

1 Title

2

3 Probing the chemical-biological relationship space with the Drug Target Explorer

4

5 Authors

6

7 Robert Allaway<sup>1</sup>, Salvatore La Rosa<sup>2</sup>, Justin Guinney<sup>1</sup>, Sara Gosline<sup>1\*</sup>

8

9 <sup>1</sup>Sage Bionetworks, 1100 Fairview Avenue N, Seattle, WA, 98109

10 <sup>2</sup>Children's Tumor Foundation, New York, NY, 10005

11 \*corresponding author: [sara.gosline@sagebionetworks.org](mailto:sara.gosline@sagebionetworks.org)

12

13 Abstract

14

15 Modern phenotypic high-throughput screens (HTS) present several challenges including  
16 identifying the target(s) that mediate the effect seen in the screen, characterizing 'hits' with a  
17 polypharmacologic target profile, and contextualizing screen data within the large potential  
18 space of drugs and biological screening model combinations. To address these challenges, we  
19 developed an interactive web application that enables exploration of the chemical-biological  
20 interaction space. Compound-target interaction data from public resources were quantified for  
21 over 280,000 molecules. Each molecule was annotated with a name and chemical structure,  
22 and every target was annotated with gene identifiers. The Drug-Target Explorer allows users to  
23 query molecules within this database of experimentally-derived and curated compound-target  
24 interactions and identify structurally similar molecules. It also enables network-based  
25 visualizations of the compound-target interaction space, and incorporates comparisons to  
26 publicly-available *in vitro* HTS datasets. Users can also identify compounds given one or more  
27 targets of interest. The Drug Target Explorer is a multifunctional platform for exploring chemical  
28 space as it relates to biological targets, and may be useful at several steps along the drug  
29 development pipeline including target discovery, structure-activity relationship, and lead  
30 compound identification studies.

31

32

33 Keywords

34 Drug targets, polypharmacology, webapp, phenotypic drug screen, compound-target  
35 network

36

37

38

39

40

41

42

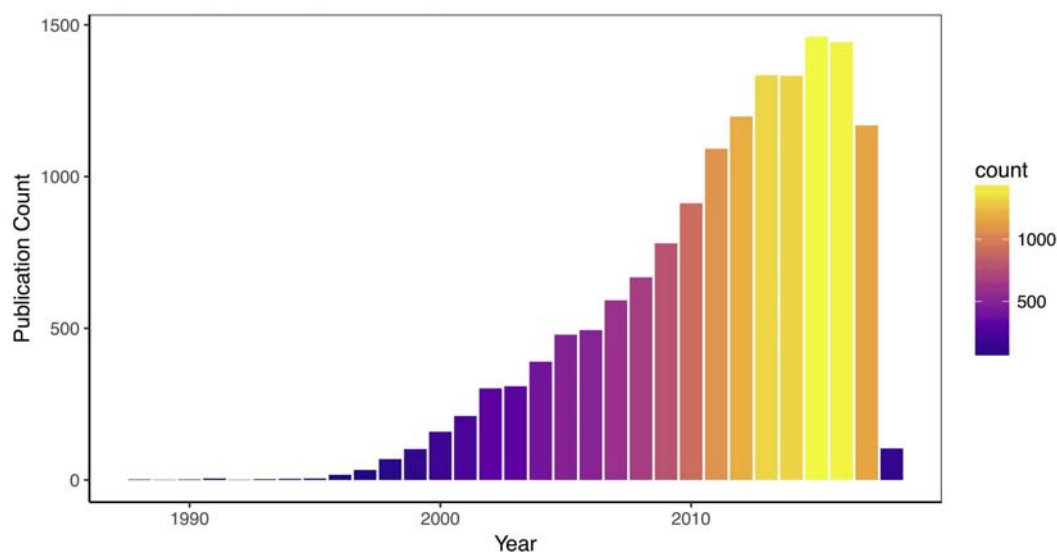
43

44

## 45 Introduction

46

47 In the modern drug discovery and development process, high-throughput screens (HTS)  
48 of drugs have become a common and important step in the identification of novel treatments for  
49 disease. In the past decade, studies describing or citing high throughput drug screening are  
50 increasingly prevalent, topping 1000 per year for the past 5 years (Figure 1) and span many  
51 disease domains such as cancer, neurodegenerative disease, and cardiopulmonary diseases.  
52 These screens are often phenotypic in nature whereby a large panel of compounds of known,  
53 presumed known, and/or unknown mechanisms of action are tested in a biological model of  
54 interest and generate phenotypic readouts such as apoptosis or proliferation. While these types  
55 of screens facilitate the rapid identification of biologically active drugs or chemical probes, they  
56 also present several challenges.



57  
58 **Figure 1 - High throughput drug screening is an increasingly common experimental**  
59 **approach.** Yearly count of Pubmed-indexed publications that appear with the search term “high  
60 throughput drug screening.” Search performed on January 30, 2018.

61  
62 One prevailing challenge is the identification of the specific biological mechanisms within  
63 a cell that determine the response in a screen. The search for novel drugs constantly pushes  
64 the pharmaceutical researchers to include novel chemical sets in phenotypic screens, with the  
65 caveat that the underlying mechanism of action (MoA) of a particular compound cannot usually  
66 be gleaned from the phenotypic screens. (1) Most of the time, identifying the MoA requires  
67 additional experimentation, particularly if the molecule represents a novel or understudied  
68 chemical entity. Another challenge is that the polypharmacologic nature of many small  
69 molecules can make it difficult to interpret HTS results as a given drug may affect multiple  
70 targets with a range of efficacy. This, in turn, presents the difficulty of consolidating multiple  
71 targets into a unified biological mechanism or set of mechanisms leading to poorly annotated  
72 targets, misunderstood MoAs (2), and unknown or ambiguous off-targets with potential deadly

73 side effects (3,4). A final challenge is that identification of related molecules and their targets is  
74 not always straightforward; in the context of HTS analysis, structurally and functionally related  
75 molecules that are not contained in a screening library might be useful to explore.

76 Multiple tools and databases have attempted to address various aspects of the  
77 challenges outlined above (see Table 1). These tools allow the user to explore known  
78 polypharmacology of small molecules. Many also allow users to explore compound-target  
79 relationships by querying either by molecule or by target: DGIdb, DT-Web, BindingDB,  
80 Polypharmacology Browser, STITCH, and SuperTarget allow users to identify MoAs/targets of a  
81 given compound by evaluating a query drug (5–10), while DT-Web, BindingDB,  
82 Polypharmacology Browser, and STITCH allow users to search by chemical similarity using any  
83 query molecule (Table 1). Probe Miner, alternatively, is designed primarily to handle target-  
84 based queries (11). All tools listed in Table 1 allow users to identify molecules with known  
85 polypharmacology, but only two, STITCH and SuperTarget, provide the ability to summarize  
86 these targets into biological pathways/mechanisms using a gene list enrichment approach  
87 (9,10). The final challenge - identifying structurally or functionally related molecules - is  
88 addressed by DT-Web, BindingDB, Polypharmacology Browser, and STITCH (6–9).

89 While several of the tools listed address one or more of these challenges, there are  
90 some gaps (Table 1). For example, ChEMBLSpace does not have a web interface and therefore  
91 requires installation on a compatible system before use (12). In addition, not all of these tools  
92 are open-source (STITCH, SuperTarget, and BindingDB). An easy to modify open-source  
93 application could enable users to create features that are helpful for their specific analyses.  
94 While most tools allow both drug-based and target-based queries, none appear to facilitate  
95 queries for molecules that affect several targets, which may be useful for users who want to  
96 leverage polypharmacology by employing drugs that inhibit multiple biological mechanisms.  
97 While multiple targets can be queried at one time in STITCH, it is not straightforward to identify  
98 single molecules that affect all query targets. In addition, DGIdb and ChEMBLSpace cannot be  
99 used to explore similar chemical space to the query molecule. These two, plus SuperTarget,  
100 also cannot be queried using molecules that are not in the database; a feature that might help  
101 users with novel preclinical candidate drugs. With the exception of DT-Web and STITCH, these  
102 tools do not allow visualization of drug-target networks, which may help users address the  
103 challenge of identifying structurally or functionally related drugs. No tools other than STITCH  
104 perform gene list enrichment, which may help users interpret the biological MoAs of  
105 polypharmacologic molecules.

106 To address these gaps, we developed the Drug-Target Explorer. Specifically, the Drug-  
107 Target Explorer enables the user to (1) look up targets for individual molecules and groups of  
108 molecules, (2) explore networks of targets and drugs, (3) perform gene list enrichment of targets  
109 to assess target pathways of compounds, (4) compare query molecules to cancer cell line  
110 screening datasets, and (5) discover bioactive molecules using a query target and exploration of  
111 these networks. We anticipate that the users will include biologists and chemists involved in  
112 drug discovery who are interested in performing hypothesis generation of human targets for  
113 novel molecules, identifying off-targets for bioactive small molecules of interest, and exploring of  
114 the polypharmacologic nature of small molecules.

	Drug-Target Explorer	Probe Miner	DGIdb v3.0	DT-Web	BindingDB	Polypharmacology Browser	STITCH	ChEMBLSpace	SuperTarget
Web app?	X	X	X	X	X	X	X		X
Open-source software?	X		X	X - underlying R package only	unknown	X		X	
Search by targets to find drugs?	X	X	X	X	X	X - only by PDB-listed ligands	X	X	X
Search by drugs to find targets?	X		X	X	X	X	X		X
Identification of molecules that are associated with multiple query targets?	X			unknown					
Drug structure input?	X			X	X	X	X		
Drug name/ID input?	X		X	X	X	X	X		X
Visualize drug-target networks?	X			X, with user provided drug-target networks	not currently functioning		X		
Identify chemically similar drugs?	X			X, with user provided drug-target networks	X	X	X		X
Allows queries using molecules not in database?	X			X, with user provided drug-target networks	X	X	X		
Target organism?	human	human	human	human	human and others	human and others	human and others	unknown	human and others
Target space?	3.6k	2.2k	6.1k	3.8k	>7k	4.6k	9.6mil	unknown	>6k
Chemical space?	280k	400k	10k	4.4k	>642k	870k	500k	unknown	>196k
Quantitative interactions?	X	X		unknown	X	X	X	X	X
Qualitative interaction?	X	X	X	X			unknown		X
Explore polypharmacology?	X	X	X	X	X	X	X	X	X
Polypharmacologic target enrichment?	X						X		X
Comparison of query molecule to HTS drug response datasets?	X								
Target prediction?				X	X	X			
Database access?	Open	Open	Open	Open	Open	Open	Full database requires license	unknown	unknown
Last known update	2018	2018	2018	2018	2018	2016	2016	2015	2012

115  
116

117 **Table 1 - Summary of selected features/uses of databases and applications for exploring**  
118 **molecule-target relationships and their overlapping features with the Drug-Target**  
119 **Database.** Related tools include Probe Miner (11), DGIdb (5), DT-Web (6), BindingDB (7),  
120 Polypharmacology Browser (8), STITCH (9), ChEMBLSpace (12), and SuperTarget (10).

121  
122

## 123 Results

124

125 The Drug Target Explorer was designed to facilitate the following use-cases: hypothesis  
126 generation of targets for newly-discovered molecules, identification of off-targets for bioactive  
127 research molecules, and exploration of the polypharmacologic nature of many drugs. Below, we  
128 include vignettes highlighting how the Drug-Target Explorer can facilitate analysis in these  
129 areas.

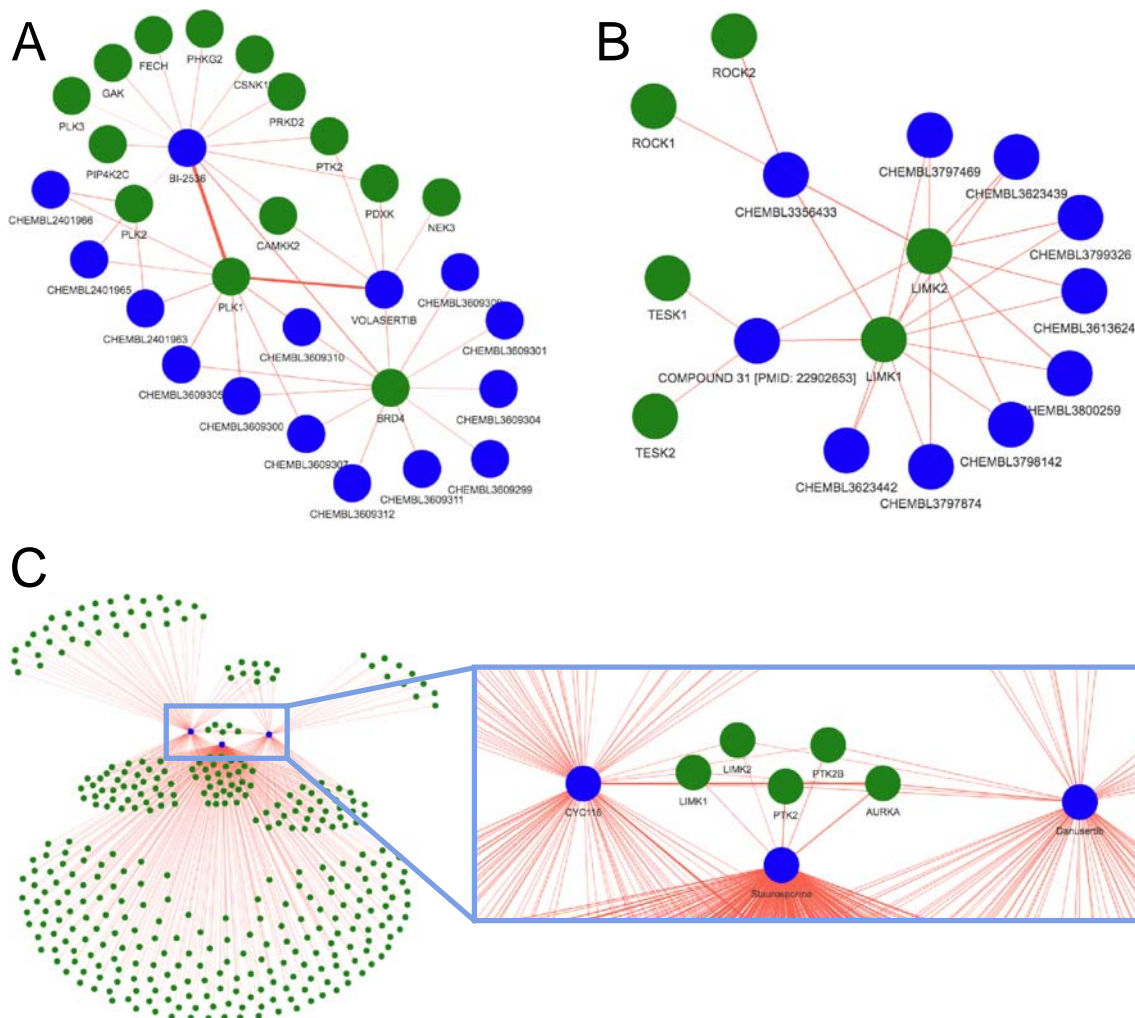
130

### 131 *Identifying potential off-target effects of novel molecules*

132

133 To highlight the use of this app to find potential off-targets of a novel molecule, we  
134 queried the Drug-Target Explorer for C21, a recently-published Polo like kinase (PLK) inhibitor  
135 that is not captured in our database (13). This molecule inhibits Plk2 and Plk1 in the low nM  
136 range, and Plk3 in the low uM range (13). Using a Tanimoto similarity of 0.65 or greater, we  
137 identified 14 molecules (Figure 2A, Supplemental Table 1). PLK1, PLK2, and PLK3 are known  
138 targets of several of these molecules, such as BI 2536 and volasertib. Curiously, CAMKK,  
139 BRD4, PDXK, and PTK2 are also targeted by molecules in this chemical set, with pChEMBL  
140 values >6-8. A plausible hypothesis could be that these targets are affected by this family of  
141 molecules, including the query molecule, in the 10-1000 nM range, which would indicate that  
142 further research is needed to determine the selectivity of C21 or other structurally related  
143 molecules.

144



145  
146  
147 **Figure 2 - Molecule-target networks highlight targets within chemical families.** (A)  
148 Using the novel Plk inhibitor C21 as a query with a Tanimoto cutoff of 0.65 (SMILES:  
149 CCNC(=O)C1=CC2=C(C=C1)N(C=C2)C1=NC=C2N(C)C(=O)[C@@H](CC)N(C3CCCC3)C2=N1), we identify 14 related  
150 molecules (blue vertices), and observe several targets (green vertices) common to multiple  
151 members of this family, including PLK1, PLK2, BRD4, CAMKK2, PTK2, and PDXX. (B) A gene-  
152 based query for two targets (green vertices), LIMK1 and LIMK2, identifies 10 molecules (blue  
153 vertices), as well as other targets affected by these molecules. (C) A query for multiple targets  
154 relevant to tumors caused by neurofibromatosis type 2 identifies three promiscuous molecules  
155 that have associations with these targets.

156  
157 *Identifying off targets of existing molecules*

158  
159 This app may also be useful in identifying off-targets of existing molecules in a  
160 preclinical or exploratory research setting. In order to confidently interrogate the role of cellular  
161 targets, one must use compounds with specificity for those targets. A well-known example of a  
162 non-specific inhibitor is imatinib. This molecule, developed for use in the treatment of chronic



163 myelogenous leukemia, was initially considered a selective inhibitor of Abl (14). More recently,  
164 several other targets have been identified for imatinib such as KIT, PDGFRA, and PDFGRB  
165 (15). Querying the Drug Target Explorer indicates that there is evidence for 61 targets of  
166 imatinib, several of which have pChEMBL values within a reasonable range of Abl, PDGFRA,  
167 and PDFGRB (Supplemental Table 2). These targets must all be considered when evaluating  
168 imatinib in human model systems.

169 A more recent example is the tool compound G-5555, a selective PAK1 inhibitor (16).  
170 This compound has been used to demonstrate the role of PAK1 in cellular processes such as  
171 invasion (17). A search of the Drug-Target Explorer database showed that this molecule not  
172 only binds PAK1 (mean pChEMBL = 8.01), but there is qualitative evidence for effects on  
173 PAK2/3, and quantitative evidence suggesting an effect on SIK2, MAP4K5, and PAK2 at similar  
174 concentrations of G-5555 (mean pChEMBLs 8.05, 8, and 7.96 respectively, Table 2). G-5555  
175 also may have an effect on STK family proteins (STK3, STK24, STK25, STK26) and LCK.  
176 Therefore, any findings with G-5555 with regards to PAK1 inhibition must be validated with other  
177 selective inhibitors or genetic approaches, as Jeannot and colleagues did (using other PAK  
178 inhibitors such as FRAX597 and FRAX1036, as well as PAK1 silencing RNA), to confirm that  
179 the effects observed are PAK1 specific (17).

180

Molecule Name	HGNC Symbol	Mean pChEMBL	n Quantitative	n Qualitative	KSI	Confidence
CHEMBL3770443	PAK1	8.01	3	1	0.1	11
CHEMBL3770443	PAK2	7.96	2	1	0.1	9.96
CHEMBL3770443	SIK2	8.05	1		0.1	9.05
CHEMBL3770443	MAP4K5	8	1		0.1	9
CHEMBL3770443	STK26	7.7	1		0.1	8.7
CHEMBL3770443	STK25	7.47	1		0.1	8.47
CHEMBL3770443	STK24	7.37	1		0.1	8.37
CHEMBL3770443	STK3	7.37	1		0.1	8.37
CHEMBL3770443	LCK	7.28	1		0.1	8.28
CHEMBL3770443	PAK3			1	0.1	1

181

182 **Table 2 – Targets of G-5555 found in the Drug-Target Explorer Database.**

183

184 *Identifying polypharmacologically-targeted pathways and drugs with similar biological*  
185 *effects*

186

187 In order to provide biological context, this app allows the user to aggregate multiple  
188 targets from compounds into functional categories. Using the previous example of G-5555, we  
189 performed enrichment analysis on the list of targets to identify potential biological pathways and  
190 MoAs that this molecule may disrupt. In doing so, we observed that G-5555 targets are enriched  
191 in several Gene Ontology terms and KEGG Pathways like T-cell receptor signaling, Ras/MAPK  
192 signaling, and Golgi-localized proteins (Supplemental Table 3). The app also allows the user to  
193 compare the query molecule to drugs in the Cancer Cell Line/CTRP and GDSC/Sanger cell line  
194 screening datasets. Specifically, the app identifies the most similar molecule available in these

195 datasets and uses that molecule as a reference to plot chemical similarity vs drug response  
196 correlation.

197

198 *Finding a drug for known targets*

199

200 Finally, the tool allows users to perform a reverse search, i.e. identify molecules that  
201 have an association with a query target or targets and assess the known selectivity of these  
202 molecules. For example, Petrilli et. al. identified LIM domain kinases as targets of interest in  
203 tumors caused by the genetic disease neurofibromatosis type 2 (NF2) (18). They found that  
204 pharmacologic (LIMK1/2 inhibitor BMS-5) and genetic modulation of LIMK1 and LIMK2 caused  
205 cell-cycle inhibition and reduced viability in merlin (*Nf2*) deficient Schwann cells (18). In the  
206 context of follow-up and validation studies, it may be beneficial to use alternate molecules that  
207 target LIMK1/2 at the same or greater potency than BMS-5. We used the Drug-Target Explorer  
208 to find molecules that target LIMK1 and LIMK2 (Supplemental Table 4, Figure 2B). For example,  
209 BMS-5 (CHEMBL2141887 in the Drug-Target Explorer) has mean pChEMBLs of 7.33 and 7.07  
210 for LIMK1 and LIMK2 respectively. A good alternative to validate the effects of this molecule  
211 might be CHEMBL3623442, a relatively structurally distinct small molecule (extended fingerprint  
212 Tanimoto similarity of 0.433 to BMS-5 in this database), with pChEMBLs of 9 and 8.52 for  
213 LIMK1 and LIMK2 respectively. Another interesting possibility is the identification of multiple  
214 molecules with overlapping desired targets and non-overlapping off-targets to reduce off-target  
215 effects, or to identify synergistic/additive single-target, multi-drug combinations as outlined by  
216 Fitzgerald et al 2006 (19). Using the above scenario with LIMK1/2, it may be possible to use  
217 structurally distinct molecules in combination or in sequence, like CHEMBL3356433 and  
218 Compound 31 highlighted in Figure 2B, to reduce off-target effects or inhibit LIMK1/2 in an  
219 additive or synergistic manner. The opposite approach could also be taken by finding a single  
220 molecule that binds multiple desired targets. In the case of merlin-deficient cells, focal adhesion  
221 kinases (FAKs) such as PTK2 (FAK2) and PTK2B, as well as Aurora kinase A (AURKA) have  
222 been highlighted as potential targets of interest (18,20,21). Using the Drug-Target Explorer, we  
223 can identify molecules that target LIMK1/2, PTK2/2B, and AURKA (Supplemental Table 5,  
224 Figure 2C). Using this information, a rational hypothesis might be that CYC116 or danusertib  
225 could be effective and selective for *NF2*-deficient tumor cells; to our knowledge, the use of these  
226 molecules in this setting has yet not been explored.

227

228

## 229 Discussion

230

231 In the present study, we demonstrate that the Drug-Target Explorer enables the user to  
232 look up targets for novel and known molecules such as C21, G-5555, and imatinib, as well as  
233 explore networks of these drugs and their targets. Users can perform target enrichment to  
234 consolidate multiple targets into pathways, compare query molecules to screening datasets,  
235 and identify bioactive molecules given a query target.

236 Several future directions are envisioned for this application. The code and database has  
237 been designed in such a way that any database with structural information and drug-gene target  
238 information (qualitative associations, or quantitative associations that can be coerced to

239 pChEMBL values) can be harmonized and integrated into the database. Therefore, as new  
240 datasets become available, such as the recently-published Drug-Target Commons (22), they  
241 can be integrated and released. We also envision occasional errors being identified as the  
242 database is explored and vetted by users and have included a feedback form for users to  
243 suggest new data to integrate, as well as to highlight necessary corrections to the dataset.  
244 Currently, the query molecule to full database similarity calculation is computationally intensive.  
245 One solution to speed up calculation times may be to implement a locality sensitive hashing  
246 method in future versions of the database and web app, such as the method devised by Cao et  
247 al 2010 (23). An additional planned feature for this app is the implementation of a bulk  
248 annotation feature to allow users to annotate HTS data with targets and/or putative targets of  
249 identical or structurally related molecules. Finally, the integration of a predictive framework for  
250 identifying targets of query drugs based on drug and target feature data would enable users to  
251 quantitatively predict targets of novel molecular entities rather than manually exploring  
252 structurally similar molecules.

253 The Drug-Target Explorer enables users to explore known molecule-human target  
254 relationships as they relate to chemical similarity rapidly and with minimal effort. We anticipate  
255 that users such as biologists and chemists using chemical probes or studying preclinical  
256 therapeutics will find this tool useful in several areas. Specifically, this tool may aid drug  
257 discovery efforts by accelerating hypothesis generation, simplifying the transition from  
258 phenotypic HTS results to mechanistic studies, and streamlining the identification of candidate  
259 molecules that target a protein or mechanism of interest.

260

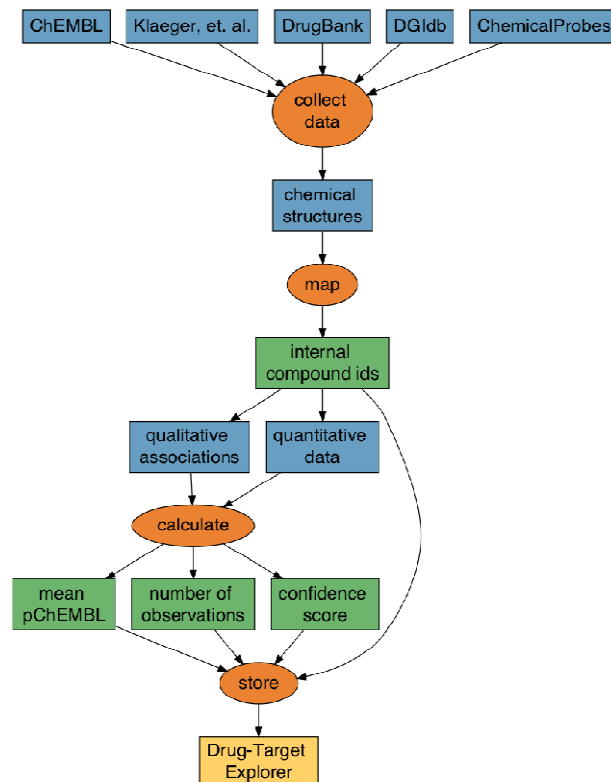
## 261 **Methods**

262

263 To build the database of known compound-target interactions, we aggregated five data  
264 sources containing qualitative and quantitative interactions (Figure 3). We considered qualitative  
265 interactions to be curated compound-target associations with no associated numeric value.  
266 Quantitative interactions were defined as compound-target information with a numeric value  
267 indicating potency of compound-target binding or functional changes. Qualitative compound-  
268 target associations were retrieved from the DrugBank 5.0.11 XML database, the DGIdb v3.0.1  
269 interactions.tsv file, and ChemicalProbes.org (acc. Jan 17 2018) (5,24,25). pChEMBL, IC50,  
270 C50, EC50, AC50, Ki, Kd, and potency values for *Homo sapiens* targets were retrieved from the  
271 ChEMBL v23 MySQL database (26). Kd values were also obtained from Klaeger et al 2017, in  
272 which the authors determined the Kd of 244 kinase inhibitors against 343 kinases (27). For all  
273 quantitative and qualitative data sources, compound structural information (SMILES) was  
274 retrieved when available. When not available, it was batch annotated using the Pubchem  
275 Identifier Exchange Service, or, in some cases, manually annotated via PubChem and  
276 ChemSpider search (28,29).



277



278

279 **Figure 3 - Process for developing the Drug-Target Explorer.** Molecule-target and chemical  
280 structure data were collected from public sources. In the case of DGldb, chemical structures  
281 were assigned using the PubChem Chemical Identifier Exchange or manually assigned using  
282 ChemSpider and PubChem. Chemical structures were converted to circular fingerprints and the  
283 databases were mapped to internal Drug-Target Explorer identifiers. Qualitative and quantitative  
284 data were summarized by calculating several summary statistics, and these data were stored  
285 together with the internal identifiers to form the Drug-Target Explorer database.

286

287 To consolidate data for “identical” molecules within and across multiple databases, the  
288 functional connectivity fingerprint (FCFP6)-like ‘circular’ fingerprint for each SMILES was  
289 calculated using the R interface (rcdk) to the Java Chemical Development Kit (CDK) (30–32).  
290 The package was modified to use the latest version of the CDK (2.1.1), which enables  
291 perception of chiral centers, enabling differentiation between isomeric molecules. Each unique  
292 circular fingerprint and all external IDs and SMILES associated with that fingerprint were then  
293 assigned an internal identifier, so that groups of molecules with identical fingerprints were  
294 assigned to the same internal ID. The internal molecular IDs were then mapped to each  
295 database to permit their aggregation. All datasets were combined and summaries were  
296 generated for each compound-target comparison using functions from the R ‘tidyverse’ (33).

297

298 The summary metrics described in Table 3 were calculated. One of these metrics,  
pChEMBL, is used to convey the efficacy of a given molecule. It is calculated from one of

299 several semi-comparable values in the ChEMBL database, and is defined as the negative log  
300 10 molar of the IC50, XC50, EC50, AC50, Ki, Kd, or potency (26). For example, a pChEMBL  
301 value of 7 would indicate that there is a measurable effect on a given target in the presence of  
302 100 nM of molecule. To harmonize the data from Klaeger et al with ChEMBL data, the Kd  
303 values were converted to pChEMBLs. The mean pChEMBL was calculated for every molecule-  
304 target combination, as well as the number of quantitative and qualitative associations found in  
305 the source databases.  
306

Metric	Unit	Meaning
mean IC50/AC50/EC50/C50/Potency/Ki/Kd	nM	mean of values obtained from quantitative datasets; available in database but not app
mean pChEMBL	-log <sub>10</sub> (nM)	mean -log <sub>10</sub> (nM) of all semi-comparable quantitative values
n_qualitative	count	number of qualitative associations identified
n_quantitative	count	number of quantitative associations identified
known selectivity score	N/A	1 divided by the number of known targets, lower is less selective
confidence score	N/A	mean pChEMBL, multiplied by n_quantitative, plus n_qualitative

307  
308 **Table 3 - Drug-target association metrics summarized in the Drug-Target Explorer**  
309 **database.**  
310

311 We calculated a known selectivity score for each molecule, which we defined as 1  
312 divided by the total number of targets for that molecule (lower values correspond to lower  
313 molecule selectivity), and a confidence score for each molecule-target relationship, which we  
314 defined as the mean pChEMBL multiplied by the number of quantitative measurements, in  
315 addition to the number of qualitative annotations. A larger confidence score indicates greater  
316 confidence in this relationship; this confidence is weighted by the potency to give increased  
317 preference to high-potency compound-target interactions.

318 This resulted in a database containing 3645 human targets (represented by HUGO gene  
319 symbols), ~280,000 small molecules, and ~623,500 molecule-target relationships summarized  
320 from ~598,000 quantitative associations and ~25,000 qualitative associations. Finally, this  
321 database as well as fingerprints and chemical aliases for each molecule were saved as R binary  
322 files and stored on Synapse. All of the data, as well as snapshots of the source databases used  
323 to build the Drug Target Explorer database (with the exception of DrugBank, which requires a  
324 license to access) are accessible at [www.synapse.org/dtexplorer](http://www.synapse.org/dtexplorer). The Drug-Target Database is  
325 licensed under CC BY-SA 4.0.  
326  
327  
328  
329  
330  
331  
332



358 <https://github.com/Sage-Bionetworks/polypharmacology-db>. The source code is licensed under  
359 Apache 2.0.

360

361

362 Supplemental table legends:

363

364 **Supplemental Table 1 – Targets of C21-like compounds in the Drug-Target Explorer**  
365 **Database.**

366

367 **Supplemental Table 2 – Targets of imatinib in the Drug-Target Explorer Database.**

368

369 **Supplemental Table 3 – Target enrichment analysis of G-5555 highlights putative**  
370 **mechanistic effects.** G-5555 targets were enriched in multiple Gene Ontology terms and  
371 KEGG pathways.

372

373 **Supplemental Table 4 – Molecules targeting LIMK1/2.** The database was queried for  
374 molecules that may modulate LIMK1 and LIMK2; this analysis revealed a large set of putative  
375 tool compounds.

376

377 **Supplemental Table 5 – Identification of multi-kinase-targeting molecules for NF2.** A query  
378 of the database for molecules that target several kinases of interest in NF2 (AURKA, LIMK1/2,  
379 PTK2/2B) identified 3 polypharmacologic compounds.

380

381 References

382

- 383 1. Wagner BK. The resurgence of phenotypic screening in drug discovery and development.  
384 Expert Opin Drug Discov [Internet]. 2016 Feb 10 [cited 2018 Mar 13];11(2):121–5.  
385 Available from: <http://www.tandfonline.com/doi/full/10.1517/17460441.2016.1122589>
- 386 2. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive  
387 map of molecular drug targets. Nat Rev Drug Discov [Internet]. 2017 Jan 2 [cited 2018  
388 Mar 26];16(1):19–34. Available from: <http://www.nature.com/articles/nrd.2016.230>
- 389 3. van Esbroeck ACM, Janssen APA, Cognetta AB, Ogasawara D, Shpak G, van der Kroeg  
390 M, et al. Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor  
391 BIA 10-2474. Science [Internet]. 2017 Jun 9 [cited 2018 Mar 26];356(6342):1084–7.  
392 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28596366>
- 393 4. Roy M, Dumaine R, Brown AM. HERG, a primary human ventricular target of the  
394 nonsedating antihistamine terfenadine. Circulation [Internet]. 1996 Aug 15 [cited 2018  
395 Mar 26];94(4):817–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8772706>
- 396 5. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a  
397 redesign and expansion of the drug–gene interaction database. Nucleic Acids Res  
398 [Internet]. 2017 Nov 16 [cited 2018 Mar 20]; Available from:  
399 <http://academic.oup.com/nar/article/doi/10.1093/nar/gkx1143/4634012>
- 400 6. Alaimo S, Bonnici V, Cancemi D, Ferro A, Giugno R, Pulvirenti A. DT-Web: a web-based  
401 application for drug-target interaction and drug combination prediction through domain-  
402 tuned network-based inference. BMC Syst Biol [Internet]. 2015 [cited 2018 Mar  
403 20];9(Suppl 3):S4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26050742>
- 404 7. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public

- 405 database for medicinal chemistry, computational chemistry and systems pharmacology.  
406 Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2018 Mar 20];44(D1):D1045–53. Available  
407 from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1072>
- 408 8. Awale M, Reymond J-L. The polypharmacology browser: a web-based multi-fingerprint  
409 target prediction tool using ChEMBL bioactivity data. J Cheminform [Internet]. 2017 Dec  
410 21 [cited 2018 Mar 20];9(1):11. Available from:  
411 <http://jcheminf.springeropen.com/articles/10.1186/s13321-017-0199-x>
- 412 9. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5:  
413 augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic  
414 Acids Res [Internet]. 2016 Jan 4 [cited 2018 Mar 20];44(D1):D380-4. Available from:  
415 <http://www.ncbi.nlm.nih.gov/pubmed/26590256>
- 416 10. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, et al. SuperTarget  
417 and Matador: resources for exploring drug-target relationships. Nucleic Acids Res  
418 [Internet]. 2008 Jan [cited 2018 Mar 20];36(Database issue):D919-22. Available from:  
419 <http://www.ncbi.nlm.nih.gov/pubmed/17942422>
- 420 11. Antolin AA, Tym JE, Komianou A, Collins I, Workman P, Al-Lazikani B. Objective,  
421 Quantitative, Data-Driven Assessment of Chemical Probes. Cell Chem Biol [Internet].  
422 2018 Feb 15 [cited 2018 Apr 25];25(2):194–205.e5. Available from:  
423 <http://www.ncbi.nlm.nih.gov/pubmed/29249694>
- 424 12. Fechner N, Papadatos G, Evans D, Morphy JR, Brewerton SC, Thorner D, et al.  
425 ChEMBLSpace—a graphical explorer of the chemogenomic space covered by the  
426 ChEMBL database. Bioinformatics [Internet]. 2013 Feb 15 [cited 2018 Mar 20];29(4):523–  
427 4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23257198>
- 428 13. Zhan M-M, Yang Y, Luo J, Zhang X-X, Xiao X, Li S, et al. Design, synthesis, and  
429 biological evaluation of novel highly selective polo-like kinase 2 inhibitors based on the  
430 tetrahydropteridin chemical scaffold. Eur J Med Chem [Internet]. 2018 Jan 1 [cited 2018  
431 Feb 1];143:724–31. Available from:  
432 <https://www.sciencedirect.com/science/article/pii/S0223523417309613>
- 433 14. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, et al. Effects of a  
434 selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat  
435 Med [Internet]. 1996 May [cited 2018 Feb 8];2(5):561–6. Available from:  
436 <http://www.ncbi.nlm.nih.gov/pubmed/8616716>
- 437 15. Pardananani A, Tefferi A. Imatinib targets other than bcr/abl and their clinical relevance in  
438 myeloid disorders. Blood [Internet]. 2004 Oct 1 [cited 2018 Feb 8];104(7):1931–9.  
439 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15166033>
- 440 16. Ndubaku CO, Crawford JJ, Drobnick J, Aliagas I, Campbell D, Dong P, et al. Design of  
441 Selective PAK1 Inhibitor G-5555: Improving Properties by Employing an Unorthodox  
442 Low-p  $K_a$  Polar Moiety. ACS Med Chem Lett [Internet]. 2015 Dec 10 [cited 2018 Feb  
443 8];6(12):1241–6. Available from:  
444 <http://pubs.acs.org/doi/10.1021/acsmedchemlett.5b00398>
- 445 17. Jeannot P, Nowosad A, Perchey RT, Callot C, Bennana E, Katsube T, et al. p27Kip1  
446 promotes invadopodia turnover and invasion through the regulation of the PAK1/Cortactin  
447 pathway. Elife [Internet]. 2017 Mar 13 [cited 2018 Feb 8];6. Available from:  
448 <http://www.ncbi.nlm.nih.gov/pubmed/28287395>
- 449 18. Petrilli A, Copik A, Posadas M, Chang L-S, Welling DB, Giovannini M, et al. LIM domain  
450 kinases as potential therapeutic targets for neurofibromatosis type 2. Oncogene  
451 [Internet]. 2014 Jul 12 [cited 2018 Feb 20];33(27):3571–82. Available from:  
452 <http://www.ncbi.nlm.nih.gov/pubmed/23934191>
- 453 19. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination  
454 therapy in the quest for clinical efficacy. Nat Chem Biol [Internet]. 2006 Sep 18 [cited  
455 2018 Feb 21];2(9):458–66. Available from:



- 456 <http://www.nature.com/doi/10.1038/nchembio817>
- 457 20. Poulidakos PI, Xiao G-H, Gallagher R, Jablonski S, Jhanwar SC, Testa JR. Re-  
458 expression of the tumor suppressor NF2/merlin inhibits invasiveness in mesothelioma  
459 cells and negatively regulates FAK. *Oncogene* [Internet]. 2006 Sep 1 [cited 2018 Apr  
460 11];25(44):5960–8. Available from: <http://www.nature.com/articles/1209587>
- 461 21. Shapiro IM, Kolev VN, Vidal CM, Kadariya Y, Ring JE, Wright Q, et al. Merlin deficiency  
462 predicts FAK inhibitor sensitivity: a synthetic lethal relationship. *Sci Transl Med* [Internet].  
463 2014 May 21 [cited 2018 Feb 21];6(237):237ra68. Available from:  
464 <http://www.ncbi.nlm.nih.gov/pubmed/24848258>
- 465 22. Tang J, Tanoli Z-U-R, Ravikumar B, Alam Z, Rebane A, Vähä-Koskela M, et al. Drug  
466 Target Commons: A Community Effort to Build a Consensus Knowledge Base for Drug-  
467 Target Interactions. *Cell Chem Biol* [Internet]. 2018 Feb 15 [cited 2018 Mar  
468 21];25(2):224–229.e2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29276046>
- 469 23. Cao Y, Jiang T, Girke T. Accelerated similarity searching and clustering of large  
470 compound sets by geometric embedding and locality sensitive hashing. *Bioinformatics*  
471 [Internet]. 2010 Apr 1 [cited 2018 Feb 14];26(7):953–9. Available from:  
472 <http://www.ncbi.nlm.nih.gov/pubmed/20179075>
- 473 24. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a  
474 major update to the DrugBank database for 2018. *Nucleic Acids Res* [Internet]. 2018 Jan  
475 4 [cited 2018 Mar 20];46(D1):D1074–82. Available from:  
476 <http://www.ncbi.nlm.nih.gov/pubmed/29126136>
- 477 25. ChemicalProbes.org [Internet]. 2018 [cited 2018 Jan 17]. Available from:  
478 [www.chemicalprobes.org](http://www.chemicalprobes.org)
- 479 26. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL  
480 database in 2017. *Nucleic Acids Res* [Internet]. 2017 Jan 4 [cited 2018 Mar  
481 20];45(D1):D945–54. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27899562>
- 482 27. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, et al. The target  
483 landscape of clinical kinase drugs. *Science* [Internet]. 2017 Dec 1 [cited 2018 Mar  
484 20];358(6367):eaan4368. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29191878>
- 485 28. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance  
486 and Compound databases. *Nucleic Acids Res* [Internet]. 2016 Jan 4 [cited 2018 Mar  
487 20];44(D1):D1202–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26400175>
- 488 29. Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. *J Chem*  
489 *Educ* [Internet]. 2010 Nov [cited 2018 Mar 20];87(11):1123–4. Available from:  
490 <http://pubs.acs.org/doi/abs/10.1021/ed100697w>
- 491 30. Guha R. Chemical Informatics Functionality in R. *J Stat Softw*. 2007;18(6).
- 492 31. Willighagen EL, Mayfield JW, Alvarsson J, Berg A, Carlsson L, Jeliaskova N, et al. The  
493 Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and  
494 substructure searching. *J Cheminform* [Internet]. 2017 Dec 6 [cited 2018 Mar 22];9(1):33.  
495 Available from: <http://jcheminf.springeropen.com/articles/10.1186/s13321-017-0220-4>
- 496 32. Guha R. fingerprint: Functions to Operate on Binary Fingerprint Data [Internet]. 2017.  
497 Available from: <https://cran.r-project.org/package=fingerprint>
- 498 33. Wickham H. tidyverse: Easily Install and Load the “Tidyverse” [Internet]. 2017. Available  
499 from: <https://cran.r-project.org/package=tidyverse>
- 500 34. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. shiny: Web Application Framework  
501 for R [Internet]. 2017. Available from: <https://cran.r-project.org/package=shiny>
- 502 35. R Core Team. R: A Language and Environment for Statistical Computing [Internet].  
503 Vienna, Austria; 2018. Available from: <https://www.r-project.org/>
- 504 36. Szöcs E. {webchem}: retrieve chemical information from the web [Internet]. Zenodo.  
505 2015. Available from: <http://dx.doi.org/10.5281/zenodo.33823>
- 506 37. Kuleshov M V, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr:

507 a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids*  
508 *Res* [Internet]. 2016 [cited 2018 Mar 20];44(W1):W90-7. Available from:  
509 <http://www.ncbi.nlm.nih.gov/pubmed/27141961>  
510 38. Jawaid W. enrichR: Provides an R Interface to “Enrichr” [Internet]. 2017. Available from:  
511 <https://cran.r-project.org/package=enrichR>  
512