



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Chemogenomic library design strategies for precision oncology, applied to phenotypic profiling of glioblastoma patient cells

### Citation for published version:

Athanasiadis, P, Ravikumar, B, Elliott, R, Dawson, JC, Carragher, NO, Clemons, PA, Johanssen, T, Ebner, D & Aittokallio, T 2023, 'Chemogenomic library design strategies for precision oncology, applied to phenotypic profiling of glioblastoma patient cells', *iScience*, vol. 26, no. 7, pp. 107209.  
<https://doi.org/10.1016/j.isci.2023.107209>

### Digital Object Identifier (DOI):

[10.1016/j.isci.2023.107209](https://doi.org/10.1016/j.isci.2023.107209)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

iScience

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

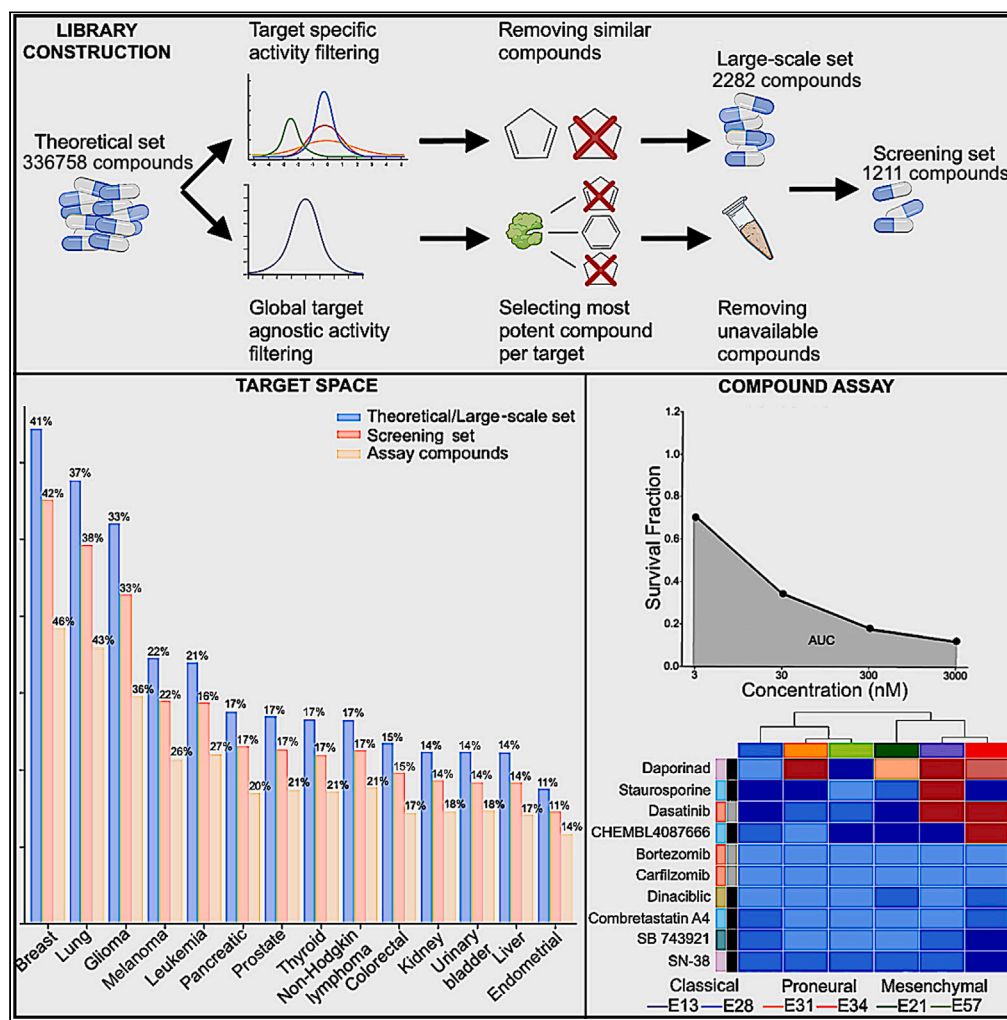
### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Article

# Chemogenomic library design strategies for precision oncology, applied to phenotypic profiling of glioblastoma patient cells



Paschalis Athanasiadis, Balaguru Ravikumar, Richard J.R. Elliott, ..., Timothy Johanssen, Daniel Ebner, Tero Aittokallio

tero.aittokallio@helsinki.fi

Highlights

Systematic strategies for designing targeted anticancer small-molecule libraries

Minimal screening library of 1,211 compounds for targeting 1,386 anticancer proteins

Pilot screening study identified patient-specific vulnerabilities in glioblastoma

Compound and target annotations and the pilot screening data made freely available

Athanasiadis et al., iScience  
26, 107209  
July 21, 2023 © 2023 The Author(s).  
<https://doi.org/10.1016/j.isci.2023.107209>



## Article

## Chemogenomic library design strategies for precision oncology, applied to phenotypic profiling of glioblastoma patient cells

Paschalis Athanasiadis,<sup>1,2,7</sup> Balaguru Ravikumar,<sup>3,7</sup> Richard J.R. Elliott,<sup>4</sup> John C. Dawson,<sup>4</sup> Neil O. Carragher,<sup>4</sup> Paul A. Clemons,<sup>5</sup> Timothy Johanssen,<sup>6</sup> Daniel Ebner,<sup>6,8</sup> and Tero Aittokallio<sup>1,2,3,8,9,\*</sup>

## SUMMARY

**Designing a targeted screening library of bioactive small molecules is a challenging task since most compounds modulate their effects through multiple protein targets with varying degrees of potency and selectivity. We implemented analytic procedures for designing anticancer compound libraries adjusted for library size, cellular activity, chemical diversity and availability, and target selectivity. The resulting compound collections cover a wide range of protein targets and biological pathways implicated in various cancers, making them widely applicable to precision oncology. We characterized the compound and target spaces of the virtual libraries, in comparison with a minimal screening library of 1,211 compounds for targeting 1,386 anticancer proteins. In a pilot screening study, we identified patient-specific vulnerabilities by imaging glioma stem cells from patients with glioblastoma (GBM), using a physical library of 789 compounds that cover 1,320 of the anticancer targets. The cell survival profiling revealed highly heterogeneous phenotypic responses across the patients and GBM subtypes.**

## INTRODUCTION

In the past ten years, there have been encouraging advances in the treatment and consequently, the survival rates, for many cancers. This progress has largely been driven by increased molecular understanding and classification of distinct cancer subtypes, the development of novel therapeutics focusing on targets associated with specific disease subtypes, and advances in both the type and range of therapeutic molecules available for the treatment of patients. When considering molecules that have been approved or advanced in clinical trials, there are successful examples of targeted chemotherapeutics,<sup>1</sup> small interfering RNAs,<sup>2</sup> monoclonal antibodies,<sup>3</sup> microRNAs,<sup>4</sup> and virotherapy,<sup>5</sup> among other drug classes. These developments are very encouraging considering the complexity of human cancers and the difficulty in developing effective treatments for advanced disease. However, small-molecule chemotherapeutics still make up the vast majority of approved drugs available to oncologists treating cancers.<sup>6–8</sup> In the case of glioblastoma (GBM) brain tumors, small-molecule chemotherapeutics are currently the only approved treatment modalities beyond surgery and radiation<sup>9</sup> and represent the most fertile ground for future innovation of new therapeutics to address the current challenges inherent in treating brain tumors. These challenges include (i) breaching the blood-brain barrier to effectively deliver therapeutics to the tumor, (ii) developing combinatorial treatments to systematically target redundant signaling pathways and tumor vulnerabilities inherent in brain tumors which typically exhibit wide intra- and inter-tumor heterogeneity, and (iii) selectively targeting GBM stem cells, which have been shown to be the main source for cancer recurrence in GBM.<sup>10</sup>

Traditional drug development often employs high-throughput drug screening of large collections of diverse small-molecule libraries against a nominated therapeutic target to identify chemical starting points (hit compounds) for further optimization. This process has been relatively successful at the industrial level, but it tends to become less successful at the academic level, due to the prohibitive infrastructural costs required to develop and prosecute large-scale screens,<sup>11</sup> as well as the cost to develop hit compounds to viable leads and sequentially to an investigational drug candidate. Instead, many academic target discovery facilities have focused on developing more physiologically relevant phenotypic

<sup>1</sup>Institute for Cancer Research, Department of Cancer Genetics, Oslo University Hospital, 0310 Oslo, Norway

<sup>2</sup>Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, 0317 Oslo, Norway

<sup>3</sup>Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, 20520 00290 Helsinki, Finland

<sup>4</sup>Cancer Research UK Scotland Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XR, UK

<sup>5</sup>Chemical Biology and Therapeutics Science Program, Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, MA 02142, United States

<sup>6</sup>Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK

<sup>7</sup>These authors contributed equally

<sup>8</sup>Senior author

<sup>9</sup>Lead contact

\*Correspondence: [tero.aittokallio@helsinki.fi](mailto:tero.aittokallio@helsinki.fi)  
<https://doi.org/10.1016/j.isci.2023.107209>



assays, which better recapitulate disease biology,<sup>12,13</sup> and screening of smaller, more focused libraries of small compounds, such as collections developed and curated for drug repurposing,<sup>14–16</sup> probes for target discovery,<sup>5,17,18</sup> or pharmacologically active targeted compound sets for specific target classes.<sup>7,19,20</sup> This focused approach has the advantage of providing a better understanding of the molecular basis of the disease, while simultaneously exploiting existing therapeutics and compounds with known safety profiles, along with probes that possess drug-like properties and compounds with known protein targets. In complex diseases of unmet therapeutic need, where target biology is still poorly understood or where disease heterogeneity indicates multiple target pathways that contribute to disease progression (as exemplified by GBM), phenotypic screening of target-annotated compound libraries in relevant patient-derived cell models may provide a valuable strategy for empirical identification of druggable targets or drug combinations. By circumventing major pitfalls, such as poor selectivity, cellular activity, and biological or target space diversity, these targeted libraries have great potential to accelerate the drug discovery process.

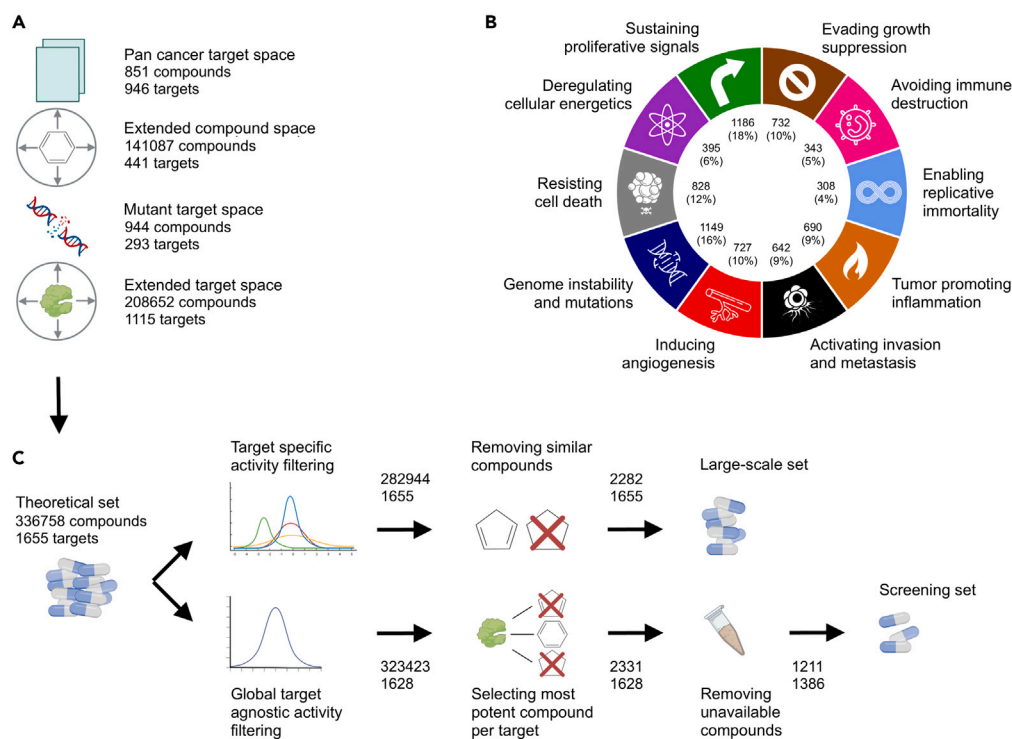
Here, we describe the construction of a comprehensive anticancer target-annotated compound library, designed to interrogate a wide range of potential cancer targets in phenotypic screening. The library design is approached as a multi-objective optimization (MOP) problem, where the aim is to maximize the cancer target coverage, while guaranteeing compounds' cellular potency and selectivity, and minimize the number of compounds arrayed into the final screening library. To do this, we used two target-based design strategies. First, we searched for small molecules against the druggable cancer targets among approved and investigational compounds (AICs) identified from the literature, drug databases, and existing oncology collections. Second, to expand the target-annotated compound library, we surveyed several pan-cancer studies to identify anticancer compound-target pairs and then expanded the chemical space around those novel targets by identifying additional bioactive compound probes through database queries. Importantly, cancer-mutated proteins, nearest neighbors, and influencer targets were further investigated for potential small-compound interactors, which generated a large *in silico* probe set collection. Finally, we refined the probe set collection by applying several filters, with adjustable activity and similarity thresholds, and removed redundant structures and compounds which could not be readily sourced at the time of library curation, to yield a sufficiently diverse, focused, and target-annotated compound library for phenotypic screening purposes, named the Comprehensive anti-Cancer small-Compound Library, or C<sup>3</sup>L. In the pilot application of C<sup>3</sup>L to cell survival profiling of patient-derived GBM stem cell models, we discovered widely heterogeneous patient-specific vulnerabilities and target pathway activities. All the compound libraries and their target and compound annotations, as well as the pilot screening data, are freely available as data spreadsheets and through an interactive web platform ([www.c3explorer.com](http://www.c3explorer.com)).

## RESULTS

### Identifying and curating small-molecule inhibitors of cancer-associated targets

Our first design objective was to define a comprehensive list of protein targets associated with the development and progression of cancers to form the basis of the anticancer small-molecule library. We first defined a list of proteins known to be implicated in cancers using The Human Protein Atlas<sup>21</sup> and nominal targets of pan-cancer studies from the PharmacoDB,<sup>22</sup> leading to the target space of 946 oncoproteins (Figure 1A; see STAR Methods for details). We then expanded the target space by using additional pan-cancer studies linked back to cancer-related targets to define the full set of 1,655 proteins or other cancer-associated gene products. Our target space was designed to span a wide range of protein families, cellular functions, and cancer phenotypes, and it covers all the categories of "hallmarks of cancer"<sup>23,24</sup> (Figure 1B).

After defining the comprehensive list of cancer-associated targets, our next objective was to identify and curate a small-molecule collection of compounds targeting these proteins. Since the compounds ranged from investigational and experimental probe compounds (EPCs) to approved drugs, we took a systematic approach to defining each source, through sorting the compounds targeting the cancer-associated targets, ranking the compounds for activity, diversity, and availability, and finally, producing a sortable and searchable database of the screening library along with its target and other annotations. The screening library construction started from >300,000 small molecules and ended up with 1,211 optimized for physical library size, cellular activity, chemical diversity, and target selectivity, leading to 150-fold decrease in compound space, yet still covering 84% of the cancer-associated targets.



**Figure 1. Target-based pipeline for designing of the three experimental probe compound (EPC) sets (Theoretical, Large-scale, and Screening set)**

(A) Four target-based compound collections were defined: (i) the pan-cancer collection includes compounds and their annotated targets from various pan-cancer studies; (ii) the extended compound space consists of additional compounds that have off-target activity against the annotated targets based on bioactivity data from public drug/target repositories such as ChEMBL,<sup>25</sup> Drug Target Commons (DTC),<sup>26</sup> and DrugBank;<sup>27</sup> (iii) the mutant target space compounds have activity against the mutant variants of the annotated targets extracted from the COSMIC database;<sup>28</sup> (iv) the extended target space includes compounds with activity against cancer-related targets established through a nearest neighbor approach.<sup>29</sup>

(B) The targets of the compounds in the theoretical set cover all the hallmarks of cancer. Note: each target protein may belong to multiple hallmarks explaining why they do not add up to 1,655.

(C) The filtering criteria for designing the large-scale and physical screening sets. See STAR Methods for the step-by-step procedures for the construction of the various probe set collections.

### Target-based approach: EPC collection

Using the target-based design, we extracted compound-target interactions manually from public databases, leading to chemical probes and investigational compounds in three nested subsets of gradually decreasing sizes; (i) the **theoretical set** is an *in silico* set curated from established target-compound pairs covering the expanded target space of 1,655 cancer-associated proteins, (ii) the **large-scale set** is a broader screening collection of filtered compounds covering the same target space as theoretical set, and (iii) the **screening set** is the final set of most potent probes arrayed into the physical library. The screening set is the smallest subset due to the limitations in compound commercial availability for screening purposes. Figure 1C shows a schematic diagram of the construction of these three target-based compound sets, with their compound and target numbers.

The **theoretical set** contains 336,758 unique compounds from four probe sets for pan-cancer target space, pan-cancer collection with extended compound space, mutant target space, and mutant target collection with extended target space. The **large-scale set** contains a subset of 2,288 compounds from the theoretical set, filtered to reduce the number of molecules in the library while still covering the same target space, using both the activity and similarity filtering procedures with pre-defined cutoff values (see Figure S4 for the filtering parameters used in this study, but the cutoff parameters are freely adjustable for other studies and library constructions). The large-scale set was based on both the on- and off-target profiles of the 2,282 small-molecule

compounds, which could be used in larger-scale screenings campaigns in academic or industrial projects (Table S1).

The **screening set** contains a smaller number (1,211) of compounds that are more easily purchasable and used for screening applications, thus making the probe set suitable for routine exploration of oncology-associated biological target space in complex phenotypic assays and identification of potential candidates for further drug development. This probe subset was obtained by subjecting the theoretical set to three filtering procedures (see Figure 1C). These filtering procedures involved (i) global target-agnostic activity filtering to remove 13,335 non-active probes, (ii) selecting the most potent compounds for each target to reduce the library to 2,331 compounds, and (iii) filtering by the availability of the compounds, which reduced the library size by 52%, while the target coverage remained at 86% and the target activity distributions were relatively unchanged ( $p > 0.05$ ; Kolmogorov-Smirnov [K-S] test; Figure S1).

### Drug-based approach: AIC collection

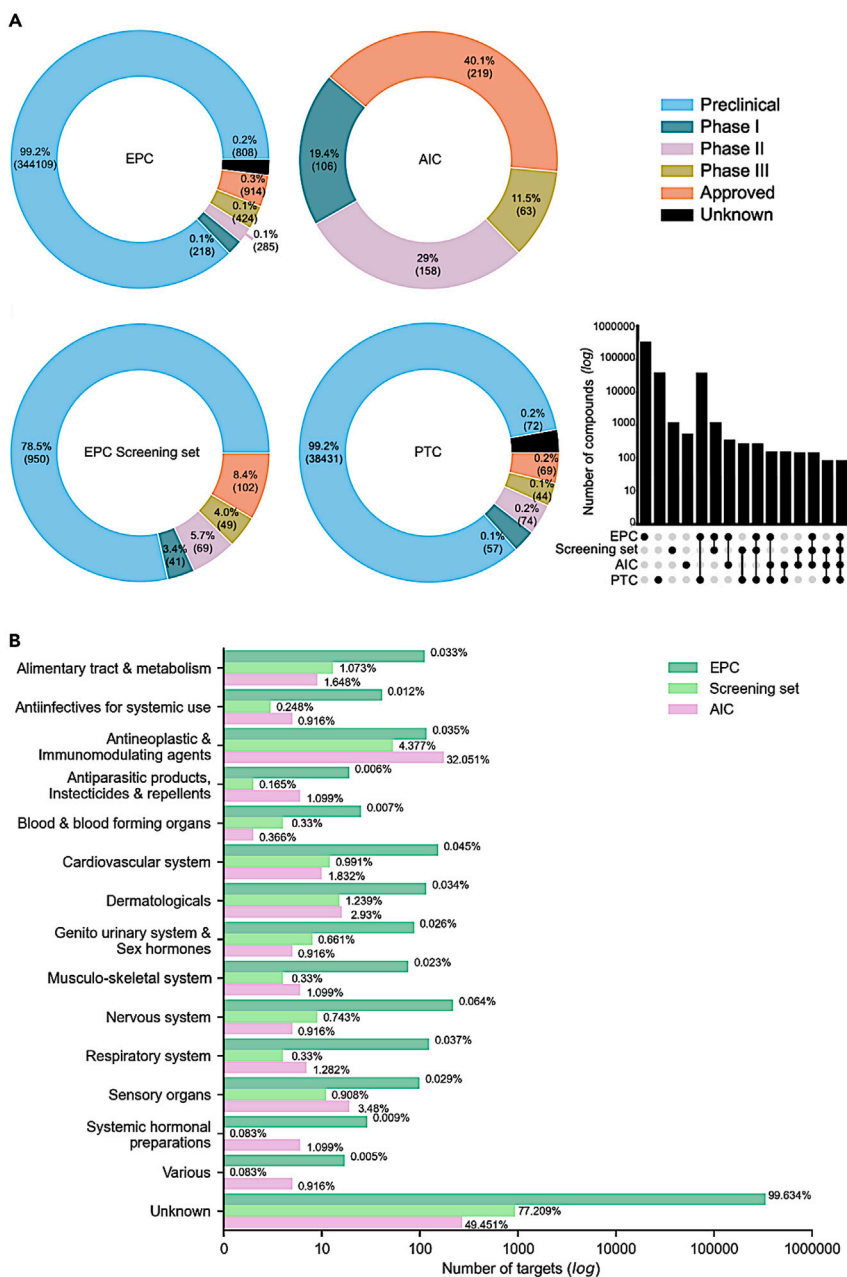
As expected from the target-based approach (i.e., identifying established potent small molecules for respective targets), EPC collections included mostly compounds that are currently in preclinical stages (Figure 2A). We next compared EPC collection with an additional set of small molecules that are currently approved for clinical use, also including anticancer compounds in various clinical development stages that might be candidates for drug repurposing applications. This compound-based design strategy led to a complementary AIC collection, named AIC collection, manually curated from several public compound sources and clinical trials. The AIC collection was further subjected to removal of duplicate molecules and similarity searches using the extended connectivity fingerprint (ECFP4/6) and molecular ACCess system (MACCS) fingerprints, where Dice similarity for ECFP4/6 and Tanimoto similarity for MACCS keys with cutoff of  $\geq 0.99$  were used to identify and remove structurally highly similar compounds (e.g., doxorubicin and epirubicin).

The AIC collection consisted of 546 unique compounds, out of which 74 (14%) are labeled as standard chemotherapeutics (Figure S2). Most of the molecularly targeted compounds in AIC collections are kinase inhibitors (50%), but the collection also includes various other drug classes, such as immunomodulatory agents (3.7%) and metabolic modifiers (2.9%). As expected, the AIC compounds contain a large number of approved drugs (40.1%), and the compounds in this collection only partly overlapped with those of the EPC collections (64% with EPC and 27% with screening set; Upset plot in Figure 2A). However, even if the total number of compounds is rather different in the collections, the AIC and EPC collections cover the same anatomical and therapeutic classes (ATCs), and the proportions of ATCs in the screening set were somewhere in between the AIC and EPC collections (Figure 2B).

To further compare the target-based EPC and compound-based AIC collections to a functional genomics-based approach to construct an anticancer compound library, we identified potent inhibitors of the 628 priority targets extracted from the CRISPR-Cas9 screens of the DepMap project,<sup>30</sup> using a similar procedure to the one implemented for the EPC collection (see STAR Methods). The priority target compound (PTC) collection of 38,747 compounds included mainly preclinical compounds (Figure 2A), with a large number of compounds overlapping with the EPC collection (99%). The four compound collections had a core set of 85 common compounds shared by all the sets (Figure 2A, upset plot). Strikingly, PTC collection covers only 113 out of the 628 priority targets (18%), indicating that most of the loss-of-function screen targets are not currently druggable. The EPC screening set has more approved compounds (102, 8.4%), compared to the PTC collection (69, 0.2%).

### Characterization of the compound and target spaces of the probe collections

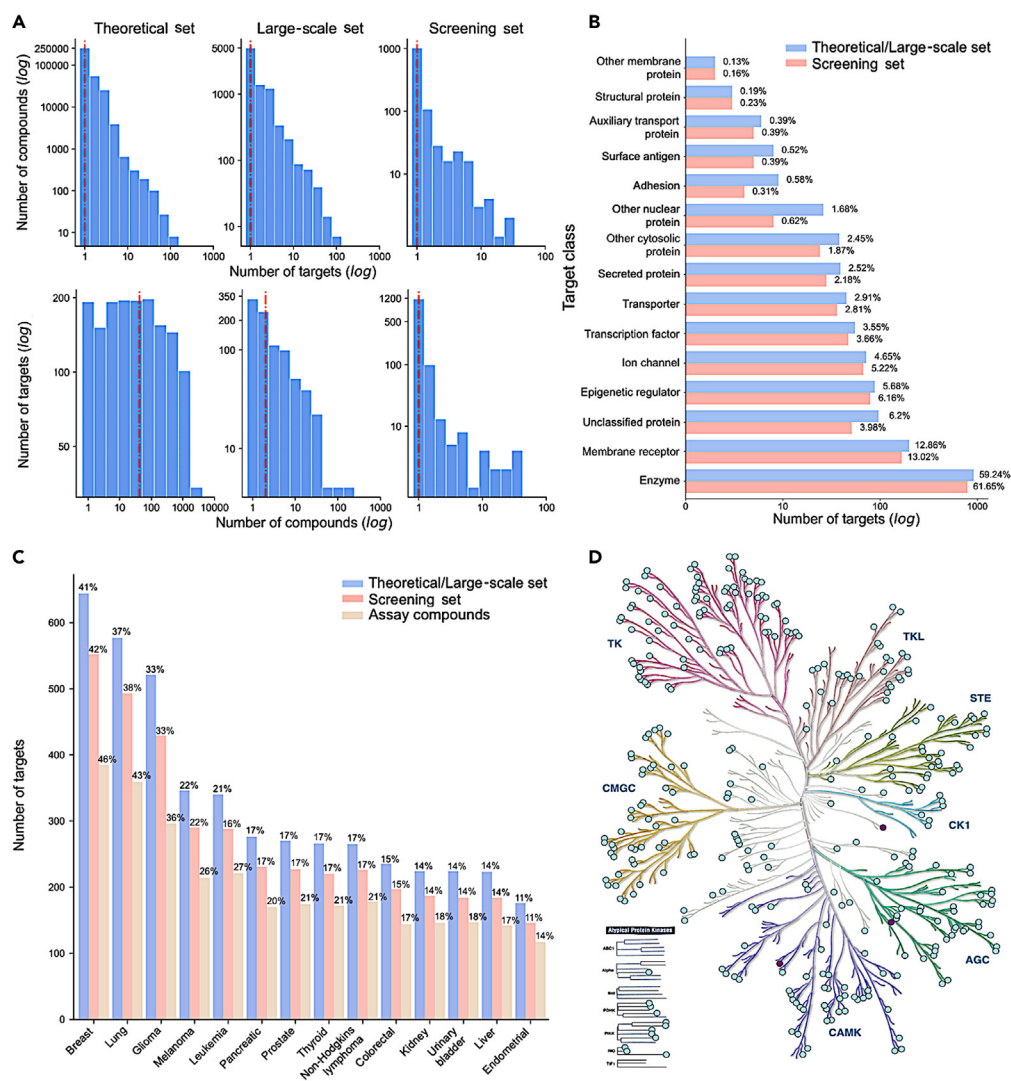
After comparing the EPC collections to the approved/investigational and priority target collections, we next analyzed the compound and target spaces of the three EPC collections, designed using the target-based procedures and selected parameter values for filtering. The target distributions of the three probe collections remained relatively similar, whereas the screening set showed reduced numbers of multi-target compounds by its design (Figure 3A, upper panel). However, the median number of targets per compound was one for each probe collection, indicating that the sets include relatively selective compounds. For most of the targets, the screening set contains only a single potent compound, again per its design principles, whereas in the theoretical and large-scale sets the median number of compounds per target was 42 and 2, respectively (Figure 3A, lower panel).



**Figure 2. Compound spaces of the target-based and compound-based collections**

(A) Clinical development phases of the compound collections: target-based approach - experimental probe compound (EPC) collection, compound-based approach - approved and investigational compound (AIC) collection, and CRISPR9-based approach - priority target compound (PTC) collection (compounds targeting the 628 priority targets from the DepMap project). The clinical development phase was extracted from ChEMBL.<sup>25</sup> Numbers in parentheses indicate the number of compounds in each category. The upset diagram shows the size of the intersections between the compound sets. The sets that are part of an intersection are filled in, and when only a single set is filled in, that indicates the number of compounds in that set.

(B) The distribution of compounds in the collections in terms of the anatomical therapeutic chemical (ATC) classes, which distinguishes the organ or system on which the compounds act and their therapeutic, pharmacological, and chemical properties.



**Figure 3. Target spaces of the compound collections**

(A) The number of targets per compound reported in drug databases (upper panel) and the number of compounds per target (lower panel) in the three compound collections. The dashed line indicates the median. The x axis and y axis are  $\log_{10}$ -scaled in each panel, while the numbers present the non-logged counts (note the different scales between the panels). The counts are based on the target activity threshold of  $\leq 1000\text{nM}$  (see Figure S7 for the other activity thresholds). (B) The distribution of target classes of compounds in the screening set and theoretical/large-scale sets (that contain the same set of targets). The numbers next to the bars indicate the percentage of targets in the two collections of different sizes. The target classification of proteins was extracted from ChEMBL.<sup>25</sup>

(C) The number of targets associated with cancer types in the theoretical/large-scale, screening set and in the physical compound set that was tested in the patient-derived GBM stem cells in an imaging-based assay. The numbers above the bars indicate the percentage of targets in the two collections of different sizes. The disease associations were extracted from OpenTargets,<sup>31</sup> with overall association score  $>0.5$  (<https://www.opentargets.org/>).

(D) Kinase families covered by the targets in the probe sets. The kinases colored in purple indicate the targets present only in the theoretical/large-scale sets, but not in the screening set, while the light-green targets are covered by all the three collections. KinMap<sup>32</sup> web-tool was used for the creation of the illustration, reproduced courtesy of Cell Signaling Technology, Inc. ([www.cellsignal.com](http://www.cellsignal.com)).

The target class distribution of the screening set is similar to that of the larger probe collections of the theoretical and large-scale sets (K-S test,  $p = 0.19$ ; Figure 3B). This indicates that the compound filtering did not miss any target class. As expected, targeted inhibitors of kinases and other enzymes are the most frequent, but there are also other well-covered target classes, such as membrane receptors, epigenetic regulators,



and ion channels (Figure 3B). When zooming into kinases, the screening set covers all the major kinase families, and only 3 of the kinase targets present in the theoretical/large-scale sets are missed in the smaller screening set (Figure 3D). Importantly, the compound sets cover similar proportions of targets implicated in multiple types of cancers (K-S test,  $p = 1.0$ ), with breast and lung cancers having the largest numbers of targeted compounds followed by glioma (Figure 3C).

The target distribution across cancer types remained very similar in the constructed physical library, which included 325 of the screening set compounds, and was used here in the pilot screening study on patient-derived GBM stem cells (assay compounds, Figure 3C), described in the next section. At the time of the curation, 325 compounds from the optimized library were sourced first from collaborators and then from multiple commercial providers. The remaining 886 compounds could not be sourced as they were not available from any vendor, were prohibitively expensive, or the lead time for acquisition of the compound was excessive. In these cases, we selected the most potent and selective inhibitor instead, yielding a completed C<sup>3</sup>L of 789 small molecules, covering 1,320 protein targets and 79% of the cancer-associated targets.

### Identification of compound responses across GBM subtypes in primary patient cells

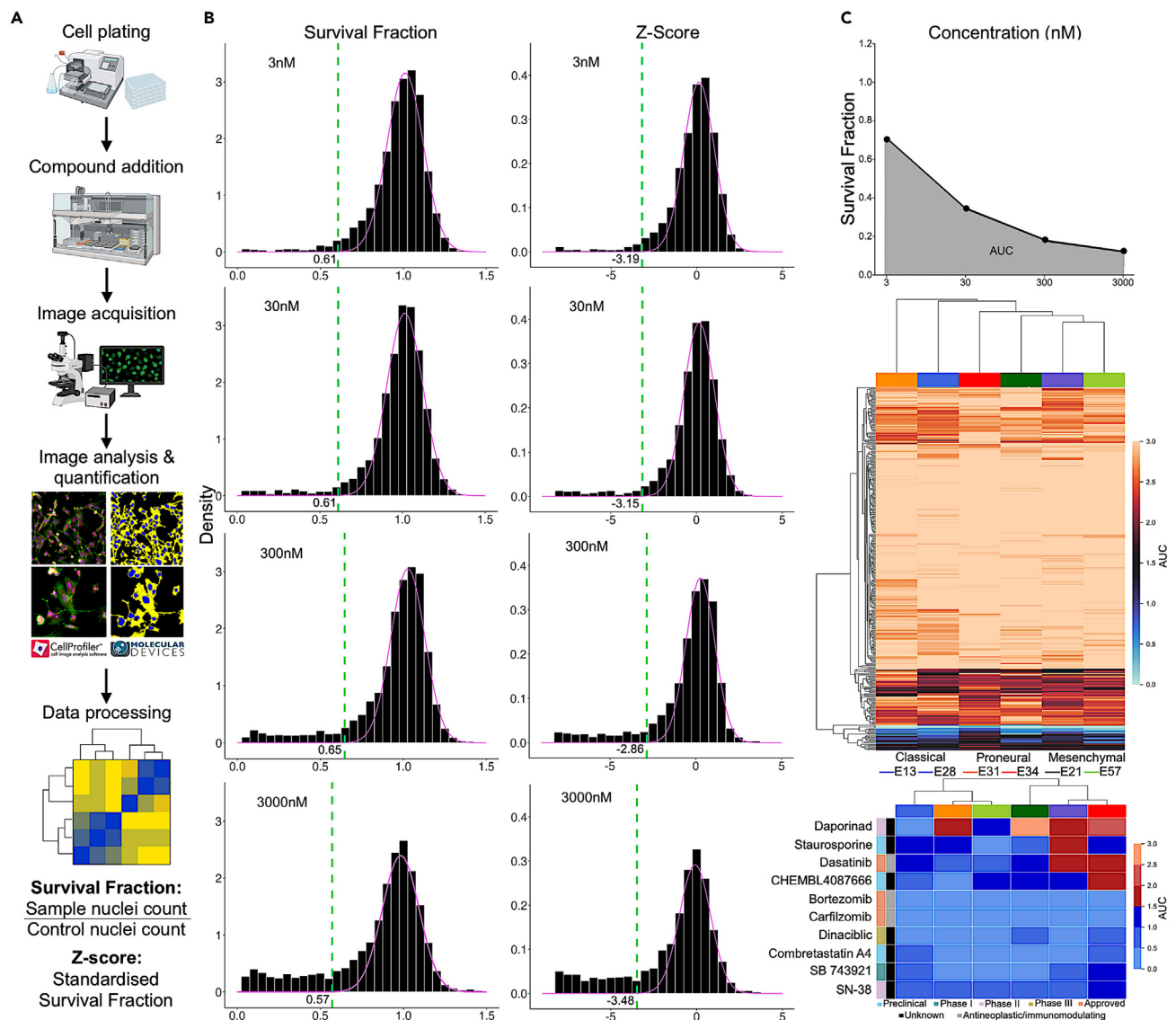
In a pilot screening application to identify therapeutic vulnerabilities in GBM, we carried out phenotypic profiling of patient-derived stem cell lines from 6 GBM patients covering the three GBM subtypes:<sup>33</sup> classical ( $n = 2$ ), proneural ( $n = 2$ ), and mesenchymal ( $n = 2$ ). The imaging-based compound testing assay quantifies the phenotypic effects of each compound by tracking cell proliferation or survival in response to the 789 compounds in the C<sup>3</sup>L physical library, each tested with 4 concentrations (see STAR Methods for details). These compounds and their 1,320 anticancer targets cover 27% and 79% of the compound and target spaces of the screening set, respectively. In this pilot screening study, we used a single phenotypic readout for further analysis: the survival fraction and its Z score, calculated based on nuclei counts in treated, untreated, and negative control DMSO samples (Figure 4A). The lower these readouts are, the more effective is the compound in the sense that fewer cancer cells survived the tested treatment.

We developed a data-driven method to define an activity threshold separately for each patient and dose, based on mixture modeling and parameter estimation (Figure 4B, see STAR Methods for details). Using a cutoff of 0.01 percentile of the background distribution, we identified as strong hits those compounds with responses under the data-driven activity thresholds across all the 4 concentrations. When using either the survival fraction or Z score at 0.01 percentile, we identified 10 strong response compounds from the screening set that cover multiple target classes. The benefit of such a data-driven threshold is that it provides an objective way to select patient-specific compounds that show strong activity at various concentration levels since it is estimated from the distribution of the measurements. Notably, the activity thresholds varied between the patients and dose levels, supporting the need for data-driven approach (Figure 4B).

To summarize the activity of a compound across the 4 tested concentrations, we calculated the area under the dose-response curve (AUC) for each compound-patient pair using the survival fraction as the response readout (Figure 4C, upper panel). The patient-specific response profiles showed a relatively large variability between the cell lines, even from the same GBM subtype, and the compound response profiles did not cluster according to the GBM subtypes (Figure 4C, middle panel). This supports the notion that GBM is a heterogeneous disease, also in terms of phenotypic responses to compound treatment, hence requiring patient-specific treatment approaches. The identified 10 strong hits included 3 antineoplastic and immunomodulating agents, such as approved multiple myeloma treatments bortezomib and carfilzomib, which showed an overall high potency in all the 6 GBM patient-derived cell models (Figure 4C, bottom panel).

### Screening hits identify patient-specific vulnerabilities and target pathway activities in GBM

When investigating the 10 strong hit compounds across the concentration levels and patient samples, one can see an expected dose-response relationships, where higher concentrations lead to increased responses of the compounds (Figure 5A). However, the compounds have differing activity profiles both within and between the GBM subtypes. For instance, dinaciclib, an investigational compound that inhibits cyclin-dependent kinases (CDKs), showed high activity in multiple patients already at lower concentrations (30nM), except for one proneural subtype patient. The broad efficacy of dinaciclib was shown recently also in 2D and 3D GBM models and in long-term culture experiments.<sup>34</sup> In contrast, another broadly active kinase inhibitor dasatinib induced more mixed responses across the GBM subtypes, in contrast to earlier studies that have reported that patients with the mesenchymal subtype show more sensitivity to dasatinib



**Figure 4. Cell survival profiling and data analysis**

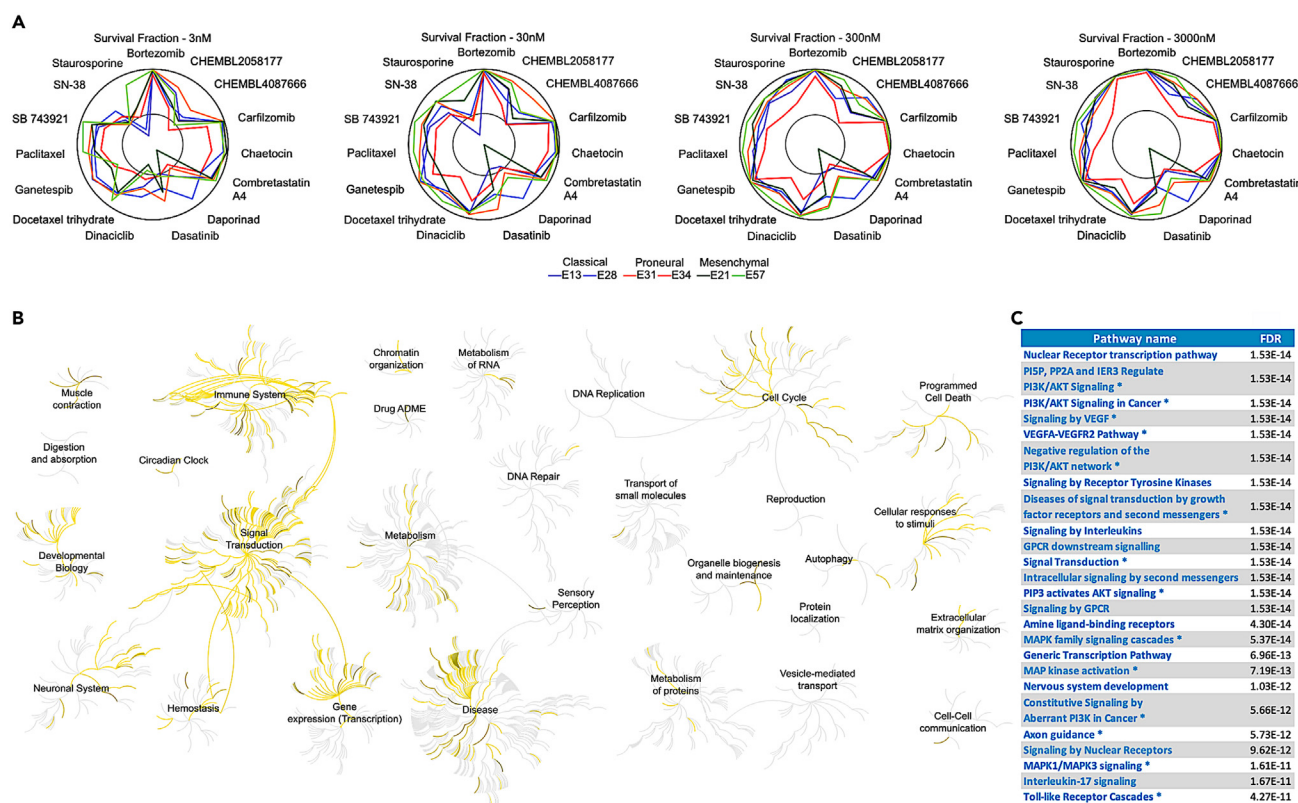
(A) The imaging-based workflow to quantify patient-specific phenotypic features (see STAR Methods for details).

(B) Examples of survival fraction distributions across 4 concentration levels of two patient-derived cell lines from two GBM subtypes: E34 (proneural) and E57 (mesenchymal). Lower values of survival fraction indicate higher compound activity. The green dashed line indicates the data-driven activity threshold for classifying a compound as a strong hit at the particular concentration level, when using the activity cutoff of 0.01 percentile of the background distribution. A total of 10 compounds with responses under the activity thresholds across all the 4 concentrations were identified as strong hits and were selected for further analysis.

(C) Calculation of the area under the dose-response curve (AUC). Lower AUC values indicate higher activity of the compound. Two heatmaps of the AUC for all the 325 compounds from the screening set (upper heatmap) and for the 10 strong hits (lower heatmap, corresponding to the lowermost cluster of the full dendrogram) across 6 glioblastoma patient primary samples. The sample colors indicate the GBM subtypes, and the compound colors mark the clinical phase and ATC classes.

compared to proneural and classical subtypes.<sup>35</sup> Notably, nicotinamide phosphoribosyltransferase (NMPRTase) inhibitor daporinad demonstrated highly selective response in a single patient of classical subtype, while the other patients had much lower responses to daporinad at all concentrations (Figure 5A).

Pathway analyses among the protein targets of strong hit compounds demonstrated that these compounds modulate multiple cellular pathways, including immune system, signal transduction, neuronal



**Figure 5. Analysis of the top hits from the patient screening**

(A) Radial plots of the 10 strong hit compounds and their survival fractions across the 4 concentration levels and 6 glioblastoma patient samples. Responses closer to the outer circle indicate lower survival fraction and therefore higher response. The inner gray circle marks the data-driven activity threshold of 0.01 percentile for the particular concentration level.

(B) Reactome target pathway activities related to the 10 strong hit compounds.

(C) The top 25 Reactome target pathways ranked by their false discovery rate (FDR). \*Pathway enrichment also identified based on the KEGG pathways.

system, cell cycle, and cellular responses to stimuli (Figure 5B). There are also less-pronounced pathway activities that originate from more selective responses; for instance, daporinad triggers autophagy. More detailed pathway enrichment results from both Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG) highlight PI3K/Akt/mitogen-activated protein kinase (MAPK) signaling as one of the top-enriched pathways (Figure 5C), which is important in GBM as ~40% of GBM patients have phosphatase and TENSin homolog (PTEN) alterations (in our panel, 4 out of 6 [66.6%] of the patient-derived cells had PTEN alteration), as well as frequent neurofibromatosis type 1 (NF1) loss (E57 cells) or alterations to the epidermal growth factor receptor (EGFR) receptor (amplification in 4 out of 6 patient-derived cells; missense mutation in the E31 cells). To enable others to explore these data and freely set other data-driven thresholds for the compound selection, we have implemented a web application ([www.c3explorer.com](http://www.c3explorer.com)) that allows the users to interactively explore the signaling and other pathways targeted by the selected compounds for additional analyses and hypothesis generation.

## DISCUSSION

Phenotypic screening provides a systematic approach to identify novel small molecules as promising future therapeutics, especially when focusing on realistic disease models and relevant target landscapes.<sup>36</sup> In this work, we developed various analytical procedures for the construction of targeted anticancer compound libraries, designed to interrogate a wide range of potential cancer-related targets for phenotypic screening. We started the systematic library construction from more than 300,000 molecules, by compiling comprehensive chemical, activity, and target information, and ended up with a physical screening set of 1,211 molecules with pre-defined cellular activity, biological and chemical diversity, target selectivity, and compound availability. This reduced the compound space by 150-fold, whereas

the remaining compounds in the screening set still covered 84% of the target space of 1,655 cancer-related proteins. The remaining 16% represent targets that need further development due to a lack of either enough specific and potent chemical probes (2%) or compound availability for sale (14% at the time of library curation).

We applied the C<sup>3</sup>L design principles to imaging-based phenotypic profiling of glioma stem cells from GBM patients, using a pilot library of 789 compounds, and identified patient-specific compound responses and target pathway activities. We anticipate these systematic library-design principles and the resulting broadly annotated small-molecule libraries will prove useful for the community in various phenotypic screening experiments in GBM and other cancers. We have made the compound libraries and their annotations available in the form of open-access data sheets (<https://github.com/PaschalisAthanas/C3L>) and implemented a web platform that enables others to interactively explore the screening library and corresponding compound-target interaction networks, as well as screening data and pathway activities using user-defined activity thresholds and other filtering criteria ([www.c3explorer.com](http://www.c3explorer.com)). This enables the community to access these data and make further analyses, e.g., drug class-specific responses or for identifying combinatorial treatments to target multiple pathway redundancies and patient-specific intra-tumor vulnerabilities that are urgently needed for treating patients with GBM.<sup>37</sup>

In the field of precision oncology, there have been several efforts to systematically catalog all of the genes implicated in cancer,<sup>38,39</sup> and several in-house, commercial, and governmental libraries of AICs targeting many of the most heavily investigated proteins are readily available.<sup>14,15,40,41,41–46</sup> However, even though these commercial or in-house libraries have been well-utilized in many anticancer screens and drug repurposing studies,<sup>16,17,47,48</sup> there is a need for a comprehensive, target-annotated, anticancer library optimized for selectivity and diversity against all known oncology targets. There are well-annotated libraries for specific classes or applications, such as for kinase inhibitors,<sup>49</sup> metabolic modifiers,<sup>20</sup> and drug repurposing,<sup>50</sup> but what is lacking is a comprehensive and high-quality probe library that provides a starting point for various anticancer screening applications. Despite the emergence of new treatment modalities, such as monoclonal antibodies, molecularly targeted small molecules are, and will most likely remain, the most prolific anticancer therapeutics for the foreseeable future that are also applicable in advanced stages of the disease, for instance, after tumor has progressed toward metastatic disease and when the cancer has become chemo- or radiotherapy resistant.<sup>6–8</sup>

The development and implementation of well-curated and comprehensive screening libraries are expected to facilitate the identification of novel anticancer drug targets, drug combinations, synthetic-lethal interactions, and therapeutic targets to overcome cancer resistance that goes beyond the recurrently mutated genes.<sup>51</sup> To enable cross-comparisons, we have carefully annotated our libraries with multiple compound IDs for molecules and target annotations; these include unique ChEMBL ID that makes it easy to extract additional information of the compounds, compound generic name of the active ingredient, chemical abstracts service (CAS) registry number for sourcing the compounds from potential vendors, and simplified molecular input line entry system (SMILES) for describing the structure of chemical species for further chemical analyses. We have also carefully annotated the active on/off-target space of the compounds using multiple databases, such as ChEMBL, DrugBank, and DrugTargetCommons, using both gene accession and gene symbol of the target proteins. Such annotations make it easy for others to further characterize the chemical and target spaces of the compound libraries, as well as compare with the existing and emerging libraries to allow for comparative analyses of their shared and unique chemical and target spaces for future library developments and screening applications.

We formulated the compound library construction as an MOP problem, where the aim is to simultaneously maximize the anticancer target diversity and compound potency (or selectivity), while minimizing the number and cost of the compounds in the physical screening set. This is important because a targeted library enables a more thorough evaluation of the anticancer potential in advanced, physiologically relevant cancer models, such as multiple genetically distinct patient-derived cell panels. With such a focused yet diverse-enough library, comprehensive screening paradigms including drug combinations, sequencing, and scheduling can be more easily incorporated into a primary phenotypic screening pipeline, achievable by many academic screening facilities.<sup>52</sup> In the present work, we used fast and heuristic procedures to select the compounds with a pre-defined potency against the selected targets that presented with sufficient structural dissimilarity, as implemented in several filtering procedures. Even though our results

show that the current heuristic approach leads to the desired results from the compound screening point of view, other approximated or exact solutions to the MOP problem could further decrease the size of the screening set, while guaranteeing sufficient potency and selectivity of the compounds against the disease-related and other targets.<sup>53</sup>

Designing an optimal compound library of small molecules is challenged by the compound promiscuity; that is, many compounds may modulate their effects through multiple protein targets with various degrees of potency in a given context. For instance, kinase inhibitors are notorious for their target promiscuity and wide polypharmacological effects across various target classes beyond kinase families.<sup>54</sup> The wider target selectivity of many compounds remains still uncharted,<sup>54,55</sup> and therefore the phenotype-driving targets of many compounds are currently still unknown in many cellular contexts or disease applications. Compared to other compound libraries that focus more on compound selectivity, C<sup>3</sup>L was designed based on dual objectives: (i) to identify compounds that are effective against GBM or other cancers (either individually or in combination) and (ii) C<sup>3</sup>L function as a target-identifying or hypothesis-generating library for follow-up studies. In the first application, target specificity is not as important as compound efficacy. In the second application, specificity is more important but not a necessary condition for hypothesis generation. Systematic study of cross-reactivity of the compounds will be needed to further investigate the target selectivity of the molecules in the current screening set across various target classes and cell contexts, beyond the rather limited data for target activity currently available in public databases.

### Limitations of the study

We highlight below some potential caveats of the present work and how the library design could be further improved in the future studies. In addition to the potency and selectivity, chemical similarity of the compounds is another important feature of library design, especially if the aim is to have a collection of not too similar small molecules that target the cancer-related proteins of interest. In addition, further phenotypic and clinical characterization of the compounds, as implemented in Drug Repurposing Hub<sup>50</sup> and ChemicalChecker,<sup>56</sup> could aid several drug discovery tasks, including target identification, mechanism of action (MoA) classification, and library characterization. Integrated analysis of chemical, molecular target, cell-based profiling, and clinical information could also be useful to provide a relevance ranking of the targets and compounds in a library, once the disease application is defined, hence adding one further dimension to the MOP problem. Additional filtering steps to aid the eventual screening applications include, for instance, removal of interference compounds, finding on-target compounds with diverse scaffold profiles, and exclusion of compounds with known adverse off-target activities. The relatively small size of the physical screening library used in the pilot screen may miss some relevant compounds and targets in GBM. To investigate and improve the relevance of the hits, follow-up phenotypic screens with full morphological cell painting profiling are warranted.

The commercial availability of the compounds was a limiting factor that reduced the physical library size significantly. Inclusion of synthesizable compounds from the theoretical set can further increase the diverse and coverage of the screening set in the future versions. Even though screening of compounds with overlapping primary targets increases the cost, having multiple compounds with overlapping primary targets may become beneficial in specific applications, e.g., for validating the screening hits from high-throughput screens, target deconvolution of the phenotypic responses, or for discovering target activity differences when screening at multiple concentrations. The future applications in phenotypic screening, either in established cell lines or patient-derived cell models, will define the usefulness of any compound collections and libraries and the cost-benefit trade-off of compound libraries with various sizes and target overlaps. We expect that the herein presented comprehensive libraries with curated target information and the pilot screening data in patient-derived cells will provide new opportunities for many exciting drug discovery applications in GBM and other cancers, including novel target identifications, target deconvolution, drug repositioning, and drug combination prediction and testing.<sup>57</sup> Follow-up studies of the screening hits may lead to the identification of chemical starting points, new target-directed drug discovery programs, or approval of new therapeutics for GBM and other cancers.

### SUPPORTING CITATIONS

63,64

## STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Tissue culture/cell maintenance
  - 384-well plating/dosing
  - 384-Well plate fix/stain/acquisition
  - Appendix A
- **METHOD DETAILS**
  - Approved and investigational compound (AIC) collection
  - Experimental probe compound (EPC) collection
  - Reducing the number of compounds
  - Cell survival profiling of glioblastoma patient cells
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **ADDITIONAL RESOURCES**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.107209>.

## ACKNOWLEDGMENTS

The authors thank Dr. David Drewry (UNC Eshelman School of Pharmacy) for his critical review and useful comments on the manuscript draft, Dr. Jani Saarela (FIMM) for his help with the library construction and compound sourcing, and LifeArc for assistance in compound sourcing. We also acknowledge Professor Steven Pollard and Dr Gillian Morrison (University of Edinburgh) and the CRUK Glioma Cellular Genetic resource for the provision of patient-derived glioma stem cells. The authors acknowledge funding support from Cancer Research UK and the Brain Tumour Charity (grant REF: C42454/A28596), Norwegian Cancer Society (grant 216104), Helse Sør-Øst (grant 2020026), Radium Hospital Foundation, and the Finnish Cancer Foundation. P.A.C. was supported in part by the National Cancer Institute's Cancer Target Discovery and Development (CTD<sup>2</sup>) Network (U01CA217848).

## AUTHOR CONTRIBUTIONS

Conceptualization, D.E. and T.A.

Methodology, P.A., B.R., R.J.R.E., J.C.D., N.O.C., P.A.C., T.J., and D.E.

Software, P.A.

Formal Analysis, P.A. and T.A.

Investigation, P.A., B.R., R.J.R.E., and J.C.D.

Resources, N.O.C, P.A.C, D.E.

Data Curation, P.A. and B.R.

Writing – Original Draft, P.A., B.R., T.J., D.E., and T.A.

Writing – Review & Editing, P.A., B.R., R.J.R.E., J.C.D., N.O.C., P.A.C., T.J., D.E., and T.A.

Visualization, P.A.

Supervision, D.E., N.O.C., and T.A.

Funding Acquisition, N.O.C., P.A.C., D.E., and T.A.

## DECLARATION OF INTERESTS

P.A.C. is an advisor to Pfizer, Inc. and Belharra Therapeutics. N.O.C. is founder and shareholder of PhenoTherapeutics Ltd.

Received: January 29, 2023

Revised: February 21, 2023

Accepted: June 21, 2023

Published: June 25, 2023

## REFERENCES

- Doebele, R.C., Drilon, A., Paz-Ares, L., Siena, S., Shaw, A.T., Farago, A.F., Blakely, C.M., Seto, T., Cho, B.C., Tosi, D., et al. (2020). Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. *Lancet Oncol.* *21*, 271–282. [https://doi.org/10.1016/S1470-2045\(19\)30691-6](https://doi.org/10.1016/S1470-2045(19)30691-6).
- Schultheis, B., Strumberg, D., Kuhlmann, J., Wolf, M., Link, K., Seufferlein, T., Kaufmann, J., Feist, M., Gebhardt, F., Khan, M., et al. (2020). Safety, Efficacy and Pharmacokinetics of Targeted Therapy with The Liposomal RNA Interference Therapeutic Atu027 Combined with Gemcitabine in Patients with Pancreatic Adenocarcinoma. *Cancers* *12*, 3130. <https://doi.org/10.3390/cancers12113130>.
- Zahavi, D., and Weiner, L. (2020). Monoclonal Antibodies in Cancer Therapy. *Antibodies* *9*, 34. <https://doi.org/10.3390/antib9030034>.
- van Zandwijk, N., Pavlakis, N., Kao, S.C., Linton, A., Boyer, M.J., Clarke, S., Huynh, Y., Chrzanowska, A., Fulham, M.J., Bailey, D.L., et al. (2017). Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol.* *18*, 1386–1396. [https://doi.org/10.1016/S1470-2045\(17\)30621-6](https://doi.org/10.1016/S1470-2045(17)30621-6).
- Macedo, N., Miller, D.M., Haq, R., and Kaufman, H.L. (2020). Clinical landscape of oncolytic virus research in 2020. *J. Immunother. Cancer* *8*, e001486. <https://doi.org/10.1136/jitc-2020-001486>.
- Crans, D.C., Henry, L., Cardiff, G., and Posner, B.I. (2019). Small Molecules: The Past or the Future in Drug Innovation? *Met. Ions Life Sci.* *19*, 203–230. <https://doi.org/10.1515/9783110527872-008>.
- van der Zanden, S.Y., Luimstra, J.J., Neeffjes, J., Borst, J., and Ovaas, H. (2020). Opportunities for Small Molecules in Cancer Immunotherapy. *Trends Immunol.* *41*, 493–511. <https://doi.org/10.1016/j.it.2020.04.004>.
- Pharmaceuticals B. Small Molecules to Treat Cancer - Research into Novel Therapies 2020. <https://www.research.bayer.com/en/cancer-causes-cancer-research-new-drug-products-chemotherapy.aspx>.
- Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* *352*, 987–996. <https://doi.org/10.1056/NEJMoa043330>.
- Prager, B.C., Bhargava, S., Mahadev, V., Hubert, C.G., and Rich, J.N. (2020). Glioblastoma Stem Cells: Driving Resilience through Chaos. *Trends Cancer* *6*, 223–235. <https://doi.org/10.1016/j.trecan.2020.01.009>.
- Spear, K.L., and Brown, S.P. (2017). The evolution of library design: crafting smart compound collections for phenotypic screens. *Drug Discov. Today Technol.* *23*, 61–67. <https://doi.org/10.1016/j.ddtec.2017.05.001>.
- Horvath, P., Aulner, N., Bickle, M., Davies, A.M., Nery, E.D., Ebner, D., Montoya, M.C., Östling, P., Pietiäinen, V., Price, L.S., et al. (2016). Screening out irrelevant cell-based models of disease. *Nat. Rev. Drug Discov.* *15*, 751–769. <https://doi.org/10.1038/nrd.2016.175>.
- Bray, M.-A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* *11*, 1757–1774. <https://doi.org/10.1038/nprot.2016.105>.
- Drug Synthesis and Chemistry Branch (DSCB). [https://dtp.cancer.gov/organization/dscb/obtaining/available\\_plates.htm](https://dtp.cancer.gov/organization/dscb/obtaining/available_plates.htm).
- Selleck Chemicals. Anti-cancer Compound Library 2020 [A unique collection of 4046 anti-cancer compounds for multiple cancers: Breast Cancer, Leukemia, Lung Cancer, Lymphoma, etc]. <https://www.selleckchem.com/screening/anti-cancer-compound-library.html>.
- Patel, A., Seraia, E., Ebner, D., and Ryan, A.J. (2020). Adefovir dipivoxil induces DNA replication stress and augments ATR inhibitor-related cytotoxicity. *Int. J. Cancer* *147*, 1474–1484. <https://doi.org/10.1002/ijc.32966>.
- Hughes, R.E., Elliott, R.J.R., Munro, A.F., Makda, A., O'Neill, J.R., Hupp, T., and Carragher, N.O. (2020). High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery. *SLAS Discov.* *25*, 770–782. <https://doi.org/10.1177/2472555220917115>.
- Jones, L.H., and Bunnage, M.E. (2017). Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* *16*, 285–296. <https://doi.org/10.1038/nrd.2016.244>.
- Wells, C.I., Al-Ali, H., Andrews, D.M., Asquith, C.R.M., Axtman, A.D., Dikic, I., Ebner, D., Etmayer, P., Fischer, C., Frederiksen, M., et al. (2021). The Kinase Chemogenomic Set (KCGS): An Open Science Resource for Kinase Vulnerability Identification. *Int. J. Mol. Sci.* *22*, 566. <https://doi.org/10.3390/ijms22020566>.
- Pemovska, T., Bigenzahn, J.W., Srndic, I., Lercher, A., Berghaler, A., César-Razquin, A., Kartnig, F., Kornauth, C., Valent, P., Staber, P.B., and Superti-Furga, G. (2021). Metabolic drug survey highlights cancer cell dependencies and vulnerabilities. *Nat. Commun.* *12*, 7190. <https://doi.org/10.1038/s41467-021-27329-x>.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* *347*, 1260419. <https://doi.org/10.1126/science.1260419>.
- Smirnov, P., Kofia, V., Maru, A., Freeman, M., Ho, C., El-Hachem, N., Adam, G.-A., Ba-Alawi, W., Safikhani, Z., and Haibe-Kains, B. (2018). PharmacDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res.* *46*, D994–D1002. <https://doi.org/10.1093/nar/gkx911>.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell*

- 144, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
24. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 12, 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>.
  25. Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Motowo, P., Atkinson, F., Bellis, L.J., Cibrián-Uhalte, E., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Res.* 45, D945–D954. <https://doi.org/10.1093/nar/gkw1074>.
  26. Tang, J., Tanoli, Z.-U.-R., Ravikumar, B., Alam, Z., Rebane, A., Vähä-Koskela, M., Peddinti, G., van Adrichem, A.J., Wakkinen, J., Jaiswal, A., et al. (2018). Drug Target Commons: A Community Effort to Build a Consensus Knowledge Base for Drug-Target Interactions. *Cell Chem. Biol.* 25, 224–229.e2. <https://doi.org/10.1016/j.chembiol.2017.11.009>.
  27. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>.
  28. Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet. Chapter 10*, Unit.10.11. <https://doi.org/10.1002/0471142905.hg1011s57>.
  29. Módos, D., Bulusu, K.C., Fazekas, D., Kubisch, J., Brooks, J., Marczell, I., Szabó, P.M., Vellai, T., Csermely, P., Lenti, K., et al. (2017). Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. *npj Syst. Biol. Appl.* 3, 1–13. <https://doi.org/10.1038/s41540-017-0003-6>.
  30. Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 568, 511–516. <https://doi.org/10.1038/s41586-019-1103-9>.
  31. Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., Carmona, M., Faulconbridge, A., Hercules, A., McAuley, E., et al. (2019). Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* 47, D1056–D1065. <https://doi.org/10.1093/nar/gky1133>.
  32. Eid, S., Turk, S., Volkamer, A., Rippmann, F., and Fulle, S. (2017). KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinf.* 18, 16. <https://doi.org/10.1186/s12859-016-1433-7>.
  33. Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110. <https://doi.org/10.1016/j.ccr.2009.12.020>.
  34. Riess, C., Koczan, D., Schneider, B., Linke, C., del Moral, K., Classen, C.F., and Maletzki, C. (2021). Cyclin-dependent kinase inhibitors exert distinct effects on patient-derived 2D and 3D glioblastoma cell culture models. *Cell Death Dis.* 7, 54. <https://doi.org/10.1038/s41420-021-00423-1>.
  35. Alhalabi, O.T., Fletcher, M.N.C., Hielscher, T., Kessler, T., Lokumcu, T., Baumgartner, U., Wittmann, E., Schlue, S., Göttmann, M., Rahman, S., et al. (2022). A novel patient stratification strategy to enhance the therapeutic efficacy of dasatinib in glioblastoma. *Neuro Oncol.* 24, 39–51. <https://doi.org/10.1093/neuonc/noab158>.
  36. Vincent, F., Nueda, A., Lee, J., Schenone, M., Prunotto, M., and Mercola, M. (2022). Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat. Rev. Drug Discov.* 21, 899–914. <https://doi.org/10.1038/s41573-022-00472-w>.
  37. Johanssen, T., McVeigh, L., Erridge, S., Higgins, G., Straehla, J., Frame, M., Aittokallio, T., Carragher, N.O., and Ebner, D. (2022). Glioblastoma and the search for non-hypothesis driven combination therapeutics in academia. *Front. Oncol.* 12, 1075559. <https://doi.org/10.3389/fonc.2022.1075559>.
  38. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 173, 371–385.e18. <https://doi.org/10.1016/j.cell.2018.02.060>.
  39. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93. <https://doi.org/10.1038/s41586-020-1969-6>.
  40. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41, D955–D961. <https://doi.org/10.1093/nar/gks1111>.
  41. Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* 12, 109–116. <https://doi.org/10.1038/nchembio.1986>.
  42. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* 5, 1210–1223. <https://doi.org/10.1158/2159-8290.CD-15-0235>.
  43. Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E.V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154, 1151–1161. <https://doi.org/10.1016/j.cell.2013.08.003>.
  44. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. <https://doi.org/10.1038/nature11003>.
  45. Heiser, L.M., Sadanandam, A., Kuo, W.-L., Benz, S.C., Goldstein, T.C., Ng, S., Gibb, W.J., Wang, N.J., Ziyad, S., Tong, F., et al. (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc. Natl. Acad. Sci. USA* 109, 2724–2729. <https://doi.org/10.1073/pnas.1018854108>.
  46. Mpindi, J.P., Yadav, B., Östling, P., Gautam, P., Malani, D., Murumägi, A., Hirasawa, A., Kangaspeska, S., Wennerberg, K., Kallioniemi, O., and Aittokallio, T. (2016). Consistency in drug response profiling. *Nature* 540, E5–E6. <https://doi.org/10.1038/nature20171>.
  47. Ravikumar, B., Timonen, S., Alam, Z., Parri, E., Wennerberg, K., and Aittokallio, T. (2019). Chemogenomic Analysis of the Druggable Kinome and Its Application to Repositioning and Lead Identification Studies. *Cell Chem. Biol.* 26, 1608–1622.e6. <https://doi.org/10.1016/j.chembiol.2019.08.007>.
  48. Gautam, P., Jaiswal, A., Aittokallio, T., Al-Ali, H., and Wennerberg, K. (2019). Phenotypic Screening Combined with Machine Learning for Efficient Identification of Breast Cancer-Selective Therapeutic Targets. *Cell Chem. Biol.* 26, 970–979.e4. <https://doi.org/10.1016/j.chembiol.2019.03.011>.
  49. Elkins, J.M., Fedele, V., Szklarz, M., Abdul Azeez, K.R., Salah, E., Mikolajczyk, J., Romanov, S., Sepetov, N., Huang, X.-P., Roth, B.L., et al. (2016). Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* 34, 95–103. <https://doi.org/10.1038/nbt.3374>.
  50. Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., Johnston, S.E., Vrcic, A., Wong, B., Khan, M., et al. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat. Med.* 23, 405–408. <https://doi.org/10.1038/nm.4306>.
  51. Hahn, W.C., Bader, J.S., Braun, T.P., Califano, A., Clemons, P.A., Druker, B.J., Ewald, A.J., Fu, H., Jagu, S., Kemp, C.J., et al. (2021). An expanded universe of cancer targets. *Cell* 184, 1142–1155. <https://doi.org/10.1016/j.cell.2021.02.020>.
  52. Shanks, E., Ketteler, R., and Ebner, D. (2015). Academic drug discovery within the United Kingdom: a reassessment. *Nat. Rev. Drug Discov.* 14, 510. <https://doi.org/10.1038/nrd4661>.
  53. Wang, T., Pulkkinen, O.I., and Aittokallio, T. (2022). Target-specific compound selectivity for multi-target drug discovery and



- repurposing. *Front. Pharmacol.* 13, 1003480. <https://doi.org/10.3389/fphar.2022.1003480>.
54. Moret, N., Clark, N.A., Hafner, M., Wang, Y., Lounkine, E., Medvedovic, M., Wang, J., Gray, N., Jenkins, J., and Sorger, P.K. (2019). Cheminformatics Tools for Analyzing and Designing Optimized Small-Molecule Collections and Libraries. *Cell Chem. Biol.* 26, 765–777.e3. <https://doi.org/10.1016/j.chembiol.2019.02.018>.
55. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., and Overington, J.P. (2017). A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 16, 19–34. <https://doi.org/10.1038/nrd.2016.230>.
56. Duran-Frigola, M., Pauls, E., Guitart-Pla, O., Bertoni, M., Alcalde, V., Amat, D., Juan-Blanco, T., and Aloy, P. (2020). Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker. *Nat. Biotechnol.* 38, 1098. <https://doi.org/10.1038/s41587-020-0564-6>.
57. Tarasov, V.V., Svistunov, A.A., Chubarev, V.N., Zatsepilova, T.A., Preferanskaya, N.G., Stepanova, O.I., Sokolov, A.V., Dostdar, S.A., Minyaeva, N.N., Neganova, M.E., et al. (2019). Feasibility of Targeting Glioblastoma Stem Cells: From Concept to Clinical Trials. *Curr. Top. Med. Chem.* 19, 2974–2984. <https://doi.org/10.2174/156802661966619112140939>.
58. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. <https://doi.org/10.1093/nar/gky1033>.
59. PostgreSQL Global Development Group (2011). PostgreSQL 9.0 Official Documentation - Volume I: The SQL Language (Fultus Corporation).
60. Landrum, G. RDKit. <http://www.rdkit.org>.
61. Van Rossum, G., and Drake, F.L., Jr. (2011). *The Python Language Reference Manual (Network Theory)*.
62. O'Boyle, N.M., and Sayle, R.A. (2016). Comparing structural fingerprints using a literature-based similarity benchmark. *J. Cheminf.* 8, 36. <https://doi.org/10.1186/s13321-016-0148-0>.
63. Roskoski, R., Jr. (2020). Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update. *Pharmacol. Res.* 152, 104609. <https://doi.org/10.1016/j.phrs.2019.104609>.
64. Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* 166, 740–754. <https://doi.org/10.1016/j.cell.2016.06.017>.
65. Jaiswal, A., Yadav, B., Wennerberg, K., and Aittokallio, T. (2019). Integrated Analysis of Drug Sensitivity and Selectivity to Predict Synergistic Drug Combinations and Target Coaddictions in Cancer. *Methods Mol. Biol.* 1888, 205–217. [https://doi.org/10.1007/978-1-4939-8891-4\\_12](https://doi.org/10.1007/978-1-4939-8891-4_12).
66. Fazekas, D., Koltai, M., Türei, D., Módos, D., Pálfi, M., Dúl, Z., Zsákai, L., Szalay-Bekó, M., Lenti, K., Farkas, I.J., et al. (2013). Signalink 2—a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.* 7, 7. <https://doi.org/10.1186/1752-0509-7-7>.
67. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The Reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687–D692. <https://doi.org/10.1093/nar/gkab1028>.
68. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database—2009 update. *Nucleic Acids Res.* 37, D767–D772. <https://doi.org/10.1093/nar/gkn892>.
69. Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451. <https://doi.org/10.1093/nar/gkh086>.
70. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. <https://doi.org/10.1093/nar/gkt1115>.
71. Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539. <https://doi.org/10.1093/nar/gkj109>.
72. Cui, Q., Ma, Y., Jaramillo, M., Bari, H., Awan, A., Yang, S., Zhang, S., Liu, L., Lu, M., O'Connor-McCourt, M., et al. (2007). A map of human cancer signaling. *Mol. Syst. Biol.* 3, 152. <https://doi.org/10.1038/msb4100200>.
73. Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52, 345–370. <https://doi.org/10.1007/bf02294361>.
74. Irwin, J.J., and Shoichet, B.K. (2005). ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182. <https://doi.org/10.1021/ci049714+>.
75. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
76. Huang, T., Peng, H., and Zhang, K. (2017). Model selection for gaussian mixture models. *Stat. Sin.* 27, 147–169. <https://doi.org/10.5705/ss.2014.105>.
77. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. <https://doi.org/10.1093/nar/gkaa970>.
78. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Patient derived Glioma stem cells	Glioma Cellular Genetics Resource	<a href="http://gcgr.org.uk">http://gcgr.org.uk</a>
<b>Chemicals, peptides, and recombinant proteins</b>		
Cultrex 3D matrix Laminin	R&D Systems <a href="http://rndsystems.com">rndsystems.com</a>	3446-005-01
Hoechst-33342	MolProbes	H1399
<b>Software and algorithms</b>		
TIBCO Spotfire Analyst	<a href="http://spotfire.tibco.com">Spotfire.tibco.com</a>	N/A
Imageexpress IXM (with built in software/image analysis modules)	Molecular Devices, <a href="http://moleculardevices.com">moleculardevices.com</a>	N/A
PostgreSQL	Postgresql global development group	<a href="https://www.postgresql.org/support/">https://www.postgresql.org/support/</a>
Python 3.7	Van Rossum G and Drake F.L. Jr	<a href="https://www.python.org/">https://www.python.org/</a>
RDKit	Landrum G.	<a href="http://www.rdkit.org">http://www.rdkit.org</a> .
ShinyGO	Ge, S.X., Jung, D., and Yao, R.	<a href="http://bioinformatics.sdstate.edu/go/">http://bioinformatics.sdstate.edu/go/</a>
<b>Other</b>		
Microplates 384w uCLEAR Black, F-Bottom, TC, Sterile	Greiner Bio-One. <a href="http://gbo.com">gbo.com</a>	781091
Integra ViaFill Multifunctional Reagent Dispenser	Integra Biosciences	N/A
PubChem	Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al.	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>
ChEMBL	Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L.J., Cibrián-Uhalte, E., et al.	<a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a>
SelleckChem	Selleck	<a href="https://www.selleckchem.com/screening/anti-cancer-compound-library.html">https://www.selleckchem.com/screening/anti-cancer-compound-library.html</a>
PharmacoDB	Smirnov, P., Kofia, V., Maru, A., Freeman, M., Ho, C., El-Hachem, N., Adam, G.-A., Ba-Alawi, W., Safikhani, Z., and Haibe-Kains, B.	<a href="https://pharmacodb.ca/">https://pharmacodb.ca/</a>
The Human Protein Atlas	Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al.	<a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a>
DrugTargetCommons	Tang, J., Tanoli, Z.-U.-R., Ravikumar, B., Alam, Z., Rebane, A., Vähä-Koskela, M., Peddinti, G., van Adrichem, A.J., Wakkinen, J., Jaiswal, A., et al.	<a href="http://drugtargetcommons.fimm.fi/">http://drugtargetcommons.fimm.fi/</a>
DrugBank	Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al.	<a href="https://go.drugbank.com/">https://go.drugbank.com/</a>
COSMIC	Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R.	<a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a>
Signalink v2	Fazekas, D., Koltai, M., Türei, D., Módos, D., Pálffy, M., Dúl, Z., Zsákai, L., Szalay-Bekő, M., Lenti, K., Farkas, I.J., et al.	<a href="http://signalink.org/">http://signalink.org/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Reactome	Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al.	<a href="https://reactome.org/">https://reactome.org/</a>
HPRD	Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al.	<a href="https://www.hprd.org/">https://www.hprd.org/</a>
DIP	Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D.	<a href="https://dip.doe-mbi.ucla.edu/dip/Main.cgi">https://dip.doe-mbi.ucla.edu/dip/Main.cgi</a>
IntAct	Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N.,	<a href="https://www.ebi.ac.uk/intact/home">https://www.ebi.ac.uk/intact/home</a>
BioGrid	Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M.	<a href="https://thebiogrid.org/">https://thebiogrid.org/</a>
ZINC15	Irwin, J.J., and Shoichet, B.K.	<a href="https://zinc15.docking.org/">https://zinc15.docking.org/</a>
KEGG	Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M.	<a href="https://www.genome.jp/kegg/">https://www.genome.jp/kegg/</a>

**RESOURCE AVAILABILITY****Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tero Aittokallio ([tero.aittokallio@helsinki.fi](mailto:tero.aittokallio@helsinki.fi)).

**Materials availability**

This study did not generate new unique reagents.

**Data and code availability**

- All data has been deposited at Zenodo (<https://zenodo.org/record/7945379>), and assay annotations are available at GitHub (<https://github.com/PaschalisAthanasios/C3L>).
- This paper does not report original code.
- Data exploration and visualization web-platform: <http://www.c3explorer.com/>.

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

Glioma stem cells obtained from the Glioma Cellular Genetics Resource (GCGR, Edinburgh, [gcgr.org.uk](http://gcgr.org.uk)) are unique, patient derived cell lines – hence no reference sequence is available after STR typing. Classification of these cell lines was performed at the DKFZ (German Cancer Research Centre in the Helmholtz Association), Heidelberg University Hospital (see [Table S2](#)).

**Tissue culture/cell maintenance**

*Materials Required* (see [Table S3](#)).

*Cell maintenance-thawing cells*

When ready the vial should be quickly thawed in a 37°C water bath. The thawed aliquot should be dispersed into 10ml of cold wash media and centrifuge for 3 minutes at 1.5K rpm. One should aspirate as much supernatant as possible and re-suspend in complete media (8ml, containing EGF and FGF) and laminin (10ug/ml) and transfer to a T25 flask. Then, incubate overnight at 37°C and 5% CO<sub>2</sub>. Once the cells have adhered to the plastic, fresh complete media, containing EGF, FGF, and laminin (10ug/ml for E57, 4ug/ml for other cell lines) is added. Take note that it may take a few days to recover. Do not allow to become approximately >80% confluent. Usually the flask will need to be split into a T75 after 48 hours post thaw. Lastly, do not split more than 1:5. The usual ratio is 1:2 or 1:3.

### *Cell maintenance - Passaging cells*

For the cell passaging procedure, the media should be removed and washed with 15ml PBS. This is followed by aspirating PBS and adding 1ml accutase in order to coat the bottom of each flask. To ensure complete coverage the flask should be rocked. Then, incubation follows for 3-4 mins at 37°C and 5% CO<sub>2</sub>. Then, 9ml of wash media is added, and the contents are transferred to 15ml conical tube, appropriate for centrifuge (for 3 mins at 1.5 rpm). The supernatant is then aspirated and the pellet is resuspended in 10ml complete media, EGF, FGF and Laminin (4ug/mL) in a T75 flask. The final volume of media is 15ml and 45ug laminin (150ug laminin for E57 cells), but the split can be done as preferred. We note that one should not split more than 1:6 (usually 1:3 a few days before plating) and not allow the split to become (approximately) >80% confluent. Yield is approximately 3x10<sup>6</sup> cells per T75 flask and viability is usually 80-90%. We also note that doubling time is approximately 24 (E57) to 60 (other lines) hours. Cells tend to form spheroids when becoming too confluent and indicates they should be split. Spheroids can be recovered and will re-adhere. Also, even if cells do not need to be passaged, fresh media should be refreshed after 4 days to encourage healthy growth. Lastly, cells grow faster with prolonged culture. For this reason, passage should exceed 30 and the lowest passage should be used for screening.

### *Cell maintenance - Freezing cells*

The freezing media consists of wash media and 10% DMSO. Cells should be detached as described above, counted and resuspended in freezing media at approximately 1-5-2x10<sup>6</sup> cells per ml, and 1ml for thawing into a T25 flask. Then the cells should be transferred to a specialised freezing chamber and freeze overnight at -80°C, and as soon as possible to liquid nitrogen storage.

### **384-well plating/dosing**

*Materials/Equipment Required* (see [Table S4](#)).

#### *384-Well plate pre-coating with laminin*

For the pre-coating of the wells, we start first by preparing a diluted solution of laminin (10ug/ml) in complete media. Then a ViaFill dispenser is prepared under sterile conditions, and is followed by adding solution of 20ug/well and incubating at 37°C and 5% CO<sub>2</sub> for a minimum of 2 hours. Then 25uL of cell suspension (without aspiration) is added at 500-1500 cells/well.

#### *384-Well plate seeding*

Cell suspensions should be prepared as described above, followed by pelleting cells and resuspending those in wash media. The cells are then counted. A ViaFill dispenser is prepared under sterile conditions, and the cell suspension is prepared in complete media (without laminin), seeding at 1000 cells/25ul (E3, E21, E31, E34), 500 cells/25ul (E57) and 1500 cells/25ul (E28), and adding 25ul per well, resulting in a final volume of 45ul (~4ug/ml laminin). Incubation for 30 min at RT is followed, and then plates are transferred to an incubator at 37°C and 5% CO<sub>2</sub>, and are incubated overnight prior to dosing with the library. We noted that for positive control, Staurosporine (1uM) is used for cell death and all wells have final DMSO concentration of 0.1% (w/w).

#### *384-Well plate library dilution and dosing*

Library plates were supplied as assay ready microplates (10mM stock in DMSO, 150nL per well, Ebner Group, Oxford). The plates were initially diluted in complete media (50uL/well, 30uM, 1% DMSO (v/v)) using a ViaFill liquid dispenser. The plates were transferred to a BioMek Liquid handling Robot and then serially diluted 1:10 to 3uM, 300nM, 30nM, 1% DMSO (v/v)). Staurosporine (10uM) was added to column 1 of each compound plate. Finally, 5uL of each diluted compound was transferred to the cell plates (1: 10 dilution, final volume 50uL, 0.1% DMSO (v/v) in all wells including 'untreated controls', final compound concentrations 3000, 300, 30, 3nM. The plates were incubated at 37°C and 5% CO<sub>2</sub> for 72 hrs.

### **384-Well plate fix/stain/acquisition**

*Materials/Equipment Required* (see [Table S5](#)).

### Fixation

First the plates are removed from the incubator and are allowed to cool for approximately 10 minutes. A fresh solution of 8% formaldehyde in PBS is prepared. An equivalent volume of fixative is added and incubated at RT for 30 minutes. Then, the entire well is aspirated and washed 3 times with PBS (see [Table S6](#)).

### Permeabilisation

For the permeabilization, the plates are washed 2 times with PBST, incubated in PBST at RT for 20 minutes (minimum) and then aspirated (see [Table S7](#)).

### Hoechst staining

For the staining, the plates are washed 3 times in PBS. Then Hoechst is added in PBS (1:5000) and then incubated for 30 min in RT in dark. Following that, the plates are washed 3 times with PBS and Hoechst and are sealed and stored at 4°C. Nuclei counts are carried out via built in software module (Molecular Devices), followed by a preliminary univariate analysis using TIBCO Spotfire Analyst, normalized to DMSO wells/plate (see [Table S8](#)).

## Appendix A

For aspirating and dispensing settings, please see [Table S9](#).

## METHOD DETAILS

### Approved and investigational compound (AIC) collection

We manually selected and curated compounds in the AIC collection based on various sources ([Table S10](#)). The compounds' IDs (from PubChem<sup>58</sup> and ChEMBL<sup>25</sup>) and their structural description (canonical SMILES) were retrieved using PostgreSQL.<sup>59</sup> One compound (TAK-530) was found to have neither a compound ID nor a canonical SMILES, and it was therefore removed. The search for approved and investigational compounds for GBM was carried out manually by searching the clinical trials database (<https://www.clinicaltrials.gov/>), and the availability of these compounds was later determined by crosschecking the compound in SelleckChem (<https://www.selleckchem.com/>) and PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). The ECFP4/6 and MACC fingerprint descriptors of compounds were enumerated using the RDKit<sup>60</sup> cheminformatics module in Python 3.7.<sup>61</sup> Dice similarity for ECP4/6 and Tanimoto for MACCS fingerprints with cutoff of  $\geq 0.99$  was used to identify and remove highly similar compounds<sup>62</sup> (e.g., doxorubicin and epirubicin, with the first compound being removed). The resulting AIC collection consists of a total of 546 unique compounds, available in GitHub (<https://github.com/PaschalisAthar/C3L>).

### Experimental probe compound (EPC) collection

For the EPC collection, compound-target pairs were extracted manually from pan-cancer studies using public databases (PharmacDB<sup>22</sup> and The Human Protein Atlas<sup>21</sup>). The wild type and mutant variants of the targets, along with the first neighbors and influencers of other cancer-related targets,<sup>29</sup> were later used to search for additional compounds that demonstrated sufficient target activity using PostgreSQL<sup>59</sup> (see the below subsections for details). Next, several compound filtering procedures were applied to generate the three EPC collections (theoretical, large-scale and screening set) using RDKit,<sup>60</sup> where the filtering procedures involved checking the structural similarity between the compounds, and Python scripts for additional processing, such as adding extra annotations for the targets or the compounds ([Figure S3](#)). The full list of cancer-associated protein targets and their probe compounds can be found at <https://github.com/PaschalisAthar/C3L>.

### Pan-cancer probe collection (PS1)

This collection of probe compounds focuses on targets and compounds implicated in various types of cancers. To construct this collection, several comprehensive large-scale pan-cancer studies were analyzed (see [Table S11](#)). A set of 946 unique targets and 1525 compounds were curated from the pan-cancer studies using the nominal targets from the PharmacDB database,<sup>22</sup> and The Human Protein Atlas database.<sup>21</sup> After removing redundant compounds, a total of 851 unique compounds remained that covered the target space of 946 unique proteins.

### *Extending the compound space (PS2)*

To further extend the compound space, the annotated primary targets from the pan-cancer studies were queried across publicly available compound bioactivity repositories.<sup>25–27</sup> To include also potential off-targets of the compounds, we used a relatively liberal activity threshold of  $\leq 1000$  nM for multi-dose bioactivity (IC<sub>50</sub>, EC<sub>50</sub>, K<sub>i</sub> or K<sub>d</sub>) to identify the compounds having cellular activity against these targets. In case of multiple entries corresponding to the same compound-target interaction, the median bioactivity value was recorded, similar to a previous study.<sup>65</sup> This curation procedure resulted in 141,087 unique compounds against 441 targets (Table S12).

### *Collection for the mutant target space (PS3)*

In addition to investigating the wild type targets implicated in various cancer types, we also identified compounds with cellular activity against their corresponding mutant variants. The mutation information of the annotated targets of Section Pan-cancer probe collection (PS1) were retrieved from the COSMIC Database.<sup>28</sup> We used various types of variants:

- Substitution - missense
- Substitution - coding silent
- Complex - compound substitution
- Insertion - in frame
- Complex - deletion in frame
- Substitution - nonsense
- Deletion - in frame

We queried these mutant targets across the existing data repositories ChEMBL<sup>25</sup> and DTC,<sup>26</sup> and the corresponding compounds were compiled using the criteria similar to those described above. The resulting mutant collection consists of 944 unique compounds targeting 293 unique mutant targets (Table S13).

### *Extending the target space (PS4)*

To further extend the target space, we queried public repositories for cancer-related targets, their first neighbors and influencers in four cancer types with high mortality rate (colon, breast, hepatocellular, and non-small cell lung cancer). Such extended “targets” are suggested to influence cancer pathogenesis and therefore to increase the drug target space for anticancer therapies.<sup>29</sup> Here, a target was considered as a cancer-related gene, when the corresponding protein was either mutated or had a differential expression in cancer. A first neighbour of cancer-related protein was defined as a protein that is directly and physically interacting with a cancer-related protein in human interactome or signalling networks,<sup>29</sup> according to the databases Signalink v2,<sup>66</sup> Reactome,<sup>67</sup> HPRD,<sup>68</sup> DIP,<sup>69</sup> IntAct,<sup>70</sup> BioGrid,<sup>71</sup> or in a cancer signalling network.<sup>72</sup> An influencer protein was defined as a protein that has a direct interaction to one of the first neighbours. After combining these three subsets, an overlap analysis with the cancer-related targets was performed to identify a total of 1115 unique targets. The non-overlapping targets were queried in ChEMBL<sup>25</sup> and DTC<sup>26</sup> to find a total of 208 653 potent compounds following the same procedure as in the previous sections (Table S14).

### **Reducing the number of compounds**

The full theoretical EPC probe set consists of a total of 336 758 compounds against 1655 unique protein targets (Figure 1). We next reduced the number of compounds to a more manageable size and constructed a compound library that is more attractive to academic screening projects using several filtering procedures, each with freely adjustable cut-off parameters that determine the stringency of the compound filtering process. Figure S4 summarizes the filtering steps and the specific filtering parameters we set in this study to produce the large-scale probe set.

### *Target-specific activity filtering*

The first compound filtering technique was to apply a target-specific activity threshold to reduce the number of compounds from the theoretical set with the following steps:

1. All the activity values were  $\log_{10}$ -transformed.
2. Repeat for each target:
  - 2.1. Target's bioactivity distribution ( $IC_{50}$ ,  $EC_{50}$ ,  $K_i$  or  $K_d$  readouts) was normalized to zero mean and unit variance (z-score normalization).
  - 2.2. A threshold was selected such that 80% of the target's activities (80% percentile) remain within that threshold (see an example in [Table S15](#)).
  - 2.3. Compounds with activities higher than the selected threshold were removed (i.e. these show less potency toward the particular target).

### Compound structural similarity filtering

The next compound filtering step was to find the similarity threshold above which two compounds were considered sufficiently similar. This procedure was based on the assumption that similar compounds are expected to have similar activity distributions. The cut-off value for the Tanimoto similarity was identified with the Akaike Information Criterion, which is an estimator for the degree of information that is lost when using a candidate model, where smaller values indicate low information loss.<sup>73</sup> More specifically, the following procedure was used for defining the similarity cut-off:

1. A portion (here, 10%) of the total number of compounds was chosen for the similarity cut-off estimation to reduce compound space and computation time.
2. Each compound's Dice similarity for ECFP4/6 and Tanimoto coefficient for MACCS fingerprints was computed with the rest of the selected compounds.
3. A coefficient threshold was varied from 0.1 to 0.99 (with 0.01 step size), and for each threshold:
  - 3.1. Compounds that had similarity value equal or greater than the particular threshold were identified.
  - 3.2. The similarity distributions of those compounds were compared using the Kolmogorov-Smirnov (K-S) test.
  - 3.3. The Akaike Information Criterion (AIC) value was calculated based on the K-S test statistic values.
4. The optimal similarity threshold was selected based on the smallest AIC value (see [Figure S5](#) for an example).
5. Steps 1-4 were repeated three times using three different random seeds.

### Global activity filtering

The first step when reducing the size of the physical screening library was to apply an activity threshold similar to that described above, but instead of being target-specific, here the same activity threshold was used across all the targets (i.e., target-agnostic, global filtering). Since two of the theoretical sets are already small in size (PS1 and PS3), the global activity filtering was implemented on the two larger subsets only (PS2 and PS4). More specifically, the procedure followed the steps:

1. The profile of quantitative bioactivity values for each target were recorded when creating the theoretical probe set (see Section 2.1 for details).
2. The bioactivity values were  $\log_{10}$ -transformed and normalized to zero mean.
3. The target with the highest activity standard deviation was identified.
4. An activity threshold was selected such that 95% of the bioactivities of the targets from step 3 are within the selected threshold (see [Figure S6](#)).
5. The selected activity threshold was applied to all the targets to remove the compounds with bioactivities larger than the selected threshold.

### *Reducing the compound space*

The second step in the physical screening library size reduction was to pick only a single compound for each target, with the aim to come up with the smallest number of most potent compounds that cover the same target space. More specifically, the compound that had the lowest bioactivity value among the multi-dose activity types (IC<sub>50</sub>, EC<sub>50</sub>, K<sub>i</sub>, K<sub>d</sub>) was selected for the particular target, since that compound is assumed to have the highest binding potency against the target. The different activity types were treated equally in this process due to sparsity of bioactivity readout data.

### *Compound availability filtering*

The final step for designing the physical screening library was to include only those compounds that are available for sale from at least one vendor using the information from ZINC15.<sup>74</sup> More specifically, if a particular compound was not available for sale, at the time of library curation, then the next most potent compound that is available for sale replaced the original one in the screening library. Even though this step reduced the number of compounds in the final screening library by 52% of the original size, the targets' coverage remained at 86% of the original target space (see [Table S16](#)). Furthermore, the target activity distributions remained relatively unchanged ([Figure S1](#)), and the differences originated mainly from the few of the most potent compounds.

### **Cell survival profiling of glioblastoma patient cells**

We performed high-throughput screening of the GBM patient glioma stem cells with an image based, nuclei staining technique. In this pilot application of the C3L design principles, the cell-based assay was performed using 789 compounds in the physical library, which included 325 of the screening set compounds (27%). The rest of the screening set compounds were not timely available at the time of library curation, or they were prohibitively expensive to purchase from any vendor. To compensate for this, we sourced additional 464 theoretical set compounds, which were selected to maximize the target space. In total, the pilot physical library covered 1320 out of the 1655 targets (80%) of the screening set. In this pilot screening study, we sourced the compounds first through collaborators, and then from commercial providers, by focusing on the most potent compounds against as many of the 1655 protein targets for comprehensive cell-based testing.

### *Patient samples and cell-based assay*

The 789 compounds were tested on 6 patient-derived GBM stem cell models from Glioma Cellular Genetics Resource (<http://gcgr.org.uk>), with 2 patients from each molecular subtype of GBM that are characterized by differences in gene expression and genetic alterations:<sup>33</sup> classical (n=2), proneural (n=2), and mesenchymal (n=2). The protocol comprises of the following steps<sup>13</sup> (see [Figure 4A](#)):

1. Cells were seeded (500-1500 cells/well) onto pre-coated laminin (10 ug/mL) 384-well plates (Greiner Bio-One microclear plates, 781091) using an INTEGRA VIAFILL dispenser. Final laminin concentration was 4ug/mL.
2. The cells were incubated overnight at 37°C, 5% CO<sub>2</sub>.
3. Assay ready compound plates were serially diluted in media using a Biomek liquid handling robot (Beckman). The final concentrations used included 3-30-300-3000nM with DMSO (0.1% w/w) as negative control and staurosporine (1uM) as positive control (for cell death). All wells had a final DMSO concentration of 0.1% (w/w).
4. The cells were incubated for 72 hours.
5. Cells were fixed *in situ* with formaldehyde and then permeabilized with TX-100 (0.01% w/w in PBS). Between steps, plates were washed using a BioTek plate washer (Agilent).
6. Nuclei staining (Hoechst 33342, 2ug/mL in PBS, 20 minutes, room temp).
7. Plates were washed and sealed with aluminium microplate seals (StarSeal, Starlab Group).
8. Image acquisition using ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices).
9. Image analysis and quantification using MetaXpress Software (Molecular Devices, v. 6.7.2.290), analysis module 'Count Nuclei'.



10. Data processing and univariate analysis using TIBCO Spotfire Analyst (PerkinElmer, version 10.10.3.17), normalized to DMSO wells.

#### *Survival fraction, Z score and AUC calculation*

While there are multiple imaging features, such as shape and granularity, that can be quantified for each cellular compartment (nuclei, cytoplasm, etc), here we used only one feature for further analysis: the survival fraction and its z-score (see Figure 4A) from staining the nucleus with Hoechst 33342. The survival fraction is calculated by dividing the nuclei count of the treated sample by the nuclei count of the untreated control sample (DMSO 0.1% (v/v)), and the z-score is the number of standard deviations from the mean survival fraction normalized to DMSO. The lower the survival fraction or the z-score the more effective is the compound in the sense that fewer cancer cells survived the treatment or that the survival score is lower than the mean survival score of the negative controls, respectively. To summarize the activity of a compound across the 4 tested concentrations, we calculated the area under the dose-response curve (AUC) using trapezoidal rule in NumPy<sup>75</sup> for each compound-patient pair using the survival fraction as the response readout (see Figure 4C).

#### *Data-driven activity threshold definition*

To set an activity threshold for strong hits based on the nuclei count measurements, we applied a mixture modelling<sup>76</sup> on the survival fraction distributions, separately for each concentration and patient (but combining all the GBM subtypes due to the small number of samples in each subtype, n=2). As expected for molecularly-targeted compounds, most of the responses appear around 1 in terms of survival fraction (see Figure 4B). To identify extreme compound responses falling outside of these background distributions, we estimated a mixture model with two Gaussian components, one component for the background responses (centered around 1 after parameter estimation), and the other distribution for the extreme responses (parameters freely estimated from the patient-specific survival fraction data). To identify compound responses far enough from the background distribution, we chose a cut-off of 0.01 percentile of the background distribution as the activity threshold to classify a compound as a strong hit (green dotted lines in Figure 4B). In comparison to a fixed threshold based on SF levels, or their z-scores, such data-driven activity threshold ensures that the hit selection does not only depend on the full distribution of the measurements, but the activity threshold is also patient- and concentration-specific, and therefore takes into account potential cell growth and dose-response differences. Finally, the compounds with responses under the activity thresholds across all the 4 concentrations were classified as strong hits for further analyses.

#### *Pathway analyses and web-based data platform*

The target annotations available in C3L were used to map the strong hit compounds to molecular pathways and cellular processes. We used Reactome pathway tool<sup>67</sup> and KEGG pathways<sup>77</sup> (through ShinyGO tool<sup>78</sup>) to explore the pathways that are modulated by the strong hits using the pre-selected activity cut-off of 0.01 percentile of the background distribution (see the section [Data-driven activity threshold definition](#) above). We have also implemented an interactive web-application ([www.c3explorer.com](http://www.c3explorer.com)) that enables users to freely set various compound selection criteria, including data-driven activity threshold for the active compound selection, and then interactively explore drug-target interactions and pathways targeted by the selected compounds for additional analyses and hypothesis generation.

## QUANTIFICATION AND STATISTICAL ANALYSIS

In the section [Compound availability filtering](#), Kolmogorov-Smirnov test was used in Python to test whether there was any difference on the distribution of the compound-target activities before and after replacing the unavailable compounds in the screening set. A significance value of  $p < 0.05$  was considered to indicate a statistically significant difference.

## ADDITIONAL RESOURCES

Data exploration and visualization web-platform: <http://www.c3explorer.com/>.