28	ANXA3	SCLC NE-LOW	P12429								
29	ALDH2	SCLC NE-LOW	P05091								
30	FGFBP1	SCLC NE-LOW	Q14512								
31	ITGAM	SCLC NE-LOW	P11215								
32	BIRC3	SCLC NE-LOW	Q13489								
33	CD27	SCLC NE-LOW	P26842								

**Fig. 2** Input table for molecular targets. Users can enter selected targets' HUGO name (black rectangle), label (blue dashed rectangle) and Uniprot ID (green dashed rectangle) in columns **A**, **B** and **C**. Hitting "Start Rendering" will initiate the Clue.io search (red arrowhead). Progress can be traced by clicking on hyperlink in cells H6-K6 (black arrow). Clicking on the hyperlink in cell F2-I2 reveals the results page

for all listed targets and the average number of compounds per target. The amount of found compounds are classified according to their pre- and clinical phase as well. In the "molecular background" section details from the retrieved molecular background data is evaluated according to the number of found Reactome- or KEGG pathways and subcellular localizations, String interactors and GO molecular functions, biological processes and. Finally, every listed target is separately detailed of their compound entries in PubMed, PubChem, ChEMBL and DrugBank.

The platform creates a table for every target, where different columns indicate the mechanism of action (MoA), clinical status (preclinical, phase 1, phase 2, phase 3, or launched), and the search resources from PubMed EMA and the direct entry from Clue.io. Furthermore, the query table includes hyperlinks with DrugBank, PubChem, and ChEMBL IDs to quickly access the compounds' chemical and pharmacological properties (Fig. 3).

*EZCancerTarget* also gives a comprehensive, highly structured overview of the selected targets regarding their molecular biology data, including molecular function (Gene Ontology), their connectome (STRING), participation in pathways, and cellular localization (Reactome and KEGG) retrieved from various databases. Hyperlinks to GeneCards and DrugBank Target Search are also available but differently structured as for UniProt entries. The "STRING" entry opens a static string map for the target and provides a hyperlink to string-db.org (Fig. 4). The following entry carries the "Molecular Functions / Subcellular Localisations" title, where the two main hyperlinks (source) lead to UniProt's "Function" and "Subcellular Localization" pages. Molecular function entries and target localizations are also provided as text separately, where hyperlinks lead to the QuickGo platform to obtain further information about relevant compartment-specific molecular pathways (Fig. 4). The last entry named "Pathways" provides links to every Reactome database, where the

**Drug Targets - Has CLUE.IO Entries** Created at 2021-06-09 17:23:34 UTC

GeneCards UniProtKB DrugBank Target Search

BIRC3 :: Compounds

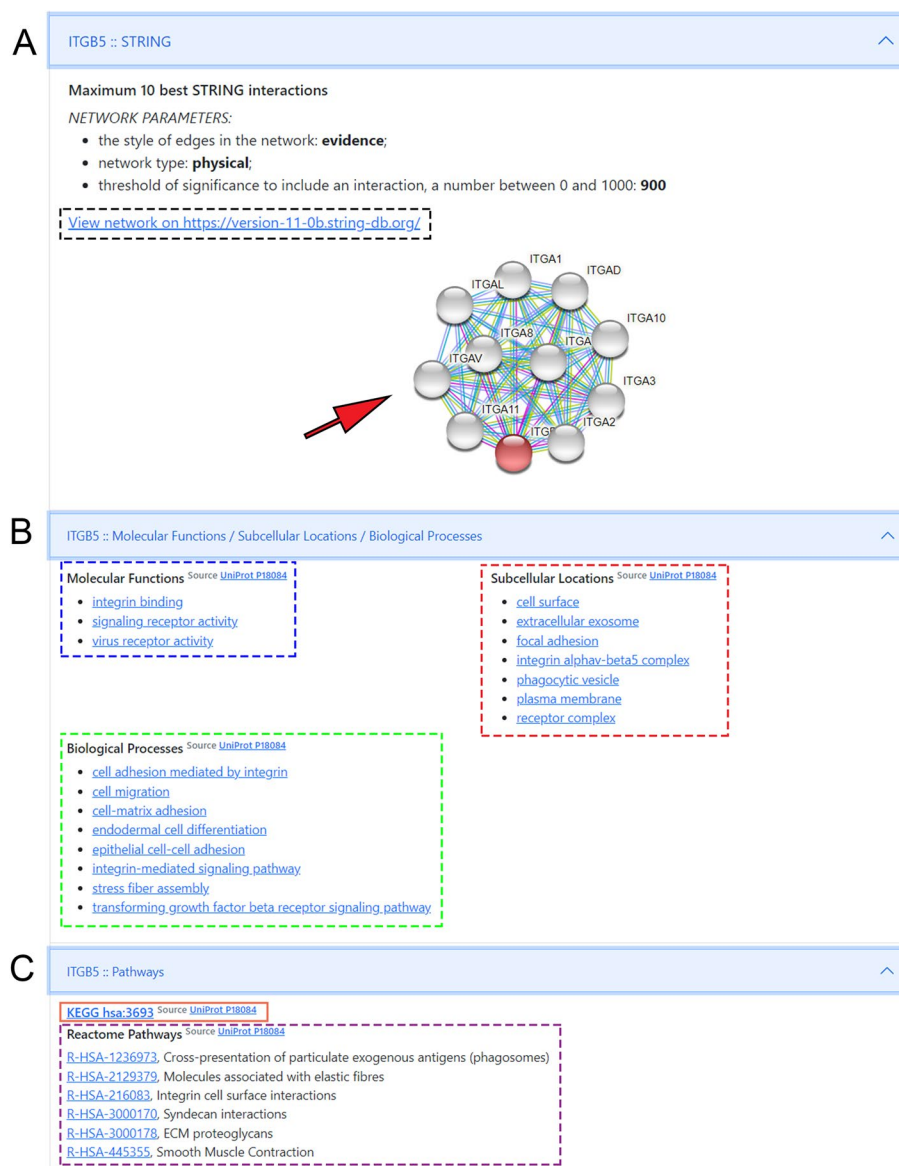
Compound	MoA <sup>1</sup>	Clinical Status	Resources <sup>1</sup>	DrugBank/PubChem/ ChEMBL
AT-406	IAP antagonist	Phase 1	<b>From PubMed:</b> <a href="#">PubMed 33460574</a> <a href="#">PubMed 25716544</a> <a href="#">PubMed 27915971</a> <b>From EMA:</b> <a href="#">Found PDF</a> <a href="#">CLUE.IO</a> <a href="#">ClinicalTrials</a>	<a href="#">ChEMBL2158051</a> <a href="#">PubChem: 25022340</a>
AZD5582	XIAP inhibitor	Preclinical	<b>From PubMed:</b> <a href="#">PubMed 33007800</a> <b>From EMA:</b> <a href="#">Found PDF</a> <a href="#">CLUE.IO</a>	<a href="#">ChEMBL3108827</a> <a href="#">PubChem: 49847690</a>
birinapant	XIAP inhibitor	Phase 2	<b>From PubMed:</b> <a href="#">PubMed 31766284</a> <a href="#">PubMed 26333881</a> <a href="#">PubMed 30631429</a> <a href="#">CLUE.IO</a> <a href="#">ClinicalTrials</a>	<a href="#">ChEMBL3039522</a> <a href="#">PubChem: 49836020</a> <a href="#">DB11782</a>
GDC-0152	XIAP inhibitor	Phase 1	<b>From PubMed:</b> <a href="#">PubMed 22623056</a> <b>From EMA:</b> <a href="#">Found PDF</a> <a href="#">CLUE.IO</a> <a href="#">ClinicalTrials</a>	<a href="#">ChEMBL2063869</a> <a href="#">PubChem: 46940575</a> <a href="#">DB12380</a>

**Fig. 3** List of drug targets. Clicking on the labels of selected targets (column on left side) unveils available compound list (black box) describing also mechanism of action (MoA, dashed box), clinical status (red box), resources of information on PubMed (green box) and DrugBank/PubChem/ChEMBL entries (blue box)

target's participation is visualized in every relevant metabolic pathway (Fig. 4). The hyperlink to the entry of the KEGG database is also displayed here, without being broken down to individual links to pathways. Supplementary Video 1 shows a short tutorial about the functionality of the program and the main steps to generate a query.

### Current content

This current content includes 145 molecular that focus on the three most aggressive malignities with limited therapeutical prospect: small-cell lung cancer (SCLC) [25], triple-negative breast cancer (TNBC) [26–28] and pancreatic adenocarcinoma (PADC) [29–31]. The starting content provides a template for researchers and serves as an example to demonstrate the software's functionality. The list can be further expanded or replaced with any UniProt-listed gene and can be customized to the needs of the research group using the platform. Users can also give titles to genes in the "label" column to address targets and classify them into different groups. When a researcher would like to use the platform, they can edit the gene list according to their goals after giving the administrator access to the spreadsheet. The original content is saved in the github main page (<https://cycle20.github.io/EZCancerTarget>), therefore can be retrieved and reset anytime. When the researcher finishes constructing their target list, the administrator refreshes the cache for the optimization of the search and generates a new result page. The results page displays a "summary" section, where the list of target genes with- or without relevant drug repurposing entries (retrieved from clue.io) are available. Molecular background data is only displayed for genes with at least one valid clue.io compound entry.



**Fig. 4** Details on the molecular background of druggable targets. **A** shows the network map from String. db with static string map and hyperlink to String-db entry. **B** displays hyperlinks to “molecular function”, “biological processes” and “subcellular localisation” to browse the UniProt database on molecular background. By clicking directly on the titles, we can access a specific function. **C** shows hyperlinks to visualize “KEGG” and “Reactome” pathways of the selected target. For Reactome, clicking on individual pathway titles we can directly access the infographic of the given pathway

## Discussion

Open access journals and databases are an essential basis for drug target developments. Endeavours, like the TCGA database and Oncomine [32] has contributed vastly to accelerate drug research in oncology the latest decade and concurrently multiple other enterprises emerged to assist researchers with valuable genomic, transcriptomic and proteomic data in the pursue for novel cancer biomarkers [33–36]. However, the information is not well structured for specific diseases, including rare and highly progressive cancers [37, 38]. Moreover, it is challenging and time-consuming to associate the latest



biomarkers with drugs to pick the optimal way and contribute to the field. Only a few research groups with diverse expertise can participate, leaving many researchers without an equal opportunity of involvement. Also, the individual interest of these groups might not represent the optimal way to examine diseases. Therefore, we propose a novel, optimal target selection methodology. Notably, after the success of PD-L1-inhibitors in non-small cell lung cancer (NSCLC), where 5-year-survival in extensive-stage disease increased from 2 to 25%, there has been a keen interest to expand on immunotherapy utilization in small cell lung cancer (SCLC) as well. Two anti-PD-1 immunotherapies, nivolumab and pembrolizumab, have had their FDA approval [39, 40], but they were withdrawn after the confirmatory phase III trials, because statistical significance for overall survival was not significant compared to control groups. Nevertheless, PD-L1 expression in SCLC has never been unequivocally correlated with the response.

Open-access data on the latest research requiring further validation is of high interest to the field. In low-prevalence and highly aggressive cancers, scarcity of available tissue samples limit research, so there is an unmet need to share and optimize resources in the field. It has been decades with only modest therapeutic advancements for highly progressive cancers such as pancreatic cancer, triple-negative breast cancer, glioblastoma multiforme, or SCLC, with an unmet need for advances. To enhance drug target development, we believe that *EZCancerTarget* can serve as an easy-to-use, semi-comprehensive data-mining platform for drug-repurposing and can assist significantly smaller research groups in the fight against malignancies. *EZCancerTarget* only provides a tool for researchers to keep their target list organized and to aid these researchers in selecting the best targets to start validation with. Our software aims to assist the decision-making process of these projects druggable and clinically relevant targets for validation.

For example, our previous study revealed a set of molecular targets showing overexpression in immun-infiltrated SCLC [25], which might be clinically valuable for targeted immunotherapies in the future. However, not every target molecule is readily druggable, or clinical trials with small molecule inhibitors might have already failed against a selected substance. This was the case with MMP7, a matrix-metalloprotease exhibiting upregulation by manifolds in infiltrated SCLC. *EZCancerTarget* readily provided information with PubMed links to the failed Phase III clinical trial (performed almost 20 years ago), proving that its inhibitor, marimastat, is ineffective in this type of cancer. In contrast, in the case of molecular target ITGB6 (Integrin-beta 6) from the same screen [25], the software presented us with an available integrin antagonist compound, GSK3008348, used for an utterly different indication (pulmonary fibrosis), but with successful results. Through the STRING interaction, subcellular localization and pathway entries provided by *EZCancerTarget*, we learned that ITGB6 is structurally and functionally connected to the tumor microenvironment's proteoglycan components. Tenascin (TNC) was shown to be one of its closest interactors, whose role in lung cancer as an immunosuppressive agent promoting tumor recurrence has been reported by multiple studies [41, 42]. Since there is no available substance under study for the direct inhibition of TNC (source: clue.io), targeting ITGB6 might impede the whole pathway, promoting cancer propagation, making ITGB6 a druggable and biologically valid target.

Protein expression, subcellular localization and involved biological pathways matter for comprehensive experimental validation and clinical context. Altogether, if a research

team finds a molecular target for validation that has already clinically proven pharmaceutical agonists/antagonists, it is still essential to put the molecule in a biological context that Clue.io does not provide alone. EZCancerTarget has a filter function and does not display targets with at least one available compound supported by a preclinical study or a clinical trial. Thus, researchers begin with a target list already narrowed down based on drug availability.

The landscape of cancer research has been transformed by open-access genomic, proteomic databases, such as The Cancer Genome Database (TCGA) [43], OncoPrint [32], Human Protein Atlas [44], Cancer Cell Line Encyclopedia [45] or DepMap [46] and a multitude of drug-repurposing databases [3]. Still, casual users can face difficulties orchestrating high-throughput screenings and interpreting the vast amount of data generated using these platforms. Furthermore, researchers require the assistance of an expert bioinformatician to conduct systematic searches and data retrieval if multiple targets are eligible, which is not always feasible for smaller research teams.

Open-access EZCancerTarget obtains information based on the search engine of the groundbreaking drug-repurposing platform Clue.io. However, on the level of a casual user (like most researchers in the biomedical field without coding or IT experience), Clue.io can provide crisp information on one target at once and the users need to go through one-by-one their target list. In contrast, our software can scan through hundreds of targets simultaneously and provide the same information as Clue.io, but supplemented with references citing the relevant literature of preclinical studies or clinical trials. Sources regarding the biochemical details of the found compounds with their DrugBank, PubChem, and ChEMBL entries are also directly available for the users with one click. Another valuable addition to our software is that it provides a comprehensive overview of the molecular background of every target with an eligible search result. A concise and organized knowledge bank of these targets' pathways, molecular- and biological functions, and subcellular localization can serve educational purposes, but can also enhance the decision-making process for planned preclinical- or clinical validation studies.

An outstanding advantage of *EZCancerTarget* compared to other drug-repurposing platforms is that it does not require any experience in database handling and provides information on the biological background of our target molecules in a processed way that is easy to understand. The latter feature can be used for educational purposes as well. We believe that *EZCancerTarget* is a useful addition to the field of drug-repurposing in cancer science and oncology and will be particularly useful for smaller research groups with limited expertise in database handling.

#### Abbreviations

SCLC	Small-cell lung cancer
EMA	European Medicines Agency
TTD	Therapeutic Target Database
STITCH	Sequencing To Imputation Through Constructing Haplotypes
KEGG	Kyoto Encyclopedia of Genes and Genomes
FDA	U.S Food and Drug Administration
SIDER	Side Effect Resource
FAERS	FDA Adverse Event Reporting System
API	Application Programming Interface
REST	Representational State Transfer
MoA	Mechanism of Action

NSCLC	Non small-cell lung cancer
TNBC	Triple Negative Breast Cancer
PADC	Pancreatic Adenocarcinoma
PD-L1	Programmed Death Ligand 1

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13040-022-00307-9>.

### Additional file 1.

### Acknowledgements

The authors thank Glen J. Weiss, MD, MBA for critical review and feedback on this manuscript.

### Authors' contributions

Conceptualization, DD and ZL; Methodology and Software, CG and GS; Validation and Formal analysis, DD, TD and CG; Investigation, DD and CG; Resources, DD and ZL; Data curation, ZL, GS and CG; Visualisation, DD and TD; Supervision and Project administration, ZL, DD and TD; Funding acquisition, ZL; Writing-original draft, DD and ZL; Writing-reviewing and editing, all authors. All authors have read and agreed to the published version of the manuscript.

### Funding

Zoltan Lohinai acknowledge funding from the Hungarian National Research, Development and Innovation Office (OTKA #124652, OTKA #129664 and OTKA #128666). David Dora acknowledge funding from Semmelweis University's "Start-up" grant. Open access funding provided by Semmelweis University.

### Availability of data and materials

All the R scripts and description of versions and runtime environment is freely accessible at: <https://github.com/cycle20/EZCancerTarget>.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 17 January 2022 Accepted: 18 September 2022

Published: 1 October 2022

### References

- Gns HS, Gr S, Murahari M, Krishnamurthy M. An update on Drug Repurposing: Re-written saga of the drug's fate. *Biomed Pharmacother.* 2019;110:700–16. <https://doi.org/10.1016/j.biopha.2018.11.127> Epub 2018 Dec 12 PMID: 30553197.
- Parvathaneni V, Kulkarni NS, Muth A, Gupta V. Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discov Today.* 2019;24(10):2076–85. <https://doi.org/10.1016/j.drudis.2019.06.014> Epub 2019 Jun 22 PMID: 31238113.
- Tanoli Z, Seemab U, Scherer A, Wennerberg K, Tang J, Vähä-Koskela M. Exploration of databases and methods supporting drug repurposing: a comprehensive survey. *Brief Bioinform.* 2021;22(2):1656–78. <https://doi.org/10.1093/bib/bbaa003> PMID:32055842;PMCID:PMC7986597.
- UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480–9. <https://doi.org/10.1093/nar/gkaa1100> PMID:33237286;PMCID:PMC7778908.
- Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, InyStein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Current Protocols Bioinformatics.* 2016;54:1.30.1-1.30.33. <https://doi.org/10.1002/cpbi.5>.
- Chen X, Ji ZL, Chen YZ. TTD: Therapeutic Target Database. *Nucleic Acids Res.* 2002;30(1):412–5. <https://doi.org/10.1093/nar/30.1.412>.
- Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. *Nucleic acids Res.* 2018;36(Database issue):D684–8. <https://doi.org/10.1093/nar/gkm795>.
- Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 2021;30(1):187–200. <https://doi.org/10.1002/pro.3978>. Epub 2020 Nov 23.

9. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131> PMID:30476243;PMCID:PMC6323986.
10. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.PMID:10592173;PMCID:PMC102409.
11. Rodchenkov I, Babur O, Luna A, Aksoy BA, Wong JV, Fong D, Franz M, Siper MC, Cheung M, Wrana M, Mistry H, Mosier L, Dlin J, Wen Q, O'Callaghan C, Li W, Elder G, Smith PT, Dallago C, Cerami E, Gross B, Dogrusoz U, Demir E, Bader GD, Sander C. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* 2020;48(D1):D489–97. <https://doi.org/10.1093/nar/gkz946> PMID:31647099;PMCID:PMC7145667.
12. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol Cell Proteomics.* 2020. <https://doi.org/10.1074/mcp> PubMed PMID: 32907876.
13. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2019;49(D1):D1388–95. <https://doi.org/10.1093/nar/gkaa971>.
14. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Li D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2017. <https://doi.org/10.1093/nar/gkx1037>.
15. Fu C, Jin G, Gao J, Zhu R, Ballesteros-Villagrana E, Wong ST. DrugMap Central: an on-line query and visualization tool to facilitate drug repositioning studies. *Bioinformatics.* 2013 Jul 15;29(14):1834–6. doi: <https://doi.org/10.1093/bioinformatics/btt279>. Epub 2013 May 15. PMID: 23681121; PMCID: PMC3702253.
16. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* 2015. <https://doi.org/10.1093/nar/gkv1075>.
17. Tanoli Z, Alam Z, Ianevski A, Wennerberg K, Vähä-Koskela M, Aittokallio T. Interactive visual analysis of drug-target interaction networks using Drug Target Profiler, with applications to precision medicine and drug repurposing. *Brief Bioinform.* 2018. doi: <https://doi.org/10.1093/bib/bby119>. Epub ahead of print. PMID: 30566623.
18. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, Tusquets I, Albanell J, Rodon J, Taberner J, de Torres C, Dienstmann R, Gonzalez-Perez A, Lopez-Bigas N. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* 2018;10(1):25. <https://doi.org/10.1186/s13073-018-0531-8> PMID:29592813;PMCID:PMC5875005.
19. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* 2019;47(W1):W357–64. <https://doi.org/10.1093/nar/gkz382>.PMID: 31106366;PMCID:PMC6602486.
20. Hecker N, Ahmed J, von Eichborn J, Dunkel M, Macha K, Eckert A, Gilson MK, Bourne PE, Preissner R. SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Res.* 2012;40(Database issue):D1113–7. <https://doi.org/10.1093/nar/gkr912>. Epub 2011 Nov 8. PMID: 22067455; PMCID: PMC3245174.
21. Eichelbaum M, Altman RB, Ratain M, Klein TE. New feature: pathways and important genes from PharmGKB. *Pharmacogenet Genomics.* 2009;19:403.
22. Josephs KS, Berner A, George A, Scott RH, Programme HEEGE, Firth HV, Tatton-Brown K. Genomics: the power, potential and pitfalls of the new technologies and how they are transforming healthcare. *Clin Med (Lond).* 2019;19(4):269–72. <https://doi.org/10.7861/clinmedicine.19-4-269>.
23. Corsello S, Bittker J, Liu Z, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med.* 2017;23:405–8. <https://doi.org/10.1038/nm.4306>.
24. Bryan J, Citro C, Wickham H (2022). gargle: Utilities for Working with Google APIs. <https://gargle.r-lib.org>, <https://github.com/r-lib/gargle>.
25. Dora D, Rivard C, Yu H, Pickard SL, Laszlo V, Harko T, Megyesfalvi Z, Dinya E, Gerdan C, Szegvari G, Hirsch FR, Dome B, Lohinai Z. Characterization of Tumor-Associated Macrophages and the Immune Microenvironment in Limited-Stage Neuroendocrine-High and -Low Small Cell Lung Cancer. *Biology.* 2021;10(6):502. <https://doi.org/10.3390/biology10060502>.
26. Marra A, Trapani D, Viale G, Criscitiello C, Curigliano G. Practical classification of triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug resistance, and novel therapies. *NPJ Breast Cancer.* 2020;16(6):54. <https://doi.org/10.1038/s41523-020-00197-2> PMID:33088912;PMCID:PMC7568552.
27. Neophytou C, Boutsikos P, Papageorgis P. Molecular Mechanisms and Emerging Therapeutic Targets of Triple-Negative Breast Cancer Metastasis. *Front Oncol.* 2018;22(8):31. <https://doi.org/10.3389/fonc.2018.00031>.PMID:29520340; PMCID:PMC5827095.
28. Newton EE, Mueller LE, Treadwell SM, Morris CA, Machado HL. Molecular Targets of Triple-Negative Breast Cancer: Where Do We Stand? *Cancers (Basel).* 2022;14(3):482. <https://doi.org/10.3390/cancers14030482> PMID:35158750;PMCID:PMC8833442.
29. Wang S, Zheng Y, Yang F, Zhu L, Zhu XQ, Wang ZF, Wu XL, Zhou CH, Yan JY, Hu BY, Kong B, Fu DL, Bruns C, Zhao Y, Qin LX, Dong QZ. The molecular biology of pancreatic adenocarcinoma: translational challenges and clinical perspectives. *Signal Transduct Target Ther.* 2021;6(1):249. <https://doi.org/10.1038/s41392-021-00659-4> PMID:34219130;PMCID:PMC8255319.
30. Qian Y, Gong Y, Fan Z, Luo G, Huang Q, Deng S, Cheng H, Jin K, Ni Q, Yu X, Liu C. Molecular alterations and targeted therapy in pancreatic ductal adenocarcinoma. *J Hematol Oncol.* 2020;13(1):130. <https://doi.org/10.1186/s13045-020-00958-3> PMID:33008426;PMCID:PMC7532113.
31. Yan W, Liu X, Wang Y, Han S, Wang F, Liu X, Xiao F, Hu G. Identifying Drug Targets in Pancreatic Ductal Adenocarcinoma Through Machine Learning, Analyzing Biomolecular Networks, and Structural Modeling. *Front Pharmacol.* 2020;30(11):534. <https://doi.org/10.3389/fphar.2020.00534> PMID:32425783;PMCID:PMC7204992.
32. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia.* 2004;6(1):1–6. [https://doi.org/10.1016/s1476-5586\(04\)80047-2](https://doi.org/10.1016/s1476-5586(04)80047-2) PMID: 15068665; PMCID: PMC1635162.

33. Banck H, Dugas M, Müller-Tidow C, Sandmann S. Comparison of Open-access Databases for Clinical Variant Interpretation in Cancer: A Case Study of MDS/AML. *Cancer Genomics Proteomics*. 2021;18(2):157–66. <https://doi.org/10.21873/cgp.20250>.
34. Krempel R, Kulkarni P, Yim A, Lang U, Habermann B, Frommolt P. Integrative analysis and machine learning on cancer genomics data using the Cancer Systems. 2018.
35. Pantziarka P, Capistrano IR, De Potter A, Vandeborne L, Bouche G. An open access database of licensed cancer drugs. *Front Pharmacol*. 2012;12:627574. <https://doi.org/10.3389/fphar.2021.627574>.
36. Wishart DS, Bartok B, Oler E, Liang K, Budinski Z, Berjanskii M, Guo A, Cao X, Wilson M. MarkerDB: an online database of molecular biomarkers. *Nucleic Acids Res*. 2021;49(D1):D1259–67. <https://doi.org/10.1093/nar/gkaa1067>.
37. Creighton CJ. Making Use of Cancer Genomic Databases. *Current protocols in molecular biology*. 2018;121, 19.14.1–19.14.13. <https://doi.org/10.1002/cpmb.49>
38. Gadaleta E, Lemoine NR, Chelala C. Online resources of cancer data: barriers, benefits and lessons. *Brief Bioinform*. 2011;12(1):52–63. <https://doi.org/10.1093/bib/bbq010>.
39. Horn L, Mansfield AS, Szczesna A, Havel L, Krzakowski M, Hochmair MJ, Huemer F, Losonczy G, Johnson ML, Nishio M, et al. First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *N Engl J of Med*. 2018;379:2220–9 [PubMed] [Google Scholar].
40. Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, Statsenko G, Hochmair MJ, Özgüroğlu M, Ji JH, et al. Durvalumab plus platinum–etoposide versus platinum–etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2019;394:1929–39 [PubMed] [Google Scholar].
41. Gocheva V, Naba A, Bhutkar A, Guardia T, Miller KM, Li CM, Dayton TL, Sanchez-Rivera FJ, Kim-Kiselak C, Jaikhani N, Winslow MM, Del Rosario A, Hynes RO, Jacks T. Quantitative proteomics identify Tenascin-C as a promoter of lung cancer progression and contributor to a signature prognostic of patient survival. *Proc Natl Acad Sci U S A*. 2017;114(28):E5625–E5634. doi: <https://doi.org/10.1073/pnas.1707054114>. Epub 2017 Jun 26. PMID: 28652369; PMCID: PMC5514763.
42. Parekh K, Ramachandran S, Cooper J, Bigner D, Patterson A, Mohanakumar T. Tenascin-C, over expressed in lung cancer down regulates effector functions of tumor infiltrating lymphocytes. *Lung Cancer*. 2005;47(1):17–29. <https://doi.org/10.1016/j.lungcan.2004.05.016> PMID: 15603851.
43. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113–20. <https://doi.org/10.1038/ng.2764>.
44. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. doi: <https://doi.org/10.1126/science.1260419>. PMID: 25613900.
45. Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483:603–7.
46. 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. <https://doi.org/10.7554/eLife.57116>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

